مجله مدل‌سازی پیشرفته ریاضی
سال ۱۴۰۱، دوره ۱۲، شماره ۳، ص ۳۴۴-۳۵۶

شاپا چاپی: ۲۲۵۱-۸۰۸۸
DOI: 10.22055/JAMM.2022.40270.2018

برآورد منحنی مشخصه عملکرد گیرنده (ROC) با استفاده از تابع هسته برنام–سندرز

حبیب‌اله ممبینی، بهزاد منصوری ×، محمدرضا آخوند

گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز، اهواز، ایران

دبیر مسئول: غلامرضا محتشمی برزادران

تاریخ دریافت: ۱۴۰۰/۱۲/۲۰ تاریخ پذیرش: ۱۴۰۱/۶/۱۲

چکیده: بسیاری از محققان از منحنی ROC به عنوان روشی کارآمد برای نمایش، ارزیابی و مقایسه دقت آزمون‌های تشخیصی استفاده می‌کنند. متداول‌ترین روش برای برآورد منحنی ROC استفاده از برآوردگر ناپارمتری هسته برای برآورد توابع توزیع تجمعی احتمال در دو بخش حساسیت و ویژگی است. با این وجود برآوردگرهای هسته در نقاط ابتدایی و انتهایی دامنه داده‌ها که به عنوان نقاط مرزی شناخته می‌شوند، نسبت به دیگر نقاط دامنه، دارای نرخ همگرایی کندتری بوده و به مقدار واقعی تابع توزیع احتمال همگرا نیستند. این مشکل را اصطلاحا مشکل مرزی می‌گویند. یک روش برای رفع مشکل مرزی در برآوردگرهای هسته، استفاده از هسته‌های نامتقارن است. در این مقاله، یک برآوردگر جدید برای منحنی ROC، بر اساس تابع هسته نامتقارن برنام–سندرز (B-S) پیشنهاد شده و همگرایی مجانبی برآوردگر پیشنهادی نشان داده شده است. علاوه بر این، برتری تحلیلی برآوردگر پیشنهادی نسبت به برآوردگر نوع هسته متقارن نشان داده شده است. عملکرد برآوردگر پیشنهادی از طریق یک مطالعه عددی بررسی و با دیگر برآوردگرهای مطرح منحنی ROC مقایسه شده است. نتایج نشان می‌دهد که ریسک برآوردگر پیشنهادی به صورت قابل ملاحظه‌ای پایین‌تر از سایر روش‌های معمول است. کاربرد برآوردگر جدید در یک مجموعه داده پزشکی نشان داده شده است.

واژه‌های کلیدی: تابع توزیع احتمال، برآوردگر هسته، هسته نامتقارن، منحنی ROC

رده‌بندی ریاضی: 62G07, 62P10

۱ مقدمه

منحنی مشخصه عملکرد گیرنده یا به اختصار منحنی ROC نمودار نرخ مثبت صحیح یا «حساسیت» یک آزمون را در مقابل نرخ مثبت کاذب یا «ویژگی» آن آزمون نشان می‌دهد. از منحنی ROC معمولاً برای توصیف، تجسم و مقایسه عملکرد آزمون‌های تشخیصی که در آن‌ها موضوع‌ها به یکی از دو یا چند گروه مختلف طبقه بندی می‌شوند، استفاده می‌شود. به عنوان مثال، در پزشکی، منحنی ROC ابزاری مفید

×نویسنده مسئول مقاله
رایانامه: b.mansouri@scu.ac.ir (B. Mansouri)

برای ارزیابی آزمایش تشخیص بیمار یا سالم بودن فرد است. پیدایش منحنی ROC به تشخیص بین سیگنال و نویز در مهندسی باز می‌گردد (گرین و اسوتس، ۱۹۶۶). با این وجود، سادگی و کارایی منحنی ROC کاربردهای آن را در علوم مختلف مانند پزشکی، یادگیری ماشین، داده کاوی، اقتصاد و بسیاری دیگر از زمینه‌ها را فراهم کرده است (فاوست (۲۰۰۶) و لاسکو و همکاران (۲۰۰۵)). فاوست (۲۰۰۶) منبع ارزشمندی برای مطالعه جنبه‌های مختلف منحنی ROC است.

فرض کنید که X و Y مشخص کننده امتیاز اعضا به ترتیب در گروه ۱ و گروه ۲ باشند. علاوه بر آن فرض کنید که X و Y دو متغیر تصادفی با توابع توزیع احتمال پیوسته و مجهول $F_1(x)$ و $F_2(y)$ باشند. برای تشخیص آنکه یک عضو متعلق به کدام یک از دو گروه ۱ یا ۲ است از یک آزمون استفاده می‌کنیم به طوریکه برای نقطه برش $c \in \mathbb{R}$، اگر امتیاز یک عضو بیشتر از c باشد، عضو به گروه ۱ تعلق می‌گیرد و در غیر اینصورت به گروه ۲ تعلق می‌گیرد. در اینصورت حساسیت آزمون به‌صورت $ES(c) = 1 - F_2(c)$ و ویژگی آزمون به‌صورت $SP(c) = F_1(c)$ تعریف می‌شود. از منظر آماری مسئله را می‌توان به‌عنوان یک آزمون فرض آماری دید که در آن فرض صفر (H_0) تعلق عضو به گروه ۱ و فرض مقابل (H_1) تعلق عضو به گروه ۲ است. در اینصورت $F_1(c) = 1 - \alpha$ است که در آن α احتمال خطای نوع اول است و $F_2(c) = 1 - \beta$ که در آن β احتمال خطای نوع دوم است. منحنی ROC یک گراف دو بعدی است که در آن SE(c) بر روی محور عمودی و SP(c) بر روی محور افقی نمایش داده می‌شود. به عبارت دیگر منحنی ROC توان آزمون را در برابر سطح معنی دار آزمون، به ازای مقادیر نقطه برش c در فاصله $-\infty$ تا ∞ به تصویر می‌کشد. با تعریف $t = 1 - F_1(c) = 1 - SP(c)$ و $ROC(t) = 1 - \beta = F_2(c) = SE(c)$ به ازای $0 \leq t \leq 1$ داریم

$$ROC(t) = 1 - F_2\left(F_1^{-1}(1-t)\right)$$

در عمل ROC(t) مجهول است زیرا $F_1(x)$ و $F_2(y)$ مجهول هستند.

هسیه و ترنبول (۱۹۹۶) یک منحنی ROC تجربی را پیشنهاد دادند که در آن توابع توزیع مجهول با تابع توزیع تجربی برآورد می‌شوند. اگرچه منحنی ROC تجربی یک برآورد یکنواخت سازگار از منحنی ROC را در فاصله $[0, 1]$ ارائه می‌کند اما برآوردهای ارائه شده توسط ROC تجربی هموار نیستند. روش دیگر برای برآورد منحنی ROC استفاده از برآوردگرهای نوع هسته است. برخی نویسندگان، برآورد منحنی ROC را با استفاده از برآورد توابع چگالی $f_1(x)$ و $f_2(y)$ به روش هسته و بر اساس هسته‌های متقارن، را پیشنهاد کرده‌اند (زو و همکاران ۱۹۹۷). لوید (۱۹۹۸) از روش برآورد مستقیم $F_1(x)$ و $F_2(y)$ استفاده کرد. لوید و یونگ (۱۹۹۹) درباره برتری مجانبی برآوردگر لوید (برآوردگر لوید) نسبت به ROC تجربی در یک نقطه بحث کردند. دو و تانگ (۲۰۰۹) و تانگ و همکاران (۲۰۱۰) مسئله برآورد ناپارامتری تبدیل–پایای[†] منحنی ROC را بررسی کردند. پولیت (۲۰۱۶) نیز برآوردگر جدیدی از نوع هسته را برای منحنی ROC پیشنهاد کرد که نسبت به یک تبدیل غیرنزولی[‡] پایا است. او نشان داد که برآوردگرش (برآوردگر پولیت) از نظر میانگین مجذور مربع خطای مجانبی[§] بهتر از برخی دیگر از برآوردگرهای نوع هسته است. مزیت دیگر برآوردگر پولیت این است که فقط به یک پارامتر هموار کننده بستگی دارد.

دانگ (۲۰۱۶) با توسعه برآوردگر هسته تابع توزیع چند متغیره، یک روش جدید برای برآورد تابع ROC چند متغیره پیشنهاد کرد. واقعیت آن است که عملکرد کلی هر برآوردگر نوع هسته برای منحنی ROC تا حد زیادی وابسته به نحوه برآورد دو تابع توزیع مجهول $F_1(x)$ و $F_2(y)$ است. اغلب برآوردگرهای متداول، مانند برآوردگر لوید یا برآوردگر پولیت از یک هسته متقارن برای برآورد تابع توزیع تجمعی احتمال استفاده می‌کنند. با این وجود، به واسطه اریبی مرزی، این نوع هسته‌ها برای برآورد توابعی که دامنه محدود دارند، کارا نیستند (تنریرو، ۲۰۱۳). در برآورد هسته این مشکل را با عنوان مسئله مرزی می‌شناسند و تاکنون چندین رویکرد برای رفع آن در رگرسیون ناپارامتری و برآورد چگالی ناپارامتری پیشنهاد شده است (وایزرمن، ۲۰۰۶ و ژانگ و همکاران، ۱۹۹۹). این رویکردها همچنین برای برآورد تابع توزیع با دامنه محدود نیز گسترش یافته‌اند (تنریرو، ۲۰۱۳ و ۲۰۱۸). مسئله مرزی در برآورد منحنی ROC توسط کولاچک و کارونامونی (۲۰۰۹) مطالعه شده است. آنها یک برآوردگر اصلاح شده مرزی (برآوردگر K-K) را برای منحنی ROC پیشنهاد کردند. یک مسئله دیگر در برآوردگرهای نوع هسته انتخاب پارامتر هموار کننده است. چون شکل تابع هسته در برآوردگرهای نوع هسته متقارن ثابت است؛ این نوع برآوردگرها به تغییر پارامتر هموار کننده بسیار حساس هستند. تنظیم پارامتر هموار کننده برای کشف ساختار داده‌ها در نواحی از دامنه که تراکم داده‌ها زیاد است و به طور همزمان ممانعت از بروز ساختار کاذب در جاهایی که تراکم داده‌ها کم است؛ کار دشواری است. سیلورمن (۱۹۸۶) این مشکل را با جزئیات توصیف کرده و برای رفع آن چند روش پیشنهاد کرده است. (چن ۱۹۹۹ و ۲۰۰۰) برآوردگرهای هسته نامتقارن را در برآورد ناپارامتری رگرسیون و چگالی معرفی کرد. هسته‌های نامتقارن مانند هسته گاما، هسته بتا و هسته برنام– سندرز برخلاف برآوردگرهای متقارن، دارای شکلی منعطف هستند که با نقطه طرح (نقطه‌ای که در آن تابع چگالی یا توزیع برآورد می‌شود) تغییر می‌کنند. برای داده‌هایی که دامنه محدود دارند این خصوصیت به همراه دامنه این توابع هسته که منطبق بر دامنه داده‌ها است، سبب رفع مشکل مرزی می‌شود. همچنین این توابع هسته به اندازه توابع هسته متقارن به مشکل تنظیم پارامتر هموار کننده در نواحی با تراکم کم و زیاد حساس نیستند. اخیرا ممبینی و همکاران (۲۰۲۱)، منصوری و همکاران (۲۰۲۲) و لافایه دمیشو و اویمیت (۲۰۲۱) برآوردگرهای جدیدی بر پایه هسته‌های نامتقارن را برای برآورد تابع توزیع تجمعی احتمال پیشنهاد کرده‌اند. آنها مفید بودن این برآوردگرها را در رفع مسئله مرزی در برآورد تابع توزیع نشان

[†] Transformation-invariant

[‡] Nondecreasing

[§] Asymptotic mean squared error

داده‌اند. با داشتن مشاهدات $X_1, ..., X_m$ برای برآورد تابع توزیع $F(x)$ با دامنه $[0, \infty)$ برآوردگر آنها بفرم

$$\hat{F}(x) = m^{-1} \sum_{i=1}^{m} \bar{K}_{x,b}(X_i), \tag{۱.۱}$$

است که در آن $\bar{K}_{x,b}(t) = \int_t^\infty k_{x,b}(u)du$ و $k(\cdot)$ یک تابع هسته نامتقارن با دامنه $[0, \infty)$ و b نیز پارامتر هموار کننده است. بنابراین دامنه تابع توزیع و تابع هسته یکسان است. با توجه به عملکرد بهتر برآوردگر هسته برنام- سندرز در مقایسه با دیگر برآوردگرهای موجود تابع توزیع (ممبینی و همکاران، ۲۰۲۱)، در این مقاله با استفاده از تابع هسته برنام- سندرز یک برآوردگر جدید ناپارامتری برای منحنی ROC معرفی کرده‌ایم. ما خواص مجانبی برآوردگر پیشنهادی را به دست آورده و همگرایی آن در احتمال را ثابت خواهیم کرد. علاوه بر آن بفرم تحلیلی نشان می‌دهیم که واریانس مجانبی برآوردگر جدید کوچکتر از واریانس مجانبی برآوردگر لوید است. همچنین در یک مطالعه عددی نشان می‌دهیم که ریسک برآوردگر پیشنهادی کوچکتر از ریسک برآوردگرهای مرسوم برای منحنی ROC است. مقاله در شش بخش تنظیم شده است. در بخش ۲، خواص مجانبی برآوردگر پیشنهادی (برآوردگر B-S) به صورت مختصر مرور شده است. در بخش ۳، برآوردگر جدید معرفی شده و اریبی و واریانس مجانبی آن به دست آمده و با اریبی و واریانس برآوردگر هسته متقارن لوید مقایسه شده است. بخش ۴ مقاله به مطالعه عددی اختصاص یافته و با استفاده از شبیه‌سازی عملکرد برآوردگر پیشنهادی نمایش داده شده و با دیگر برآوردگرها مقایسه شده است. در بخش ۵ مزیت برآوردگر جدید در تحلیل یک مجموعه داده واقعی نشان داده شده است. سرانجام در بخش ۶ بحث و نتیجه‌گیری ارائه شده است.

۲ خواص مجانبی برآوردگر هسته **B-S** برای تابع توزیع احتمال

در این مقاله فرض می‌کنیم که:

۱- توابع توزیع تجمعی F_i $i = 1, 2$ بر فاصله $[0, \infty)$ مطلقا پیوسته بوده و دارای مشتق مرتبه اول و دوم پیوسته و کراندار هستند.

۲- پارامترهای هموار کننده $b_i > 0, i = 1, 2$ در شرط $b_i \longrightarrow 0$ زمانیکه $n_i \longrightarrow \infty$ صدق می‌کنند.

۳- انتگرال‌های

$$\int_0^\infty (uf_i(u))^2 du \quad ، \int_0^\infty \left(u^2 f_i'(u)\right)^2 du, \quad i = 1, 2,$$

متناهی هستند.

فرض کنید که مشاهدات $X_{1,1}, ..., X_{1,n_1}$ و $X_{2,1}, ..., X_{2,n_2}$ را به ترتیب از دو گروه اول و دوم با توابع توزیع $F_1(x)$ و $F_2(x)$ در اختیار داریم. برآورد تابع توزیع احتمال $F_i(x), \quad i = 1, 2$ با استفاده از هسته $B - S$ به صورت

$$\hat{F}_i(x) = n_i^{-1} \sum_{j=1}^{n_i} \bar{K}_{B-S}\left(X_{i,j} \ ; \ x, \sqrt{b_i}\right), \quad b_i > 0, \quad i = 1, 2, \tag{۱.۲}$$

حاصل می‌شود که در آن

$$\bar{K}_{B-S}\left(X_{i,j} \ ; \ x, \sqrt{b_i}\right) = 1 - \Phi\left(\left(\sqrt{\frac{X_{i,j}}{x}} - \sqrt{\frac{x}{X_{i,j}}}\right)/\sqrt{b_i}\right), \quad i = 1, 2,$$

و $\Phi(\cdot)$ تابع توزیع نرمال استاندارد و b_i پارامتر هموار کننده است. تحت فرضیات ۳-۱، ممبینی و همکاران (۲۰۲۱) ثابت کردند که

$$\text{bias}\left(\hat{F}_i(x)\right) = \frac{b_i}{2}\left(xf_i(x) + x^2 f_i'(x)\right) + O(b_i^2), \quad i = 1, 2, \tag{۲.۲}$$

و

$$\text{var}(\hat{F}_i(x)) = n_i^{-1}\left(F_i(x)(1 - F_i(x))\right) - n_i^{-1} b_i^{\frac{1}{2}} \pi^{-\frac{1}{2}} x f_i(x) + O(n_i^{-1} b_i), \quad i = 1, 2, \tag{۳.۲}$$

ممبینی و همکاران (۲۰۲۱) با آنالیز میانگین انتگرال مربعات خطا MISE در نقاط مرزی و داخلی دامنه، نشان دادند که $\hat{F}_i(x), \ i = 1, 2$ دارای اریبی مرزی نیست و نرخ همگرایی آن در هر دو ناحیه مرزی و داخلی سریع است. می‌توان نشان داد که پارامتر هموار کننده بهینه حاصل از مینیمم کردن MISE عبارت است از

$$\hat{b}_i = \left\{\int_0^\infty xf(x)dx\right\}^{\frac{2}{3}} \left\{\sqrt{\pi}\int_0^\infty (xf(x) + x^2 f'(x))^2 dx\right\}^{-\frac{2}{3}} n^{-\frac{2}{3}}, \quad i = 1, 2. \tag{۴.۲}$$

۳ برآورد منحنی ROC با استفاده از هسته B-S

فرض کنید $\hat{R}(t) = ۱ - \hat{F}_۲\left(\hat{F}_۱^{-۱}(۱-t)\right)$ برآوردگر تابع $R(t)$ باشد. واضح است که عملکرد این برآوردگر تا حد زیادی وابسته به برآورد مناسب توابع توزیع $F_۱$ و $F_۲$ است. این امر به دلیل آن است که $\hat{R}(t)$ خواص $\hat{F}_۱$ و $\hat{F}_۲$ را به ارث می‌برد. همانگونه که در مقدمه ذکر شد برآوردگرهای نوع هسته متقارن دارای دو مشکل اریبی مرزی و نحوه تنظیم پارامتر هموار کننده هستند. اگر چه که محققین برای رفع این نواقص در هسته‌های متقارن، روش‌هایی را پیشنهاد کرده‌اند اما هسته‌های نامتقارن راه حلی ساده برای هر دو مشکل هستند. این امر به دلیل آن است که هسته‌های نامتقارن دارای شکلی منعطف بوده و دامنه آنها منطبق بر دامنه داده‌ها است. در این مقاله، برای بهره‌گیری کامل از مزیت هسته‌های نامتقارن در برآورد منحنی ROC، برآوردگر زیر را بر پایه تابع هسته B-S پیشنهاد کرده‌ایم

$$\hat{R}_{BS}(t) = ۱ - \hat{F}_۲\left(\hat{F}_۱^{-۱}(۱-t)\right), \tag{۱.۳}$$

که در آن $\hat{F}_i(x), i = ۱, ۲$ برآوردگر هسته B-S توابع توزیع $F_۱$ و $F_۲$ هستند که در معادله (۱.۲) تعریف شده‌اند. قضیه زیر نشان می‌دهد که $\hat{R}(t)$ به‌صورت مجانبی یک برآوردگر ناریب و سازگار برای $R(t)$ است.

قضیه ۱.۳. فرض کنید که فرضیات ۳-۱ بخش دوم برقرار بوده و $R''(t)$ برای $t \in (۰, ۱)$ موجود باشد. آنگاه اریبی و واریانس $\hat{R}_{BS}(t)$ عبارتند از

$$\begin{aligned}\text{bias}\left(\hat{R}_{BS}(t)\right) = \frac{R''(t)}{۲}&\left(n_۱^{-۱}t(۱-t) + b_۱q_t^۲f_۱^۲(q_t) - n_۱^{-۱}\pi^{-\frac{۱}{۲}}b_۱^{\frac{۱}{۲}}q_tf_۱(q_t)\right)\\ &+ \frac{(b_۱-b_۲)}{۲}(q_tf_۲(q_t) + q_t^۲f_۲'(q_t)) + O(b_۲) + O(n_۱^{-۱}b_۱),\end{aligned}$$

و

$$\begin{aligned}\text{var}(\hat{R}_{BS}(t)) &= n_۲^{-۱}R(t)(۱-R(t)) + n_۱^{-۱}R'^۲(t)t(۱-t)\\ &- \left\{n_۲^{-۱}\pi^{-\frac{۱}{۲}}b_۲^{\frac{۱}{۲}}q_tf_۱(q_t)R'(t) + n_۱^{-۱}R'^۲(t)\pi^{-\frac{۱}{۲}}b_۱^{\frac{۱}{۲}}q_tf_۱(q_t)\right\} + O\left(n_۲^{-۱}b_۲^{\frac{۱}{۲}}\right)\\ &+ O(n_۲^{-۱}b_۱) + O(n_۱^{-۱}n_۲^{-۱}) + O(n_۱^{-۱}b_۱) + O(b_۲),\end{aligned}$$

که در آن $\hat{q}_t = \hat{F}_۱^{-۱}(۱-t)$.

اثبات. اثباتی که در اینجا آمده است مشابه با اثبات ارائه شده توسط لوید (۱۹۹۸) است که برای هسته نامتقارن B-S وفق یافته است. فرض کنید $\hat{q}_t = \hat{F}_۱^{-۱}(۱-t)$ یک M-برآوردگر باشد که در رابطه $h(\hat{q}_t, X) = \hat{F}_۱(\hat{q}_t) - ۱ + t = ۰$ صدق می‌کند. دقت کنید که $R(t) = ۱ - F_۲(q_t)$ و $\hat{R}_{BS}(t) = ۱ - \hat{F}_۲(\hat{q}_t)$ هستند. با استفاده از لوید (۱۹۹۸) نتایج زیر برای واریانس و اریبی $\hat{q}_t$ به دست می‌آیند:

$$\text{var}(\hat{q}_t) = \frac{\text{var}(\hat{F}_۱(\hat{q}_t))}{E^۲(\hat{f}_۱(q_t))}, \tag{۲.۳}$$

و

$$-E(\hat{F}_۱(q_t) - ۱ + t) = f_۱(q_t)\ \text{bias}\ (\hat{q}_t) + \frac{۱}{۲}f_۱'(q_t)\text{MSE}(\hat{q}_t). \tag{۳.۳}$$

با جایگذاری (۲.۲) در (۳.۳) داریم

$$-\left(F_۱(q_t) + \frac{b_۱}{۲}\left(q_tf_۱(q_t) + q_t^۲f_۱'(q_t)\right) - ۱ + t + O(b^۲)\right) = f_۱(q_t)\ \text{bias}\ (\hat{q}_t) + \frac{۱}{۲}f_۱'(q_t)\text{MSE}(\hat{q}_t),$$

یا

$$\begin{aligned}\frac{b_۱}{۲f_۱(q_t)}\left(q_tf_۱(q_t) + q_t^۲f_۱'(q_t)\right) + \frac{۱}{۲}\frac{f_۱'(q_t)}{f_۱(q_t)}\text{MSE}(\hat{q}_t) &\approx \text{-bias}(\hat{q}_t),\\ \Rightarrow \frac{b_۱q_t}{۲} + \frac{f_۱'(q_t)}{۲f_۱(q_t)}\left(b_۱q_t^۲ + \text{MSE}(\hat{q}_t)\right) &\approx \text{-bias}(\hat{q}_t),\end{aligned}$$

چون $\text{MSE}(\hat{q}_t) = \text{var}(\hat{q}_t) + \text{bias}^2(\hat{q}_t) \approx \text{var}(\hat{q}_t) + O(b_1^2)$ داریم:

$$\text{bias}\,(\hat{q}_t) \approx -\left(\frac{b_1 q_t}{2} + \frac{f_1'(q_t)}{2f_1(q_t)}\left(b_1 q_t^2 + \text{var}(\hat{q}_t)\right)\right), \tag{۴.۳}$$

همچنین از جایگذاری معادله (۳.۲) در معادله (۲.۳) داریم

$$\begin{aligned}\text{var}(\hat{q}_t) &= \frac{n_1^{-1}\left(F_1(q_t)(1-F_1(q_t))\right) - n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t) + O(n_1^{-1}b_1)}{\left(f_1(q_t) + \frac{b_1}{2}\left(q_t f_1'(q_t) + q_t^2 f_1''(q_t)\right) + o(b_1)\right)^2} \\ &\approx \frac{n_1^{-1}\left(F_1(q_t)\left(1-F_1(q_t)\right)\right)}{f_1^2(q_t)} - \frac{n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t)}{f_1^2(q_t)} + O(n_1^{-1}b_1),\end{aligned} \tag{۵.۳}$$

که در آن $E\left(\hat{f}_1(q_t)\right) = f_1(q_t) + O(b_1)$ (مارچانت و همکاران (۲۰۱۳) را برای جزئیات ببینید).
چون $\hat{F}_2$ مستقل از $\hat{q}_t$ است؛ میانگین و واریانس $\hat{R}_{BS}(t) = 1 - \hat{F}_2(\hat{q}_t)$ به شرط $\hat{q}_t$ عبارتند از

$$\text{var}\left(\hat{R}_{BS}(t)|\hat{q}_t\right) = n_2^{-1}\left(F_2(\hat{q}_t)\left(1-F_2(\hat{q}_t)\right)\right) - n_2^{-1}\pi^{-\frac{1}{2}}b_2^{\frac{1}{2}}q_t f_2(q_t) + O(n_2^{-1}b_2),$$

و

$$\begin{aligned}E\left(\hat{R}_{BS}(t)|\hat{q}_t\right) &= 1 - F_2(\hat{q}_t) - \frac{b_2}{2}\left(\hat{q}_t f_2(\hat{q}_t) + \hat{q}_t^2 f_2'(\hat{q}_t)\right) + O(b_2^2). \\ E\left(\hat{R}_{BS}(t)\right) &= E\left(E\left(\hat{R}_{BS}(t)|\hat{q}_t\right)\right) = E\left(1 - F_2(\hat{q}_t) - \frac{b_2}{2}\left(\hat{q}_t f_2(\hat{q}_t) + \hat{q}_t^2 f_2'(\hat{q}_t)\right) + O(b_2^2)\right) \\ &= 1 - E\left\{F_2(\hat{q}_t) - \frac{b_2}{2}\left(\hat{q}_t f_2(\hat{q}_t) + \hat{q}_t^2 f_2'(\hat{q}_t)\right)\right\} + O(b_2^2) \\ &= 1 - \left\{F_2(q_t) + f_2(q_t)\text{bias}(\hat{q}_t) + \frac{1}{2}f_2'(q_t)\text{var}(\hat{q}_t)\right\} \\ &\quad - \frac{b_2}{2}\left(q_t f_2(q_t) + q_t^2 f_2'(q_t)\right) + O(b_2), \\ \Rightarrow E\left(\hat{R}_{BS}(t)\right) &= R(t) + f_2(q_t)\left(\frac{b_1 q_t}{2} + \frac{f_1'(q_t)}{2f_1(q_t)}\left(b_1 q_t^2 + \text{var}(\hat{q}_t)\right)\right) \\ &\quad - \frac{1}{2}f_2'(q_t)\text{var}(\hat{q}_t) - \frac{b_2}{2}\left(q_t f_2(q_t) + q_t^2 f_2'(q_t)\right) + O(b_2) \\ &= R(t) + \left(f_2(q_t)f_1'(q_t) - f_2'(q_t)f_1(q_t)\right)\frac{\text{var}(\hat{q}_t)}{2f_1(q_t)} + \frac{b_1 q_t f_2(q_t)}{2} + \frac{b_1 q_t^2 f_2(q_t)f_1'(q_t)}{2f_1(q_t)} \\ &\quad - \frac{b_2}{2}\left(q_t f_2(q_t) + q_t^2 f_2'(q_t)\right) + O(b_2) \\ &= R(t) + R''(t)\frac{f_1^2(q_t)\text{var}(\hat{q}_t)}{2} + \left(\frac{b_1 q_t f_2(q_t)}{2} - \frac{b_2 q_t f_2(q_t)}{2}\right) \\ &\quad + \left(\frac{b_1 q_t^2 f_2(q_t)f_1'(q_t)}{2f_1(q_t)} - \frac{b_2 q_t^2 f_2'(q_t)}{2}\right) + O(b_2), \\ \Rightarrow E\left(\hat{R}_t\right) &= R(t) + R''(t)\frac{f_1^2(q_t)\text{var}(\hat{q}_t)}{2} + \frac{q_t f_2(q_t)}{2}(b_1 - b_2) \\ &\quad + \frac{b_1 q_t^2}{2f_1(q_t)}(f_2(q_t)f_1'(q_t) - f_2'(q_t)f_1(q_t)) + \frac{q_t^2 f_2'(q_t)}{2}(b_1 - b_2) + O(b_2)\end{aligned}$$

$$= R(t) + R''(t)\frac{f_1^2(q_t)\text{var}(\hat{q}_t)}{2} + \frac{b_1 - b_2}{2}(q_t f_2(q_t) + q_t^2 f_2'(q_t)) + b_1 q_t^2 f_1^2(q_t)R''(t) + O(b_2)$$
$$= R(t) + \frac{R''(t)}{2}\left(n_1^{-1}t(1-t) - n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t) + b_1 q_t^2 f_1^2(q_t)\right)$$
$$+ \frac{(b_1 - b_2)}{2}(q_t f_2(q_t) + q_t^2 f_2'(q_t)) + O(b_2) + O(n_1^{-1}b_1).$$

حال واریانس $\hat{R}_{BS}(t)$ را محاسبه می‌کنیم.

$$n_2 E\left(\text{var}\left(\hat{R}_{BS}(t)|\hat{q}_t\right)\right) = E\left(F_2(\hat{q}_t)(1 - F_2(\hat{q}_t)) - \pi^{-\frac{1}{2}}b_2^{\frac{1}{2}}\hat{q}_t f_2(\hat{q}_t) + O(b_2)\right)$$
$$= F_2(q_t)(1 - F_2(q_t)) + f_2(q_t)(1 - 2F_2(q_t))\text{bias}(\hat{q}_t) - \pi^{-\frac{1}{2}}b_2^{\frac{1}{2}}q_t f_2(q_t) + O\left(b_2^{\frac{1}{2}}\right)$$
$$= R(t)(1 - R(t)) - \pi^{-\frac{1}{2}}b_2^{\frac{1}{2}}q_t f_2(q_t) + O\left(b_2^{\frac{1}{2}}\right) + O(b_1) + O(n_1^{-1}).$$

با استفاده از بسط تیلور $f_2(\hat{q}_t)$ حول q_t و این نکته که جملات شامل کواریانس متناهی هستند، داریم:

$$\text{var}\left(E(\hat{R}_{BS}(t)|\hat{q}_t)\right) = \text{var}\left(1 - F_2(\hat{q}_t) - \frac{b_2}{2}\left(\hat{q}_t f_2(\hat{q}_t) + \hat{q}_t^2 f_2'(\hat{q}_t)\right) + O(b_2^2)\right)$$
$$= (f_2(q_t) + O(b_2))^2\,\text{var}(\hat{q}_t) + O(b_2) = f_2^2(q_t)\text{var}(\hat{q}_t) + O(b_2)$$
$$= f_2^2(q_t)\left(\frac{n_1^{-1}\left(F_1(q_t)(1 - F_1(q_t))\right)}{f_1^2(q_t)} - \frac{n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t}{f_1(q_t)} + O(n_1^{-1}b_1)\right) + O(b_2)$$
$$= R'^2(t)\left\{n_1^{-1}t(1-t) - n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t)\right\} + O\left(n_1^{-1}b_1^{\frac{1}{2}}\right) + O(b_2).$$
$$\Rightarrow \text{var}\left(\hat{R}_{BS}(t)\right) = E\left(\text{var}\left(\hat{R}_{BS}(t)|\hat{q}_t\right)\right) + \text{var}\left(E\left(\hat{R}_{BS}(t)|\hat{q}_t\right)\right)$$
$$= n_2^{-1}R(t)(1 - R(t)) - n_2^{-1}\pi^{-\frac{1}{2}}b_2^{\frac{1}{2}}q_t f_1(q_t)R'(t)$$
$$+ n_1^{-1}R'^2(t)\left\{t(1-t) - \pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t)\right\} + O\left(n_2^{-1}b_2^{-\frac{1}{2}}\right) + O(n_2^{-1}b_1)$$
$$+ O\left(n_1^{-1}n_2^{-1}\right) + O\left(n_1^{-1}b_1\right) + O(b_2),$$

که در آن $R'(t) = \dfrac{f_2(q_t)}{f_1(q_t)}$ و $R''(t) = \dfrac{f_2(q_t)f_1'(q_t) - f_2'(q_t)f_1(q_t)}{f_1^3(q_t)}$. □

نتیجه ۲.۳. $\hat{R}_{BS}(t)$ به طور مجانبی نا اریب و سازگار است.

اثبات از قضیه ۱.۳ و این فرض که $b_1, b_2 \longrightarrow 0$ زمانیکه $n_1, n_2 \longrightarrow \infty$ نتیجه می‌شود.
مقایسه اریبی و واریانس برآودگر B-S و برآوردگر لوید $(\hat{R}_{Lloyd}(t))$ می‌تواند جالب باشد زیرا $\hat{R}_{BS}(t)$ و $\hat{R}_{Lloyd}(t)$ را می‌توان به ترتیب نماینده برآوردگرهای هسته نامتقارن و متقارن دانست. اریبی $\hat{R}_{BS}(t)$ برابر

$$\text{bias}\left(\hat{R}_{BS}(t)\right) \approx \frac{R''(t)}{2}\left(n_1^{-1}t(1-t) + b_1 q_t^2 f_1^2(q_t) - n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t)\right)$$
$$+ \frac{(b_1 - b_2)}{2}\left(q_t f_2(q_t) + q_t^2 f_2'(q_t)\right),$$

است و به ازای $b_1 = b_2$ عبارت فوق به

$$\text{bias}\left(\hat{R}_{BS}(t)\right) \approx \frac{R''(t)}{2}\left(n_1^{-1}t(1-t) + b_1 q_t^2 f_1^2(q_t) - n_1^{-1}\pi^{-\frac{1}{2}}b_1^{\frac{1}{2}}q_t f_1(q_t)\right), \qquad (۶.۳)$$

ساده می‌شود. اریبی $\hat{R}_{Lloyd}(t)$ به ازای مقادیر یکسان پارامتر هموار کننده برای $\hat{F}_1$ و $\hat{F}_2$ برابر

$$\text{bias}\left(\hat{R}_{Lloyd}(t)\right) \approx \frac{R''(t)}{2}\left(n_1^{-1}t(1-t) + b_1 q_t^2 f_1^2(q_t)\right), \qquad (۷.۳)$$

است (لوید، ۱۹۹۸).
دقت کنید که اریبی هر برآوردگر تابعی از پارامتر هموار کننده آن است. بنابراین برای مقایسه اریبی دو برآوردگر باید به این مسأله توجه کرد و از پارامتر هموار کننده یکسان در روابط (۶.۳) و (۷.۳) استفاده کرد. فرض کنید که h و b پارامترهای هموار کننده به ترتیب در توابع هسته متقارن و نامتقارن باشند. دقت کنید که پارامتر هموار کننده تابعی از حجم نمونه است. بر حسب نرخ همگرایی به صفر به راحتی می‌توان نشان داد که $h^2 = b$ بنابراین در جمله اریبی برآوردگر لوید، b_1 جایگزین h_1^2 می‌شود.
لوید (۱۹۹۸) جمله اریبی $\hat{R}_{Lloyd}(t)$ را مورد بحث قرار داد و نتیجه گرفت که اگر $R(t)$ محدب باشد یعنی $R''(t) < 0$ آنگاه جمله درون پرانتز در رابطه (۷.۳) همیشه منفی است بجز جاییکه $R(t) = t$ باشد. اما این مطلب در مورد برآوردگر $\hat{R}_{BS}(t)$ برقرار نیست. بنابراین بر خلاف برآوردگر لوید، برآوردگر پیشنهادی دارای اریبی سیستماتیک منفی نیست. علاوه بر آن به سادگی می‌توان نشان داد که به ازای $b_1 = b_2$ داریم

$$\text{bias}\left(\hat{R}_{BS}(t)\right) \approx \text{bias}\left(\hat{R}_{Lloyd}(t)\right) - \frac{R''(t)}{2} n_1^{-1}\pi^{-\frac{1}{2}} b_1^{\frac{1}{2}} q_t f_1(q_t).$$

با توجه به اینکه اریبی $\hat{R}_{Lloyd}(t)$ منفی است، اگر $R(t)$ محدب باشد $(R''(t) < 0)$ آنگاه اریبی برآوردگر پیشنهادی کمتر از اریبی برآوردگر لوید است بجز حالتی که قدر مطلق عبارت $\frac{R''(t)}{2} n_1^{-1}\pi^{-\frac{1}{2}} b_1^{\frac{1}{2}} q_t f_1(q_t)$ بیشتر از دو برابر قدرمطلق اریبی از برآوردگر لوید باشد یا اینکه $R(t) = t$ باشد.
با مقایسه واریانس دو برآوردگر $\hat{R}_{BS}(t)$ و $\hat{R}_{Lloyd}(t)$ (لوید، ۱۹۹۸) داریم:

$$\text{var}\left(\hat{R}_{BS}(t)\right) \approx \text{var}\left(\hat{R}_{Lloyd}(t)\right) - \pi^{-\frac{1}{2}} q_t f_1(q_t)\left\{n_2^{-1} b_2^{\frac{1}{2}} R'(t) + n_1^{-1} R'^2(t) b_1^{\frac{1}{2}}\right\}.$$

از آنجا که $R(t)$ یک تابع غیر نزولی است $(R'(t) \geq 0)$ واریانس برآوردگر پیشنهادی به صورت یکنواخت کوچکتر از واریانس برآوردگر لوید است. این نتایج نظری توسط مطالعات عددی نیز تایید می‌شوند. در بخش بعد در یک مطالعه عددی با در نظر گرفتن سناریوهای مختلف و متنوع نشان می‌دهیم که MISE برآوردگر پیشنهادی به طور قابل ملاحظه‌ای کوچکتر از MISE دیگر برآوردگرهای منحنی ROC از جمله برآوردگر لوید است.

۴ مطالعه عددی

در این بخش عملکرد برآوردگر پیشنهادی (B-S) از طریق یک مطالعه شبیه‌سازی نمایش داده شده و با عملکرد برآوردگرهای لوید (لوید، ۱۹۹۸)، پولیت (پولیت، ۲۰۱۶) و K-K (کولاچک و کارنامونی، ۲۰۰۹) مقایسه شده است. بجز برآوردگر پیشنهادی در دیگر برآوردگرها از تابع هسته اپانچنیکو استفاده شده است. در برآوردگر استفاده شده و پارامترهای آن با استفاده از روش حداکثر درست‌نمایی از داده‌ها برآورد شده است. برای انتخاب پارامتر هموار کننده در دو برآوردگر لوید و پولیت از روش پیشنهادی آلتمن و لگر (۱۹۹۵) استفاده شده است. سرانجام در برآوردگر K-K، پارامتر هموار کننده با روش پیشنهادی هوروا و همکاران (۲۰۰۸) انتخاب شده است. در هر برآوردگر سعی شده است که پارامتر هموار کننده به روش بهینه انتخاب شود تا بهترین عملکرد آن حاصل شود.
در این مطالعه پنج حالت شامل ترکیب‌های مختلف توابع توزیع $(F_1(x), F_2(x))$ را در نظر گرفته‌ایم:

حالت ۱: $(Beta(1,1), Beta(3,1))$
حالت ۲: $(Gamma(1,2), Gamma(3,2))$
حالت ۳: $(Halfnormal(0,1), Halfnormal(1,1))$
حالت ۴: $(lognormal(0,1), lognormal(1,1))$
حالت ۵: $(Normal(4,1), Normal(5,1))$

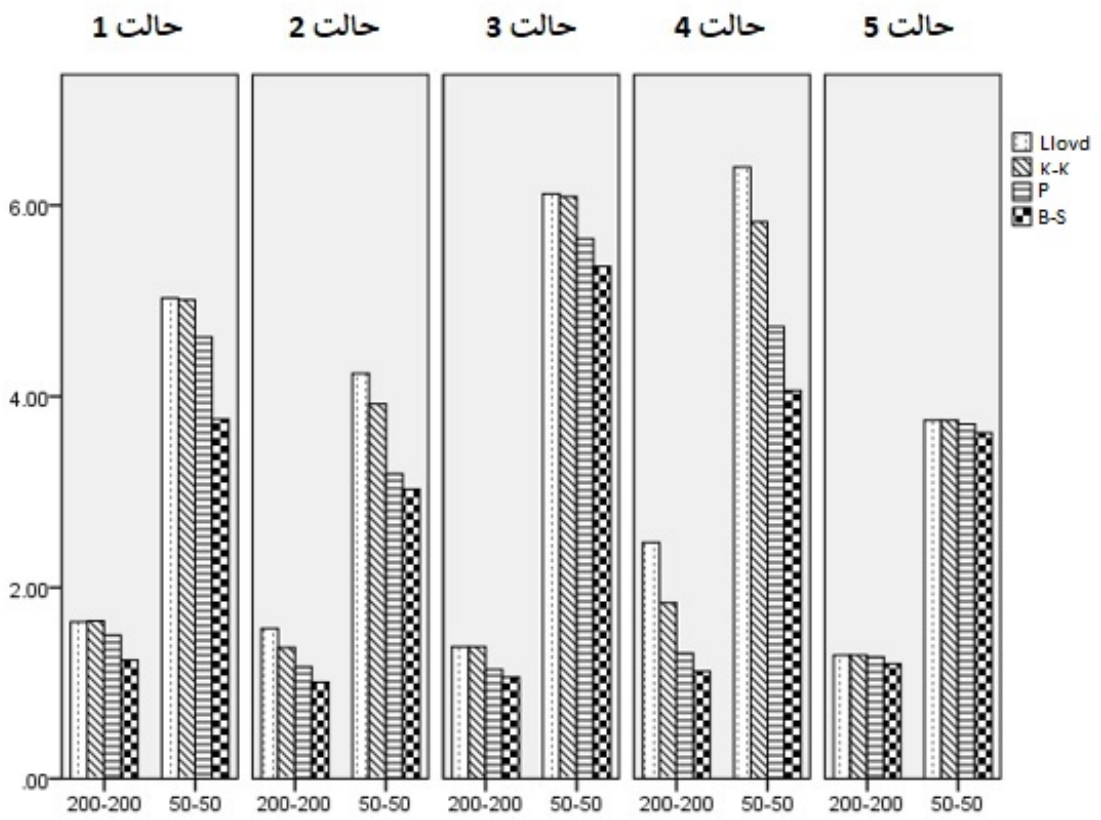
دقت کنید که حالت ۵ از دیگر حالت‌ها متفاوت است از این نظر که دو تابع توزیع با دامنه $(-\infty, \infty)$ را شامل می‌شود. این حالت را برای مشاهده عملکرد برآوردگرها زمانیکه توزیع‌های زمینه‌ای دارای دامنه غیر محدود هستند، در نظر گرفتیم. در هر مورد ۱۰۰ تکرار از دو اندازه نمونه متفاوت $(n,m) = (50,50)$ و $(n,m) = (200,200)$ از دو توزیع تولید کرده‌ایم. در تمام حالت‌ها، پارامترهای

مجهول به روش حداکثر درست‌نمایی برآورد شده‌اند. به منظور ارزیابی عملکرد برآوردگرها و مقایسه کارایی آنها از معیار انتگرال مربعات خطا $ISE_i = \int_0^1 \left(\hat{R}_i(t) - R(t)\right)^2 dt$ استفاده شده است که در آن $\hat{R}_i(x)$ به ازای $i = 1, 2, 3, 4$ به ترتیب برآورد منحنی ROC را با برآوردگر B-S، لوید، Pulit و K-K نشان می‌دهد. در عمل ما انتگرال را با مجموع تقریب می‌زنیم.

جدول (۱) میانگین و انحراف معیار ISE را در ۱۰۰ تکرار برای حالت‌های ۱ تا ۵ و به ازای دو اندازه نمونه مختلف نشان می‌دهد. با افزایش اندازه نمونه، میانگین ISE در ۱۰۰ تکرار (MISE) تمام برآوردگرها به‌صورت قابل توجهی کاهش یافته است. نتایج دلالت بر این دارد که در تمام حالات و اندازه نمونه‌های مختلف، MISE برآوردگر پیشنهادی کمتر از سه برآوردگر دیگر است. شکل (۱) MISE را برای چهاربرآوردگر به تصویر کشیده است. همانگونه که می‌توان دید در تمام ترکیب‌های مختلف توابع توزیع و به اندازه نمونه‌های مختلف همواره MISE برآوردگر پیشنهادی کوچکتر از MISE سه برآوردگر دیگر است.

جدول ۱: میانگین و انحراف معیار ISE چهار برآوردگر در برآورد منحنی ROC به ازای دو اندازه نمونه مختلف. برای جزئیات متن را ببینید.

اندازه نمونه	برآوردگر	$\times 10^{-3}$	حالت ۱	حالت ۲	حالت ۳	حالت ۴	حالت ۵
	لوید	Mean	۵٫۰۳	۴٫۲۴	۶٫۱۲	۶٫۴۰	۳٫۷۵
		std.	۴٫۹۵	۳٫۴۸	۶٫۴۲	۵٫۶۷	۴٫۳۱
n= ۵۰	K-K	Mean	۵٫۰۱	۳٫۳۰	۶٫۰۹	۵٫۸۳	۳٫۷۵
m= ۵۰		Std.	۴٫۹۷	۳٫۰۵	۶٫۴۰	۵٫۳۰	۴٫۳۱
	پولیت	Mean	۵٫۶۲	۴٫۴۹	۵٫۶۵	۴٫۷۳	۴٫۲۲
		Std.	۵٫۷۰	۵٫۲۰	۵٫۶۶	۶٫۵۸	۴٫۲۸
	B-S	Mean	۳٫۷۶	۳٫۰۳	۵٫۳۶	۴٫۰۶	۳٫۶۲
		std.	۴٫۳۹	۳٫۴۷	۵٫۹۳	۶٫۲۰	۴٫۳۰
	لوید	Mean	۱٫۶۴	۱٫۵۷	۱٫۳۸	۲٫۴۷	۱٫۲۹
		Std.	۱٫۵۸	۱٫۱۳	۱٫۳۹	۱٫۲۵	۱٫۵۵
n= ۲۰۰	K-K	Mean	۱٫۶۳	۱٫۳۷	۱٫۳۸	۱٫۸۴	۱٫۲۹
m= ۲۰۰		Std.	۱٫۵۵	۱٫۰۴	۱٫۳۸	۱٫۶۶	۱٫۵۵
	پولیت	Mean	۱٫۵۰	۱٫۱۷	۱٫۱۴	۱٫۳۱	۱٫۴۲
		Std.	۱٫۷۱	۱٫۰۳	۱٫۳۰	۱٫۱۱	۱٫۴۳
	B-S	Mean	۱٫۲۴	۱٫۰۱	۱٫۰۶	۱٫۱۲	۱٫۲۰
		Std.	۱٫۴۴	۱٫۰۳	۱٫۲۷	۱٫۰۵	۱٫۴۳

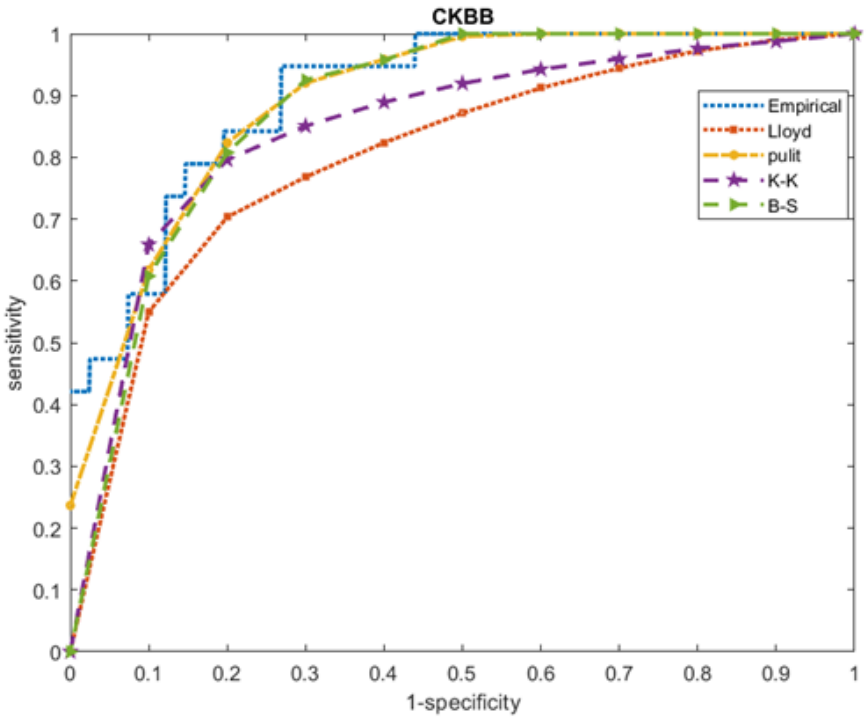


شکل ۱: نمودار MISE در برآورد منحنی ROC با چهار برآوردگر: (□) برآوردگر لوید، (▨) برآوردگر K-K، (▤) برآوردگر پولیت و (■) برآوردگر B-S برای دو اندازه نمونه مختلف.

۵ آنالیز داده واقعی

در این بخش از یک مجموعه داده واقعی برگرفته از ژو و همکاران (۲۰۰۹) صفحه ۱۴۲ برای نمایش عملکرد برآوردگر پیشنهادی استفاده شده است. این مجموعه داده حاصل یک تحقیق پزشکی است که توسط هانس و همکاران (۱۹۸۵) برای مطالعه موثر بودن اندازه گیری مایع مغزی نخاعی CKBB ایزو آنزیم (شاخص CKBB) طی ۲۴ ساعت پس از جراحت به عنوان وسیله‌ای برای پیش بینی نتیجه ضربه شدید مغزی، انجام شده است. داده‌ها شامل ۶۰ بیمار هستند که دچار ضربه شدید مغزی شده و از میان آنها ۱۹ بیمار در نهایت نتیجه خوب (بهبودی نسبی یا کامل) و ۴۱ بیمار که نتیجه بد (عدم بهبود یا بهبودی ضعیف) داشته‌اند. ژو و همکاران (۲۰۰۹) برای ارزیابی توانایی شاخص CKBB در پیش‌بینی نتیجه ضربه شدید مغزی در بیماران، منحنی ROC را برای این داده‌ها برآورد کرده‌اند.

شکل (۲) برآورد منحنی ROC را برای شاخص CKBB (به عنوان فاکتور پیش بینی کننده) با استفاده از ۵ روش نشان می‌دهد. در شکل مشخص است که برآوردگر پولیت دارای اریبی مرزی است. با مقایسه با برآورد تجربی مشخص است که دو برآوردگر لوید و K-K منحنی ROC را در نقاط انتهایی کم برآورد می‌کنند. این مطلب با نتایج تئوری لوید (۱۹۹۸) مبنی بر اریبی منفی برآوردگر لوید که در بخش قبل به آن اشاره شد، تطابق دارد. برآوردگر B-S درای مشکل اریب مرزی و کم برآوردی نیست و برخلاف برآوردگر تجربی، منحنی را به صورت هموار برآورد می‌کند.



شکل ۲: برآورد منحنی ROC برای شاخص CKBB با استفاده از پنج براوردگر: برآوردگر تجربی (▪▪▪)، برآوردگر لوید (▪■▪)، برآوردگر K-K (▪★▪)، برآوردگر پولیت (▪●▪)، و برآوردگر پیشنهادی (▪►▪)

۶ بحث و نتیجه‌گیری

در این مقاله، یک رویکرد جدید برای برآورد منحنی ROC معرفی شد. روش پیشنهادی بر پایه برآورد توابع توزیع زمینه‌ای F_1 و F_2 با استفاده از هسته نامتقارن برنام- سندرز است. آنالیز اریبی و واریانس مجانبی برآوردگر پیشنهادی دلالت بر سازگاری آن داشت. همچنین مشخص شد که برخلاف برآوردگر لوید برآوردگر پیشنهادی دارای اریبی سیستماتیک منفی نیست و در مقایسه با برآودگر لوید بطور یکنواخت واریانس کوچکتری دارد. نتایج شبیه‌سازی‌های انجام شده که دامنه متنوعی از توزیع‌های مختلف را شامل می‌شد، نشان داد که برآوردگر پیشنهادی در مقایسه با دیگر برآوردگرهای مرسوم دارای ریسک کوچکتری است. بر این اساس می‌توان در برآورد منحنی ROC خصوصا زمانیکه داده‌ها مثبت هستند، برآوردگر پیشنهادی را جایگزین برآوردگرهای ناپارامتری موجود کرد.

فهرست منابع

[1] Altman, N, Leger, C (1995) *Bandwidth selection for kernel distribution function estimation*. J Stat Plan Inference, **46**, 195-214.

[2] Chen, SX (1999) *Beta kernel estimators for density functions*. Comput Stat Data Anal, **31**, 131-145.

[3] Chen, SX (2000) *Probability density function estimation using gamma kernels*. Ann Inst Stat Math, **52**, 471-480.

[4] Du P, Tang L (2009) *Transformation-invariant and nonparametric monotone smooth estimation of ROC curves*. Stat Med, **28**, 349-359.

[5] Duong, T. (2016). *Non-parametric smoothed estimation of multivariate cumulative distribution and survival functions, and receiver operating characteristic curves*. J Korean Stat Soc, 45 (**1**), 33-50.

[6] Fawcett T (2006) *An introduction to ROC analysis*. Pattern Recognition Lett, **27**, 861-874. Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley New York.

[7] Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. Wiley New York.

[8] Hans P, Albert A, Born J, Chapelle JP (1985) *Derivation of a bioclinical prognostic index in severe head injury*. Intensive Care Med, **11**, 186-191.

[9] Horová I, Koláček J, Zelinka J, El-Shaarawi AH (2008) *Smooth estimates of distribution functions with application in environmental studies*. Article in Proceedings. Advanced topics on mathematical biology and ecology, **1**, 122-127.

[10] Hsieh F, Turnbull BW (1996) *Nonparametric and semiparametric estimation of the receiver operating characteristic curve*. Ann. Stat., **24**, 25-40.

[11] Koláček J, Karunamuni RJ (2009) *On boundary correction in kernel estimation of ROC curves*. Austrian J. Stat., **38**, 17–32-17–32.

[12] Lafaye de Micheaux, P., & Ouimet, F. (2021). *A study of seven asymmetric kernels for the estimation of cumulative distribution functions*. Mathematics, 9 (**20**), 2605.

[13] Lasko TA. Bhagwat JG, Zou KH, Ohno-Machado L (2005) *The use of receiver operating characteristic curves in biomedical informatics*. J. Biomed. Inform., **38**, 404-415.

[14] Lloyd CJ (1998) *Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems*. J Am Stat Assoc, **93**, 1356-1364.

[15] Lloyd CJ, Yong Z (1999) *Kernel estimators of the ROC curve are better than empirical*. Stat Probab Lett, **44**, 221-228.

[16] Marchant C, Bertin K, Leiva V, Saulo H (2013) *Generalized Birnbaum–Saunders kernel density estimators and an analysis of financial data*. Comput Stat Data Anal, **63**, 1-15.

[17] Mansouri, B., Atiyah Sayyid Al-Farttosi, S., Mombeni, H., & Chinipardaz, R. (2022). *Estimating Cumulative Distribution Function Using Gamma Kernel*. J. Sci. Islam. Repub. 33 (**1**), 45-54.

[18] Mombeni HA, Mansouri B, Akhoond MR (2021) *Asymmetric kernels for boundary modification in dis-tribution function estimation*. Revstat Stat. J.

[19] Pulit M (2016) *A new method of kernel-smoothing estimation of the ROC curve*. Metrika, **79**, 603-634.

[20] Silverman BW (1986) *Density estimation for statistics and data analysis*. Chapman & Hall: London.

[21] Tang L, Du P, Wu C (2010) *Compare diagnostic tests using transformation-invariant smoothed ROC curves*. J Stat Plan Inference, **140**, 3540-3551.

[22] Tenreiro C (2013) *Boundary kernels for distribution function estimation*. Revstat Stat. J., **11**, 169-190.

[23] Tenreiro C (2018) *A new class of boundary kernels for distribution function estimation*. Commun. Stat. Theory Methods, **47**, 5319-5332.

[24] Wasserman L (2006) *All of Nonparametric Statistics*, Springer: New York.

[25] Zhang S, Karunamuni RJ, Jones MC (1999) *An improved estimator of the density function at the boundary*. J Am Stat Assoc, **94**, 1231-1240.

[26] Zhou XH,. McClish DK, Obuchowski NA (2009) *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons.

[27] Zou KH. Hall W, Shapiro DE (1997) *Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests*. Stat Med, **16**, 2143-2156.

Estimating Receiver Operating Characteristic Curve Using Birnbaum-Saunders Kernel

Habiballah Mombeni[1], **Behzad Mansouri**[2,¶], **Mohammad Reza Akhoond**[3]

[1,2,3] Department of Statistics, Faculty of Mathematical Sciences & Computer, Shahid Chamran University of Ahvaz, Ahvaz, Iran

Communicated by: G. R. Mohtashami Borzadaran

Received: 2022/3/11 Accepted: 2022/9/3

Abstract: Many researchers use the receiver operating characteristic curve (ROC) as a popular way of displaying, evaluating and comparing the discriminatory accuracy of diagnostic tests. The most common approach for estimating the ROC curve is using nonparametric kernel estimates in two parts, sensitivity and specificity. Kernel estimators, however, at the beginning and end points of the data domain, known as boundary points, have a slower convergence rate than other points in the domain and are not convergent to the actual value of the probability distribution. This problem is known as the boundary problem. One way to solve the boundary problem in kernel estimators is to use asymmetric kernels. This paper proposes a new kernel estimator for the ROC curve based on the asymmetric Birnbaum-Saunders (B-S) kernel and the asymptotic convergence of the proposed estimator is shown. In addition, the analytical superiority of the proposed estimator over the corresponding symmetric kernel-type estimator is shown. The performance of the proposed estimator is illustrated via a numerical study. The results show that the proposed estimator outperforms the other commonly-used methods. The application of the proposed method to a set of medical data is also presented.

Keywords: Probability distribution function, Kernel estimator, Asymmetric kernel, ROC curve.

¶Corresponding author.
E-mail addresses: b.mansouri@scu.ac.ir (B. Mansouri)