

بررسی روش اصلاح اریبی ضربی با استفاده از هسته‌های نامتقارن

لیلا رجبی^۱، بهزاد منصوری^۲

تاریخ دریافت: ۱۴۰۱/۱۱/۲۳

تاریخ پذیرش: ۱۴۰۲/۰۲/۲۲

چکیده:

برآورد چگالی به روش هسته یک روش استاندارد برای برآورد تابع چگالی احتمال است که در بسیاری موارد عملکرد مناسبی دارد. با این حال، مشخص شده است که برای توزیع‌های نامنفی، چوله و دم‌پهن که از ویژگی‌های معمول توزیع طول عمر است، عملکرد خوبی ندارد. هدف این مقاله بررسی عملکرد روش‌های اصلاح اریبی ضربی (MBC)^۳ با استفاده از برآوردهای هسته نامتقارن و مقایسه این برآوردها با دیگر روش‌های رفع مسئله مرزی است. در این مقاله، علاوه بر معرفی روش‌های MBC در ترکیب با برآوردهای هسته نامتقارن، در یک مطالعه شبیه‌سازی به مقایسه این روش با دیگر روش‌های اصلاح اریبی پرداخته شده و نشان داده شده است که این برآوردها می‌توانند نسبت به برآوردها هسته استاندارد، در برخی موارد برازش بسیار بهتری برای برآورد چگالی فراهم کند. هم‌چنین از روش‌های MBC با استفاده از برآوردهای هسته نامتقارن برای برآورد چگالی طول عمر قرنیه‌های پیوند داده‌شده به ۱۱۹ بیمار استفاده شده است.

واژه‌های کلیدی: مسئله مرزی، هسته نامتقارن، اصلاح اریبی ضربی، چگالی طول عمر.

۱ مقدمه

برآوردها به تغییر پارامتر هموارکننده بسیار حساس هستند. تنظیم پارامتر هموارکننده برای کشف ساختار داده‌ها در نواحی از دامنه که تراکم داده‌ها زیاد است و به‌طورکلی هم‌زمان ممانعت از بروز ساختار کاذب در جاهایی که تراکم داده کم است؛ کار دشواری است. از این رو برای حل این مشکل می‌توان از روش‌های هموارسازی تطبیقی مانند روش‌های پهنای باند متغیر [۳] استفاده کرد. یک روش پیشنهادی برای رفع هم‌زمان مسئله مرزی و ناهمواری در دم توزیع استفاده از برآوردهای هسته نامتقارن است که برای اصلاح داده‌های چوله و نامنفی موفق عمل می‌کند. توابع هسته نامتقارن ابتدا توسط چن^۴ [۴] به‌عنوان ابزاری برای رفع اریبی مرزی معرفی شد و پس‌از آن توسط بسیاری از آماردانان از جمله جین و کاوژاک^۵ [۱۶]، اسکیت^۶ [۲۱] به‌عنوان یک روش جایگزین مناسب برای هسته معمولی جهت رفع هم‌زمان مسئله مرزی و ناهمواری در دم توزیع معرفی شد. در این مقاله، دو رده اصلاح اریبی ضربی (MBC) که به‌عنوان ایده‌ای در برخورد با

در بسیاری از توزیع‌های طول عمر، از لحاظ تجربی دو ویژگی وجود دارد، اول وجود مرز طبیعی در مبدأ و دوم تمرکز مشاهدات نزدیک مرز (مسئله مرزی) و داده‌های پراکنده در یک دم بلند (توزیع دم‌پهن) که این دو ویژگی سبب بروز مشکلاتی در برآورد چگالی این داده‌ها به روش هسته استاندارد می‌گردد. راه‌حل ارائه‌شده برای مشکل اول (مسئله مرزی) استفاده از روش‌های اصلاح مرزی است. مسئله مرزی اولین بار توسط گسر و مولر^۴ [۵] و سیلورمن^۵ [۲۳] مورد بررسی قرار گرفت. در برآوردها هسته با تابع متقارن روش‌های زیادی برای اصلاح مسئله مرزی وجود دارند. برخی از این روش‌ها شامل روش انعکاس داده [۲۲]، هسته‌های مرزی [۱۹] و [۲۰]، تبدیل تجربی [۱۸] و روش انعکاس داده‌های تبدیل‌شده [۲۵] است. مشکل دوم از آنجا ناشی می‌شود که شکل تابع هسته در برآوردهای هسته متقارن ثابت است؛ این نوع

^۱ کارشناسی ارشد آمار دانشگاه شهید چمران اهواز. (نویسنده مسئول: Leilarajabi88@gmail.com)

^۲ دانشیار گروه آمار، دانشگاه شهید چمران اهواز b. mansouri@scu.ac.ir

^۳Multiplicative bias correction

^۴ Gasser and Müller

^۵Silverman

^۶Chen

^۷Jin and Kawczak

^۸Scaillet

مقادیری نامنفی و اغلب در دم بالا از نوع توزیع‌هایی مانند پارتو یا لگ نرمال هستند، روش‌های MBC با استفاده از برآوردگرهای هسته نامتقارن می‌تواند مورد استفاده قرار گرفته و سبب رفع ویژگی‌هایی در داده‌ها شود که در برآوردگر چگالی هسته استاندارد مشکل‌ساز بوده و در نتیجه کیفیت برآزش در نمونه متناهی بهبود یابد. هدف این مقاله، بررسی روش‌های MBC در ترکیب با برآوردگرهای هسته نامتقارن و مقایسه عملکرد آن‌ها با دیگر برآوردگرهای اصلاح اریبی است. به این منظور دو رده از روش‌های MBC با استفاده از هسته‌های نامتقارن به‌عنوان یک روش اصلاح مسئله مرزی مطرح‌شده و کارایی این دو رده از برآوردگرها در مقایسه با دیگر برآوردگرهای اصلاح اریبی بررسی می‌شود و مشخصه‌های آن‌ها نظیر MSE، MISE، پارامتر هموارکننده، مرتبه اریبی و واریانس مورد بررسی قرار می‌گیرد. در مرحله بعد با استفاده از شبیه‌سازی، کارایی این دو رده از برآوردگرها با برآوردگرهای اصلاح اریبی دیگر مقایسه می‌شود.

این مقاله در ۵ بخش تنظیم شده است. در بخش ۲، برآوردگرهای MBC با استفاده از هسته‌های نامتقارن مطرح می‌شود. در بخش ۳، یک مطالعه شبیه‌سازی برای بررسی توانایی روش پیشنهادی در این مقاله، انجام شده است. در بخش ۴ از روش پیشنهادی برای برآورد چگالی یک مجموعه داده واقعی مربوط به طول عمر قرینه‌های پیوند داده‌شده به ۱۱۹ بیمار استفاده شده است و سرانجام در بخش ۵، بخش پایانی نتیجه‌گیری ارائه می‌شود.

نامتقارن z را به‌صورت زیر در نظر می‌گیریم:

$$\hat{f}_{j,b}(x) = \frac{1}{n} \sum_{i=1}^n K_{j(x,b)}(X_i) \quad (1)$$

که j شامل هسته‌های گاما ("GAM1")، گامای اصلاح‌شده ("GAM2")، گاوسین معکوس ("IG")، گاوسین متقابل ("RIG")، لگ نرمال ("LN") و برنهام-ساندرز ("BS") است. شکل تابعی این هسته‌ها در جدول ۱ آورده شده است. همه این هسته‌ها دارای شکلی انعطاف‌پذیر و نامنفی بوده و برای داده‌های مثبت در نقاط مختلف متغیرند، بنابراین مقدار همواری آن‌ها نیز متغیر بوده و دامنه‌شان با دامنه چگالی احتمال تحت برآورد نیز قابل تطبیق است، فاقد اریب مرزی هستند و نرخ همگرایی بهینه برای میانگین انتگرال توان دو خطا

مسئله مرزی به‌منظور کاهش نرخ همگرایی در معیار عملکرد انتگرال میانگین توان دو خطا توسط ترل و اسکات^۹ [۲۴] و جونز^{۱۰} و همکاران [۱۷] پیشنهاد شدند، معرفی می‌شوند.

ایده استفاده از توابع هسته نامتقارن به همراه دو رده ناپارامتری اصلاح اریبی ضربی (MBC) و توسعه روش‌های هموارسازی با استفاده از توابع هسته نامتقارن ابتدا توسط هیروکاوا^{۱۱} [۷] و [۹] با استفاده از هسته‌های بتا و بتا اصلاح‌شده روی فاصله واحد [۰, ۱] مطرح شده است. سپس توسط ایگاراشی^{۱۲} [۱۳] برای دو برآوردگر هسته بتا اصلاح‌شده با پارامترهای هموارسازی متفاوت با استفاده از روش‌های اصلاح اریبی به‌کاربرده شد. ایگاراشی و کاکیزاوا^{۱۳} [۱۴] و [۱۵] استفاده از روش‌های اصلاح اریبی نامنفی و جمعی را برای برخی برآوردگرهای هسته نامتقارن مورد بررسی قرار داده و روش‌های اصلاح اریبی ضربی را با استفاده از هسته‌های طبقه‌بندی‌شده، بهبود بخشیدند. برای حالت نیمه پارامتری نیز روش‌های اصلاح اریبی با استفاده از هسته‌های نامتقارن توسط هیروکاوا و ساکیودو^{۱۴} [۱۲] مورد بررسی قرار گرفته است. هیروکاوا و ساکیودو [۱۱] دو رده ناپارامتری اصلاح اریبی ضربی (MBC) را با استفاده از هسته‌های نامتقارن با دامنه‌ی $[0, \infty)$ بسط داده و نشان دادند که تحت پیش‌فرض‌های مناسب، در هر دو رده از روش‌های MBC مرتبه اریبی کاهش یافته، درحالی‌که واریانس بدون تغییر باقی می‌ماند. بر این اساس نرخ همگرایی MISE در هر دو برآوردگر MBC بهبود می‌یابد. از آنجایی‌که توزیع‌های طول عمر دارای

۲ برآوردگرهای MBC با استفاده از هسته‌های نامتقارن

فرض کنید X_1, \dots, X_n یک نمونه تصادفی از توزیع تک متغیره با چگالی f روی دامنه $[0, \infty)$ باشد. استفاده از برآوردگر چگالی هسته معمولی برای برآورد چگالی در نقطه طرح $x \in [0, \infty)$ یک هسته متقارن با دامنه \mathbb{R} یا برخی فواصل متناهی مانند $[-1, 1]$ بدون استفاده از روش‌های اصلاح مرزی به‌خوبی کار نمی‌کند. فرض کنید $K_{j(x,b)}(\cdot)$ هسته‌ای نامتقارن باشد که z به نقطه طرح $x > 0$ و پارامتر هموارکننده $b > 0$ وابسته است، در این صورت برآوردگر چگالی با استفاده از هسته

⁹Terrell and Scott

¹⁰Jones

¹¹Hirukawa

¹²Igarashi

¹³Kakizawa

¹⁴Sakudo

$f(x) = \hat{f}_{j,b}(x)\{f(x)/\hat{f}_{j,b}(x)\}$ ، برآوردگر چگالی هسته نامتقارن JLN-MBC به صورت زیر تعریف می‌شود:

$$\hat{f}_{JLN,j}(x) = \hat{f}_{j,b}(x) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{K_{j(x,b)}(X_i)}{\hat{f}_{j,b}(X_i)} \right\} \quad (۳)$$

توجه داشته باشید که عبارت درون پرانتز برآوردگر ناپارامتری اصلاح اریبی $f(x)/\hat{f}_{j,b}(x)$ است. نکته قابل توجه آن است که $\hat{f}_{TS,j}(x)$ و $\hat{f}_{JLN,j}(x)$ فاقد اریبی مرزی بوده و همیشه برآورد چگالی نامنفی تولید می‌کنند [۱۱]. هیروکاوا و ساکیودو [۱۱] نشان دادند که با فرض اینکه f تا مرتبه کافی مشتق پذیر باشد، همگرایی اریبی هر برآوردگر MBC از مرتبه $O(b)$ به مرتبه $O(b^\nu)$ می‌رسد، درحالی‌که مرتبه همگرایی واریانس در مقایسه با رابطه (۱) بدون تغییر می‌ماند، به این معنی که برای x دور از مرز هنوز $O\{(nb^\dagger)^{-1}\}$ است. در نتیجه، میانگین انتگرال توان دو خطا ("MISE") هر برآوردگر MBC برای x دور از مرز به شکل $O(b^\nu + n^{-1}b^{-\dagger})$ است؛ بنابراین، در بهترین حالت، هر برآوردگر می‌تواند به نرخ همگرایی $O(n^{-\frac{4}{3}})$ در MISE دست یابد که از $O(n^{-\frac{5}{3}})$ سریع‌تر است.

۱.۲ ویژگی‌های مجانبی برآوردگرهای MBC با استفاده از هسته‌های نامتقارن

برای مطالعه همگرایی برآوردگرهای MBC، فرضیات زیر را در نظر می‌گیریم:

فرض ۱- برای نقطه طرح $X > 0$ ، $f(X) > 0$ دارای چهار مشتق پیوسته و کران دار است.

فرض ۲- پارامتر هموارکننده $b = b_0 (> 0)$ در شرط $b \rightarrow 0$ و $nb^{\nu_j + \frac{4}{3}}$ وقتی که $n \rightarrow \infty$ صدق می‌کند و این یعنی نرخ همگرایی پارامتر هموارکننده b از نرخ همگرایی $O\{n^{-1/(r_j + \frac{4}{3})}\}$ کندتر است که r_j به صورت زیر است:

$$r_j = \begin{cases} \frac{1}{2} & j = G, MG, RIG \\ 1 & j = LN, BS \\ \frac{3}{2} & j = IG \end{cases} \quad (۴)$$

این شرایط برای کنترل مرتبه بزرگی در تقریب اریبی هر برآوردگر MBC نیاز است.

بر پایه فرض‌های ۱ و ۲ می‌توان نشان داد:

$$E\{\hat{f}_{j,b}(X)\} = f(X) + a_{1,j}(X, f)b + a_{2,j}(X, f)b^\nu + o(b^\nu),$$

(MISE) آن‌ها در رده برآوردگرهای چگالی هسته نامنفی بهینه است. به علاوه، واریانس آن‌ها هم با دور شدن نقطه طرح از مرز کاهش می‌یابد.

جدول ۱: شکل تابع هسته‌های نامتقارن

| (j) هسته | شکل تابع هسته $(u \geq 0)$ |
|-----------|--|
| GAM ν | $K_{\frac{\nu}{b}+\nu}(u) = \frac{u^{\frac{\nu}{b}} e^{-\frac{u}{b}}}{b^{\frac{\nu}{b}+\nu} \Gamma[\frac{\nu}{b} + \nu]}$ $K_{\frac{\nu}{b}+\nu}(u) = \frac{u^{\rho_b(x)-1} e^{-\frac{u}{b}}}{b^{\rho_b(x)} \Gamma[\rho_b(x)]}$ |
| GAM ν | $\rho_b(x) = \begin{cases} \frac{x}{b} & x \geq \nu b \\ \frac{1}{\nu} (\frac{x}{b})^\nu + 1 & x \in [0, \nu b) \end{cases}$ |
| IG | $K_{IG(x, \frac{1}{b})}(u) = \frac{1}{\sqrt{\nu \pi b u}} \exp(-\frac{\lambda}{\nu b x} (\frac{u}{x} - \nu + \frac{x}{u}))$ |
| RIG | $K_{RIG(\frac{1}{x-b}, \frac{1}{b})}(u) = \frac{1}{\sqrt{\nu \pi b u}} \exp(-\frac{x-b}{\nu b x} (\frac{u}{x-b} - \nu + \frac{x-b}{u}))$ |
| LN | $K_{\ln x, \nu \ln(\nu+b)}(u) = \frac{1}{\sqrt{\lambda \pi \ln(\nu+b) u}} \exp[-\frac{(\ln u - \ln x)^\nu}{\lambda \ln(\nu+b)}]$ |
| BS | $K_{b^\dagger, x}(u) = (\frac{1}{xu} + \frac{x}{u}) \exp[-\frac{1}{\nu b} (\frac{u}{x} - \nu + \frac{x}{u})]$ |

در حالت کلی برای ایجاد تمایز از برآوردگرهای MBC، برآوردگر رابطه (۱) را برآوردگر بدون اصلاح اریبی ("BU") می‌گوییم. هر دو رده اصلاح اریبی نیز مرتبه اریبی را کاهش داده، درحالی‌که مرتبه بزرگی واریانس بدون تغییر می‌ماند، میانگین توان دو خطای برآوردگرهای MBC در بهترین عملکرد به نرخ همگرایی $O(n^{-\frac{4}{3}})$ در بخش مرزی، می‌رسد.

رده اول MBC از ترکیب ضربی دو برآوردگر چگالی معرفی شده توسط ترل و اسکات [۲۴] را به طور مختصر "TS"، نشان می‌دهیم. در رده TS-MBC، از یک ترکیب ضربی دو برآوردگر چگالی با هسته یکسان و پارامترهای هموارکننده متفاوت استفاده می‌شود؛ بنابراین، اگر $\hat{f}_{j,b/c}(x)$ برآوردگر چگالی با استفاده از هسته نامتقارن j و پارامتر هموارکننده $\frac{b}{c}$ باشد که $c \in (0, 1)$ ثابت از پیش تعیین شده است و به نقطه طرح x بستگی ندارد. آنگاه برآوردگر چگالی هسته نامتقارن TS-MBC به صورت زیر تعریف می‌شود:

$$\hat{f}_{TS,j}(x) = [\hat{f}_{j,b}(x)]^{\frac{1}{1-c}} [\hat{f}_{j,b/c}(x)]^{-\frac{c}{1-c}} \quad (۲)$$

از طرف دیگر، رده دوم MBC معرفی شده توسط جونز و همکاران [۱۷] را به طور مختصر با «JLN» نشان می‌دهیم. رده JLN-MBC با استفاده از تنها یک پارامتر هموارکننده b ساخته می‌شود. با توجه به رابطه

دور از مرز داریم

$$\begin{aligned} \text{Var}\{\hat{f}_{j,b}(X)\} &= (nb^{\frac{1}{\nu}})^{-1} (\sqrt{\pi} X^{r_j})^{-1} f(X) \\ &:= (nb^{\frac{1}{\nu}})^{-1} V_j(X) f(X) \end{aligned}$$

و برای x مرزی به صورت $\text{Var}\{\hat{f}_{j,b}(X)\} = O\{(nb^{r_j+\frac{1}{\nu}})^{-1}\}$ دست می‌آید [۱۱].

که $a_{\nu,j}(X, f)$ و $a_{\nu,j}(X, f)$ توابع هسته خاصی هستند که به نقطه طرح X و مشتقات f وابسته‌اند [۱۱]. با استفاده از خواص متغیرهای تصادفی متناسب با هر هسته، شکل واضح $a_{\nu,j}(X, f)$ و $a_{\nu,j}(X, f)$ در جدول ۲ آمده است. از طرف دیگر، برای تقریب واریانس برای

جدول ۲: شکل $a_{\nu,j}(x, f)$ و $a_{\nu,j}(x, f)$ برای هسته‌های نامتقارن مختلف

| هسته (j) | $a_{\nu,j}(x, f)$ | $a_{\nu,j}(x, f)$ |
|-----------------|--|---|
| GAM \setminus | $f'(x) + \frac{x}{\nu} f''(x)$ | $f''(x) + \frac{\delta}{\rho} x f'''(x) + \frac{x^\nu}{\lambda} f^{(\nu)}(x)$ |
| GAM ν | $\begin{cases} \frac{x}{\nu} f''(x) & x \geq \nu b \\ \xi_b(x) f'(x) & x \in [0, \nu b), \\ \xi_b(x) = \rho_b(x) - \frac{x}{b} = O(1) \end{cases}$ | $\begin{cases} \frac{x}{\nu} f'''(x) + \frac{x^\nu}{\lambda} f^{(\nu)}(x) & x \geq \nu b \\ \frac{1}{\nu} \left[\xi_b^\nu(x) + \xi_b(x) + \frac{x}{b} \right] f''(x) & x \in [0, \nu b) \end{cases}$ |
| IG | $\frac{x^\nu}{\nu} f''(x)$ | $\frac{x^\delta}{\nu} f'''(x) + \frac{x^\rho}{\lambda} f^{(\nu)}(x)$ |
| RIG | $\frac{x}{\nu} f''(x)$ | $\frac{1}{\nu} f''(x) + \frac{x}{\nu} f'''(x) + \frac{x^\nu}{\lambda} f^{(\nu)}(x)$ |
| LN | $\frac{x}{\nu} f'(x) + \frac{x^\nu}{\nu} f''(x)$ | $\frac{x}{\lambda} f'(x) + \frac{\nu}{\lambda} x^\nu f''(x) + \frac{\nu}{\rho} x^\nu f'''(x) + \frac{x^\rho}{\lambda} f^{(\nu)}(x)$ |
| BS | $\frac{x}{\nu} f'(x) + \frac{x^\nu}{\nu} f''(x)$ | $\frac{\nu}{\rho} x^\nu f''(x) + \frac{\nu}{\rho} x^\nu f'''(x) + \frac{x^\rho}{\lambda} f^{(\nu)}(x)$ |

واریانس برآوردگر $TS-MBC$ به صورت زیر تقریب زده می‌شود:

$$\text{Var}\{\tilde{f}_{TS,j}(x)\} = \begin{cases} n^{-1} b^{-\frac{1}{\nu}} \lambda(c) \nu_j(x) f(x) + o\{(nb^{\frac{1}{\nu}})^{-1}\} & \text{دور از مرز} \\ O\{(nb^{r_j+\frac{1}{\nu}})^{-1}\} & \text{مرزی} \end{cases} \quad (۶)$$

که در آن $\lambda(c)$

$$\lambda(c) = \frac{(1+c^{\frac{\rho}{\nu}})(1+c)^{\frac{1}{\nu}} - \sqrt{\nu} c^{\frac{\rho}{\nu}}}{(1+c)^{\frac{1}{\nu}}(1-c)^{\frac{1}{\nu}}}$$

$\nu_i(x) = (\sqrt{\pi} x^{r_j})^{-1}$ و r_j طبق رابطه (۴) تعریف می‌شود.

قضیه ۱۰۲. اگر فرضیات ۱ و ۲ برقرار باشند، برای $a_{\nu,j}(x, f)$ اریبی تقریبی برآوردگر $TS-MBC$ با استفاده از هسته j به صورت زیر به دست می‌آید [۱۱]:

$$\begin{aligned} \text{Bias}\{\tilde{f}_{TS,j}(x)\} &\sim \frac{1}{c} \left[\frac{1}{\nu} \left\{ \frac{a_{\nu,j}^\nu(x, j)}{f(x)} \right\} - a_{\nu,j}(x, f) \right] b^\nu \\ &:= \frac{1}{c} p_j(x) b^\nu \end{aligned} \quad (۵)$$

۲.۲ ویژگی‌های موضعی برآوردگرهای MBC با استفاده از هسته‌های نامتقارن

برای x دور از مرز، MSE‌های $\tilde{f}_{JLN,j}(x)$ و $\tilde{f}_{TS,j}(x)$ به صورت زیر تقریب زده می‌شوند [۱۱]:

$$MSE \tilde{f}_{TS,j}(x) = \frac{p_j^*(x)}{c^{\frac{1}{\lambda}}} b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}} \lambda(c) \nu_j(x) f(x) + o(b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}}) \quad (۹)$$

$$MSE \tilde{f}_{JLN,j}(x) = q_j^*(x) b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}} \nu_j(x) f(x) + o(b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}}) \quad (۱۰)$$

پارامترهای هموارکننده MSE بهینه به صورت زیر هستند:

$$b_{TS,j}^* = \{c^{\frac{1}{\lambda}} \lambda(c)\}^{\frac{1}{\lambda}} \left\{ \frac{\nu_j(x) f(x)}{\lambda p_j^*(x)} \right\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \quad (۱۱)$$

$$b_{JLN,j}^* = \left\{ \frac{\nu_j(x) f(x)}{\lambda q_j^*(x)} \right\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \quad (۱۲)$$

در نتیجه MSE‌های بهینه برای دو برآوردگر به صورت زیر به دست می‌آیند:

$$MSE^* \{\tilde{f}_{TS,j}(x)\} \sim \frac{9}{\lambda^{\frac{1}{\lambda}}} \gamma(c) p_j^{\frac{1}{\lambda}}(x) \{\nu_j(x) f(x)\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \quad (۱۳)$$

$$MSE^* \{\tilde{f}_{JLN,j}(x)\} \sim \frac{9}{\lambda^{\frac{1}{\lambda}}} q_j^{\frac{1}{\lambda}}(x) \{\nu_j(x) f(x)\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \quad (۱۴)$$

که در آن $\gamma(c)$ به صورت زیر است:

$$\gamma(c) = \left\{ \frac{(1+c^{\frac{1}{\lambda}})(1+c)^{\frac{1}{\lambda}} - 2\sqrt{\lambda} c^{\frac{1}{\lambda}}}{c^{\frac{1}{\lambda}}(1+c)^{\frac{1}{\lambda}}(1-c)^{\frac{1}{\lambda}}} \right\}^{\frac{1}{\lambda}}$$

مشاهده می‌شود که پارامترهای هموارکننده MSE بهینه به صورت $O(n^{-\frac{1}{\lambda}}) = O(h^{*\frac{1}{\lambda}})$ هستند که h^* پهنای باند MSE بهینه برای برآوردگرهای هسته مرتبه چهارم، یا برآوردگرهای TS-MBC یا JLN-MBC با استفاده از هسته‌های متقارن نامنفی است. در نتیجه، MSE‌های بهینه $\tilde{f}_{JLN,j}(x)$ و $\tilde{f}_{TS,j}(x)$ برای x دور از مرز، مانند برآورد MBC با استفاده از هسته مرتبه دوم، به صورت $O(n^{-\frac{1}{\lambda}})$ است که یک نرخ همگرایی سریع‌تر از $O(n^{-\frac{1}{\lambda}})$ ، نرخ همگرایی بهینه MSE متناظر برآوردگر BU در رابطه (۱) برای x دور از مرز، است. از طرف دیگر، برای x مرزی، MSE‌های $\tilde{f}_{JLN,j}(x)$ و $\tilde{f}_{TS,j}(x)$ به صورت $O\{b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}}\}$ هستند که پارامتر هموارکننده MSE بهینه به صورت $b^{\frac{1}{\lambda}} = O\{n^{-\frac{1}{\lambda}}\}$ و $b^{\frac{1}{\lambda}} = O\{n^{-\frac{1}{\lambda}}\}$ است. در نتیجه نرخ همگرایی بهینه MSE‌ها از $O\{n^{-\frac{1}{\lambda}}\}$ ، نرخ برآوردگر BU برای x مرزی، سریع‌تر است.

قضیه ۲.۲. اگر فرضیات ۱ و ۲ برقرار باشند اریبی برآوردگر JLN-MBC با استفاده از هسته z را می‌توان به صورت زیر تقریب زد [۱۱]:

$$Bias\{\tilde{f}_{JLN,j}(x)\} \sim -f(x) a_{\lambda,j}(x, h) b^{\frac{1}{\lambda}} := q_i(x) b^{\frac{1}{\lambda}}, \quad (۷)$$

که $a_{\lambda,j}(x, h)$ با جایگذاری $f = f(x)$ در $a_{\lambda,j}(x, f) = \frac{a_{\lambda,j}(x, f)}{f(x)}$ به دست می‌آید.

برای $\nu_j(x)$ و r_j ، واریانس برآوردگر JLN-MBC به صورت زیر تقریب زده می‌شود:

$$Var\{\tilde{f}_{JLN,j}(x)\} = \begin{cases} n^{-1} b^{-\frac{1}{\lambda}} \nu_j f(x) + o\{(nb^{\frac{1}{\lambda}})^{-1}\} & \text{دور از مرز} \\ O\{(nb^{r_j+\frac{1}{\lambda}})^{-1}\} & \text{مرزی} \end{cases} \quad (۸)$$

برای دیدن اثبات کامل قضیه‌های (۱۰۲) و (۲۰۲) به [۲] مراجعه شود. از آنجایی که دامنه هسته‌های نامتقارن با چگالی واقعی f مقایسه می‌شود، برآوردگرهای TS و JLN فاقد اریبی در مرز هستند. این برآوردگرها مرتبه اریبی را از $O(b^{\frac{1}{\lambda}})$ به $O(b^{\frac{1}{\lambda}})$ کاهش می‌دهند، درحالی‌که واریانسشان برای x دور از مرز $o\{(nb^{\frac{1}{\lambda}})^{-1}\}$ و برای x مرزی $O\{(nb^{r_j+\frac{1}{\lambda}})^{-1}\}$ مشاهده می‌شود مرتبه واریانس در مقایسه با برآوردگر BU (۱) بدون تغییر باقی می‌ماند.

واریانس مجانبی مرتبه اول برآوردگرهای JLN-MBC برای x دور از مرز معادل با واریانس متناظر آن در برآوردگر BU در رابطه (۱) است. از طرف دیگر، چون ضریب $\lambda(c)$ در واریانس برآوردگر TS-MBC در $c \in (0, 1)$ از ۱ به $\frac{2\sqrt{\lambda}}{\sqrt{c}}$ افزایش می‌یابد، واریانس این برآوردگرها نسبت به برآوردگر BU در رابطه (۱) برای x دور از مرز تمایل به بزرگ‌تر شدن دارند. در نهایت، واریانس مجانبی برآوردگرهای TS-MBC و JLN-MBC برای x دور از مرز متناسب با x^{-r_j} هستند، حتی بعد از این‌که MBC ساخته می‌شود. ویژگی کاهش واریانس با x مثبت معادل روش استفاده از پهنای باند بزرگ‌تر در ناحیه دم با داده‌های پراکنده است. در نتیجه، برآوردگرهای MBC مانند برآوردگر BU ویژگی واریانس مطلوب متناسب با نقطه طرح را دارا هستند.

آن چگالی 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. هسته‌های نامتقارن با داشتن ویژگی هموارسازی تطبیقی دارای شباهت‌هایی با روش‌های پهنای باند متغیر هستند اما برخلاف این روش‌ها، متکی بودن هسته نامتقارن به تنها یک پارامتر هموارکننده برای هموارسازی تطبیقی موضعی در سرتاسر داده‌ها آن‌ها را برای کار تجربی جذاب‌تر می‌کند. هسته‌های نامتقارن در رده برآوردگرهای هسته‌های نامنفی به نرخ همگرایی بهینه (در میانگین انتگرال توان دو خطا) دست می‌یابند [۶] و [۱۰]. در آخر برخلاف هسته‌های متقارن، واریانس برآوردگرهای هسته نامتقارن با دور شدن نقطه طرح از مرز تمایل به کاهش دارند [۶] و [۱۰].

یک موضوع جذاب مقایسه روش‌های انتخاب پارامتر هموارکننده و تأثیر آن بر عملکرد برآوردگر تحت مطالعه است. یک نکته دیگر خود چگالی است که می‌خواهیم آن را برآورد کنیم و نکته دیگر تأثیر حجم نمونه است. به منظور مطالعه این موارد، مقایسه عملکرد دو رده برآوردگرهای MBC و برخی نکات دیگر، مقایسه برآوردگرها را به هسته‌های GAM1 و GAM2 به دلیل کاربرد بیشتر متمرکز کرده و دو رده معرفی شده را برای برآوردگر GAM1 در نظر گرفته و به مقایسه آن‌ها با هسته استاندارد به عنوان نمونه‌ای از رده هسته‌های متقارن می‌پردازیم. به این منظور سه توزیع گاما، لگ نرمال و وایبول را با سه اندازه نمونه شبیه‌سازی کرده‌ایم. این سه توزیع در جدول (۳) آورده شده است. نمودار مقایسه برآوردگرهای هسته استاندارد، GAM1، GAM2، TS-GAM1، JLN-GAM1 هر سه توزیع با سه اندازه نمونه ۱۰۰، ۲۰۰، ۵۰۰ به دو روش سرانگشتی سیلورمن^{۱۶} (ROT) و روش ارجاع به توزیع گاما (GR) در شکل‌های (۱) تا (۳) آورده شده است. برای رسم برآوردگرهای چگالی از نرم‌افزار R استفاده کرده‌ایم.

۳.۲ ویژگی‌های کلی برآوردگرهای MBC با استفاده از هسته‌های نامتقارن

با استدلالی که در [۴] آمده است، MISE‌های برآوردگرهای MBC به صورت زیر هستند [۱۱]:

$$\begin{aligned} MISE\{\tilde{f}_{TS,j}(x)\} &= \frac{b^{\frac{1}{\lambda}}}{c^{\frac{1}{\lambda}}} \int_0^{\infty} p_j^{\frac{1}{\lambda}}(x) dx \\ &+ \frac{\lambda(c)}{nb^{\frac{1}{\lambda}}} \int_0^{\infty} \nu_j(x) f(x) dx \\ &+ o(b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}}) \end{aligned} \quad (15)$$

$$\begin{aligned} MISE\{\tilde{f}_{JLN,j}(x)\} &= b^{\frac{1}{\lambda}} \int_0^{\infty} q_j^{\frac{1}{\lambda}}(x) dx \\ &+ \frac{1}{nb^{\frac{1}{\lambda}}} \int_0^{\infty} \nu_j(x) f(x) dx \\ &+ o(b^{\frac{1}{\lambda}} + n^{-1} b^{-\frac{1}{\lambda}}) \end{aligned} \quad (16)$$

که $p_j^{\frac{1}{\lambda}}(x)$ ، $q_j^{\frac{1}{\lambda}}(x)$ و $\nu_j(x)$ انتگرال‌پذیرند. پارامترهای هموارکننده MISE بهینه به صورت زیر به دست می‌آیند:

$$b_{TS,j}^{**} = \{c^{\frac{1}{\lambda}} \lambda(c)\}^{\frac{1}{\lambda}} \left\{ \frac{\int_0^{\infty} \nu_j(x) f(x) dx}{\lambda \int_0^{\infty} p_j^{\frac{1}{\lambda}}(x) dx} \right\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \quad (17)$$

$$b_{JLN,j}^{**} = \left\{ \frac{\int_0^{\infty} \nu_j(x) f(x) dx}{\lambda \int_0^{\infty} q_j^{\frac{1}{\lambda}}(x) dx} \right\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \quad (18)$$

بنابراین، MISE بهینه برای دو برآوردگر به صورت زیر خواهد بود:

$$\begin{aligned} MSE^{**}\{\tilde{f}_{TS,j}(x)\} &\sim \frac{9}{\lambda^{\frac{1}{\lambda}}} \gamma(c) \left\{ \int_0^{\infty} p_j^{\frac{1}{\lambda}}(x) dx \right\}^{\frac{1}{\lambda}} \\ &\times \left\{ \int_0^{\infty} \nu_j(x) f(x) dx \right\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \end{aligned} \quad (19)$$

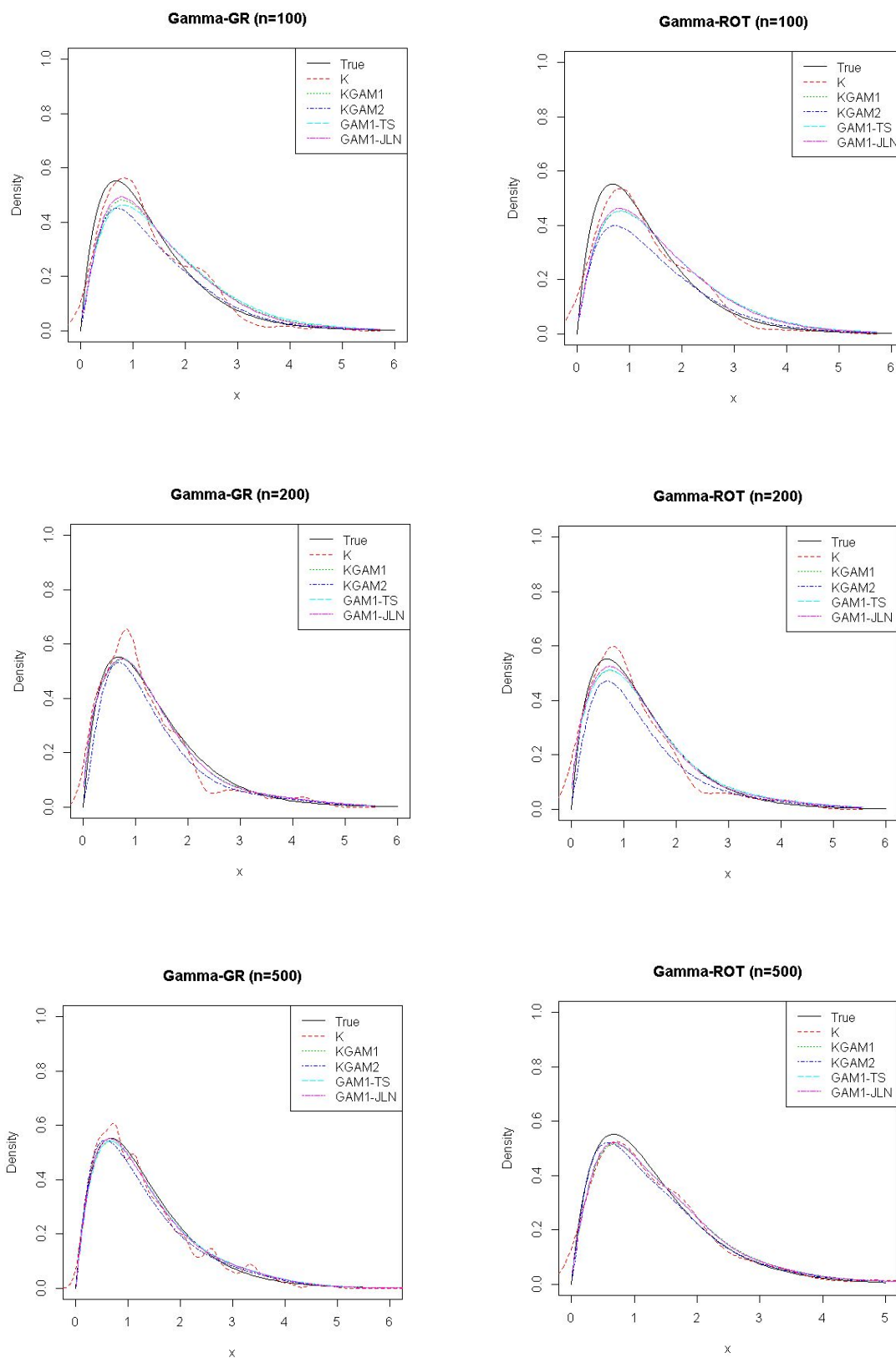
$$\begin{aligned} MSE^{**}\{\tilde{f}_{JLN,j}(x)\} &\sim \frac{9}{\lambda^{\frac{1}{\lambda}}} \left\{ \int_0^{\infty} q_j^{\frac{1}{\lambda}}(x) dx \right\}^{\frac{1}{\lambda}} \\ &\times \left\{ \int_0^{\infty} \nu_j(x) f(x) dx \right\}^{\frac{1}{\lambda}} n^{-\frac{1}{\lambda}} \end{aligned} \quad (20)$$

۳ شبیه‌سازی

هسته‌های نامتقارن مانند هسته گاما برخلاف هسته‌های متقارن دارای شکلی منعطف هستند و با توجه به موقعیت نقاط طرح^{۱۵} که در

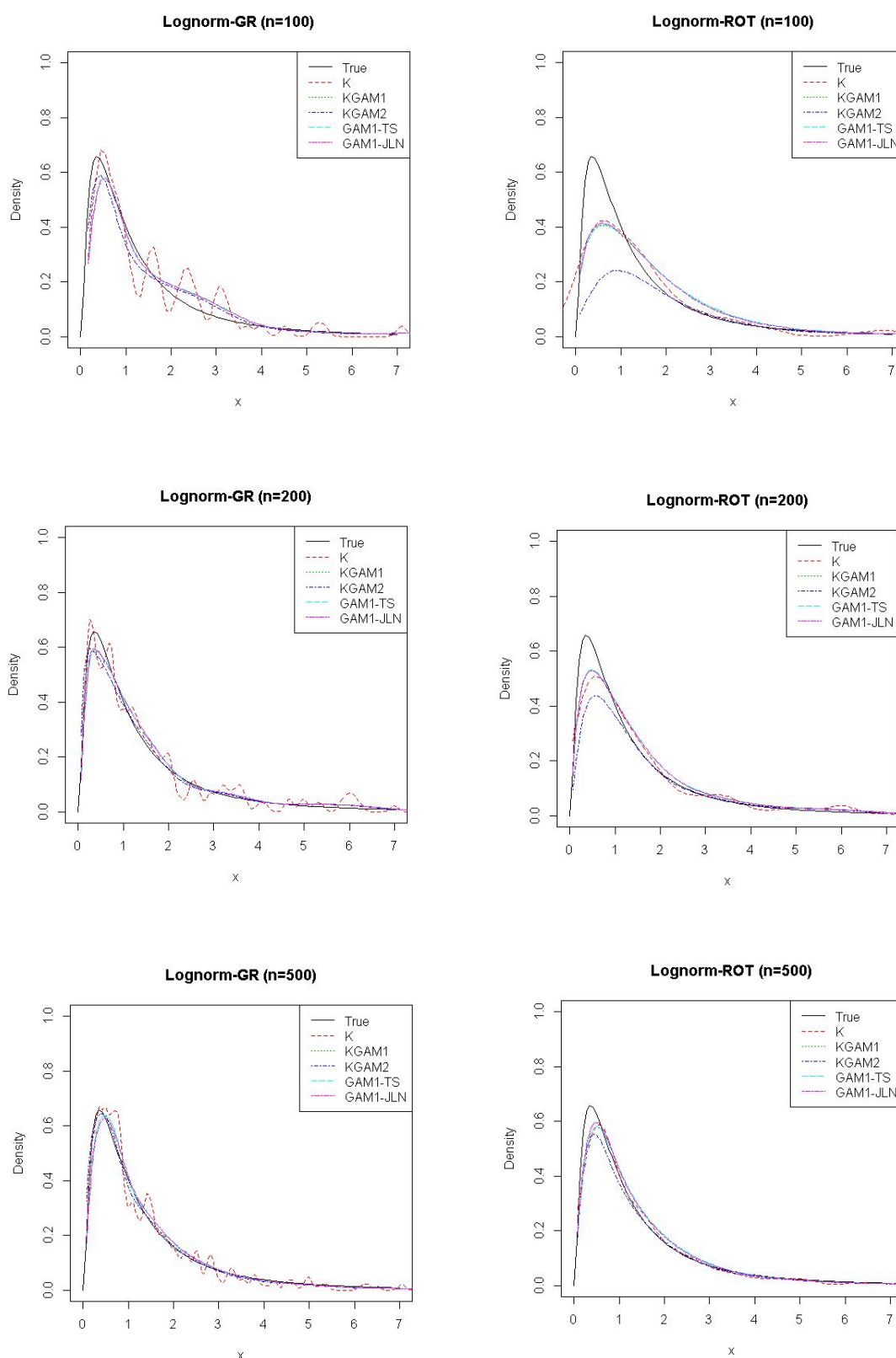
¹⁵Design points

¹⁶Silverman rule-of-thumb



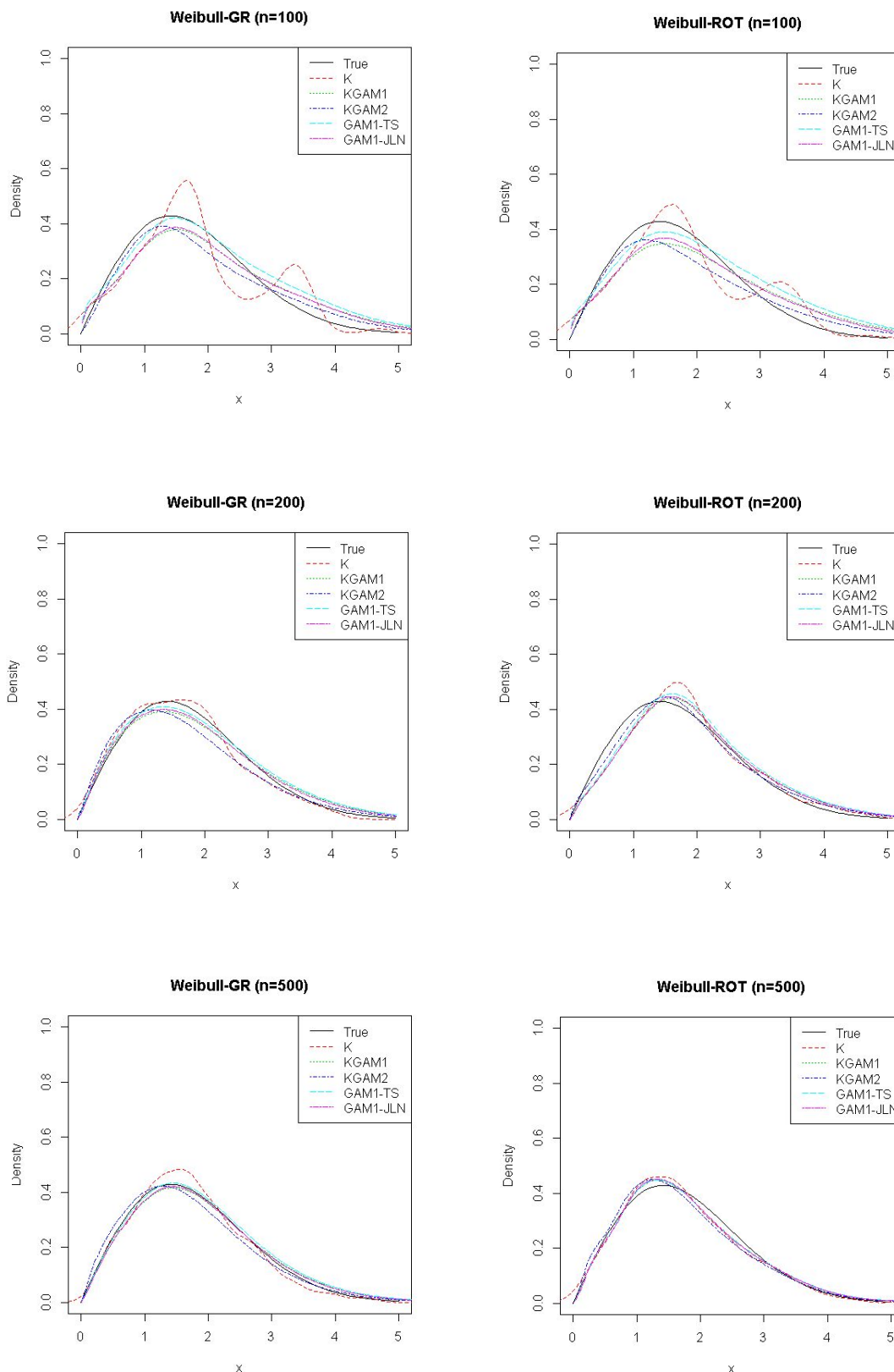
شکل ۱. شبیه‌سازی توزیع گاما $\Gamma(2, 1/5)$ با استفاده از برآوردگرهای هسته استاندارد (K) ، $(KGAM_1)$ ، $(KGAM_2)$ ، $(GAM_1 - TS)$ ، $(GAM_1 - JLN)$ و انتخاب پارامتر هموارکننده به دو روش سرانگشتی ROT و ارجاع به توزیع گاما (GR) در سه اندازه نمونه

$$n = 100, 200, 500$$



شکل ۲. شبیه‌سازی توزیع لگ نرمال $LN(0, 1)$ با استفاده از برآوردگرهای هسته استاندارد (K) ، $(KGAM_1)$ ، $(KGAM_2)$ ، $(GAM_1 - TS)$ ، $(GAM_1 - JLN)$ و انتخاب پارامتر هموارکننده به دو روش سرانگشتی ROT و ارجاع به توزیع گاما (GR) در سه اندازه نمونه

$$n = 100, 200, 500$$



شکل ۳. شبیه‌سازی توزیع وایبول $W(2, 2)$ با استفاده از برآوردگرهای هسته استاندارد (K) ، $KGAM1$ ، $KGAM2$ ، $GAM1-TS$ ، $GAM1-JLN$ و انتخاب پارامتر هموارکننده به دو روش سرانگشتی ROT و ارجاع به توزیع گاما (GR) در سه اندازه نمونه

$$n = 100, 200, 500$$

جدول ۳: توزیع‌های در نظر گرفته شده برای شبیه‌سازی به همراه چگالی آن‌ها

| توزیع | تابع چگالی $f(x)$, $x \geq 0$ |
|----------|--|
| گاما | $\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-\frac{x}{\beta})$, $(\alpha, \beta) = (2, 1.5)$ |
| لگ نرمال | $\frac{1}{x\sqrt{2\pi}\sigma} \exp\{-\frac{(\log x - \mu)^2}{2\sigma^2}\}$, $(\mu, \sigma) = (0, 1)$ |
| وایبول | $(\frac{\alpha}{\beta})(\frac{x}{\beta})^{\alpha-1} \exp\{-\frac{x}{\beta}\}$, $(\alpha, \beta) = (2, 2)$ |

جدول ۶: مقادیر MSE برآوردگرها در حالات مختلف شبیه‌سازی توزیع $(2, 2)W(\times 10^{-3})$

| ۵۰۰ | | ۲۰۰ | | ۱۰۰ | | برآوردگر/نمونه |
|-------|-------|-------|-------|-------|-------|----------------|
| ROT | GR | ROT | GR | ROT | GR | |
| ۰.۲۹۳ | ۰.۴۳۲ | ۱.۲۰۴ | ۰.۵۰۷ | ۲.۶۹۵ | ۴.۶۰۸ | هسته استاندارد |
| ۰.۱۵۰ | ۰.۱۴۷ | ۱.۰۲۳ | ۰.۴۸۱ | ۲.۶۸۲ | ۱.۸۲۳ | GAM۱ |
| ۰.۲۲۵ | ۰.۳۶۲ | ۰.۳۲۵ | ۰.۴۶۴ | ۱.۸۵۱ | ۱.۲۶۸ | GAM۲ |
| ۰.۱۴۹ | ۰.۱۳۴ | ۱.۰۳۵ | ۰.۳۷۱ | ۲.۲۹۹ | ۱.۶۴۹ | TS - GAM۱ |
| ۰.۱۶۶ | ۰.۱۰۸ | ۱.۰۲۳ | ۰.۳۳۸ | ۲.۰۷۱ | ۱.۶۸۳ | JLN - GAM۱ |

نتایج شبیه‌سازی نشان می‌دهد که با توجه به معیار MBC برای تمام توزیع‌ها، نتایج به‌طورکلی در سه اندازه نمونه مشابه بوده و با افزایش طول نمونه عملکرد برآوردگرها بهبود می‌یابد هم‌چنین در کل روش گاما-مرجع بهتر از روش سرانگشتی عمل می‌کند.

در شبیه‌سازی انجام شده برای هر ترکیب از اندازه نمونه و توزیع، روش هسته متقارن علاوه بر این‌که دارای مسئله مرزی بوده به‌هیچ‌وجه برآورد خوبی در این شبیه‌سازی ارائه نمی‌دهد و برآوردگرهای MBC به‌طورکلی دارای معیار MSE کمتری از دو برآوردگر BU هستند. این نتایج نشان می‌دهد که به‌طور یکنواخت بهترین برآوردگر MBC وجود ندارد، بلکه بهترین عملکرد برآوردگرهای MBC به توزیع‌ها بستگی دارد.

۴ داده‌های واقعی

در این بخش روش‌های اصلاح مسئله مرزی تبدیل لگاریتمی، روش انعکاسی و روش هسته مرزی و ترکیب روش‌های MBC با برآوردگرهای هسته نامتقارن، برای برآورد تابع چگالی مربوط به داده‌های طول عمر پیوند قرنیه استفاده شده است. داده‌هایی که در این مقاله مورد مطالعه

معیار MSE را به‌عنوان معیار عملکرد برای هر برآوردگر در هر سه توزیع با سه اندازه نمونه به دو روش سرانگشتی سیلورمن (ROT) و روش ارجاع به توزیع گاما (GR) محاسبه کرده و نتایج را در جدول‌های (۴) تا (۶) آورده‌ایم

جدول ۴: مقادیر MSE برآوردگرها در حالات مختلف شبیه‌سازی توزیع $(2, 1.5)\Gamma(\times 10^{-3})$

| ۵۰۰ | | ۲۰۰ | | ۱۰۰ | | برآوردگر/نمونه |
|-------|-------|-------|-------|-------|-------|----------------|
| ROT | GR | ROT | GR | ROT | GR | |
| ۰.۴۱۳ | ۰.۳۰۹ | ۱.۴۵۲ | ۰.۵۶۴ | ۱.۱۰۴ | ۱.۷۰۱ | هسته استاندارد |
| ۰.۲۲۶ | ۰.۰۶۹ | ۰.۳۲۳ | ۰.۱۶۹ | ۲.۵۶۳ | ۱.۶۸۱ | GAM۱ |
| ۰.۱۹۵ | ۰.۱۹۲ | ۲.۲۴۷ | ۱.۱۱۲ | ۴.۰۹۸ | ۲.۱۳۵ | GAM۲ |
| ۰.۱۹۰ | ۰.۰۶۷ | ۰.۵۹۹ | ۰.۱۷۶ | ۲.۵۵۵ | ۲.۲۰۸ | TS - GAM۱ |
| ۰.۱۹۲ | ۰.۰۵۳ | ۰.۲۰۰ | ۰.۱۷۷ | ۲.۲۱۵ | ۱.۴۲۰ | JLN - GAM۱ |

جدول ۵: مقادیر MSE برآوردگرها در حالات مختلف شبیه‌سازی توزیع $(0, 1)LN(\times 10^{-3})$

| ۵۰۰ | | ۲۰۰ | | ۱۰۰ | | برآوردگر/نمونه |
|-------|-------|-------|-------|-------|-------|----------------|
| ROT | GR | ROT | GR | ROT | GR | |
| ۰.۴۴۱ | ۰.۵۳۴ | ۱.۸۵۸ | ۱.۰۶۲ | ۲.۷۷۲ | ۲.۶۰۹ | هسته استاندارد |
| ۰.۴۱۶ | ۰.۱۳۳ | ۱.۲۲۵ | ۰.۳۸۷ | ۲.۶۹۵ | ۱.۳۹۳ | GAM۱ |
| ۰.۴۴۹ | ۰.۱۳۳ | ۲.۹۴۰ | ۰.۶۱۹ | ۷.۳۲۷ | ۰.۷۱۷ | GAM۲ |
| ۰.۴۱۵ | ۰.۱۳۳ | ۱.۲۱۳ | ۰.۳۷۹ | ۲.۶۶۰ | ۱.۳۷۵ | TS - GAM۱ |
| ۰.۳۱۵ | ۰.۱۶۸ | ۱.۲۱۶ | ۰.۳۲۲ | ۲.۴۹۳ | ۱.۳۸۸ | JLN - GAM۱ |

برابر است با

$$\begin{aligned} \hat{f}_X(x) &= \hat{f}_Y(\log x) \frac{1}{x} = \frac{1}{nh} \sum_{i=1}^n \phi \left(\frac{\log x - \log X_i}{h} \right) \frac{1}{x} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{xh\sqrt{\pi}} \exp \left[-\frac{1}{2} \left(\frac{\log x - \log X_i}{h} \right)^2 \right] \\ &= \frac{1}{n} \sum_{i=1}^n \ln(x; \log X_i, h) \end{aligned}$$

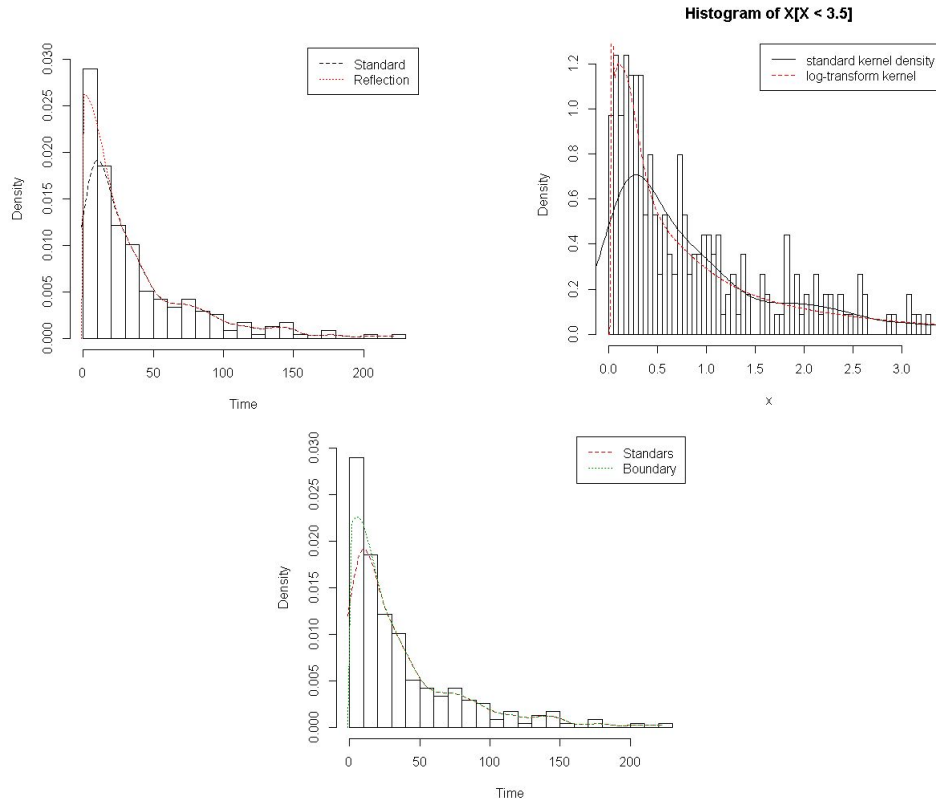
با استفاده از تبدیل لگاریتمی اعمال شده مشاهده می‌شود در این داده‌ها تابع هسته با تبدیل لگاریتمی نسبت به برآوردگر هسته استاندارد عملکرد مناسب‌تری داشته و از دقت خوبی برخوردار است، اما دارای نقاط ضعفی نیز هست، از جمله این‌که به‌طور کامل قادر به رفع مسئله مرزی نیست. در نمودار با استفاده از روش انعکاسی، برای X_i های نزدیک به ۰، مقداری از $\hat{f}(x)$ از X_i هایی که با $K_i(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$ مشخص می‌شود و در فاصله $(-\infty, 0)$ قرار می‌گیرد را با در نظر گرفتن قرینه X_i نسبت به ۰ این مقدار از دست داده‌شده را به فاصله $(0, \infty)$ برمی‌گرداند. همان‌طور که مشاهده می‌شود با توجه به مسئله مرزی برآوردگر هسته انعکاسی برآورد قابل قبولی را برای توزیع داده‌های طول عمر در مقایسه با روش هسته استاندارد بدون مسئله مرزی ارائه می‌دهد. در منحنی برآورد حاصل از روش هسته مرزی مشاهده می‌شود که برآوردگر هسته مرزی توانسته است با ارائه یک برآورد نامنفی، میزان احتمال در ناحیه مرزی را به‌خوبی برآورد کند و در خارج از یک محدوده خاص مرزی، برآوردگر هسته مرزی منطبق بر برآوردگر هسته استاندارد است. در شکل ۵ می‌توان مشاهده کرد، درحالی‌که برآوردگرهای KGAM1 و KGAM2 تا حد زیادی شبیه هم هستند به‌طور قابل‌توجهی نسبت به برآوردگر هسته استاندارد بهتر عمل می‌کنند و با ترکیب این دو هسته نامتقارن در دو رده MBC برآوردگرهای KGAM1-TS و KGAM1-JLN موجب افزایش عملکرد بهتر در پردازش می‌شوند. می‌توان مشاهده کرد که وقتی برآوردگر هسته استاندارد چگالی را کم (بیش‌تر) برازش می‌کند. برآوردگر JLN-MBC مربوط به آن در جهت عکس برازش را اصلاح می‌کند [۷]. این شکل نشان می‌دهد که KGAM1-JLN واقعاً در جهت اصلاح برازش هسته استاندارد موفق عمل کرده است.

با مقایسه برآوردگرهای هسته نامتقارن با دو رده از روش‌های MBC در شکل ۵ و برآوردگرهای روش تبدیل و انعکاسی و هسته مرزی در شکل ۴ برخی نکات مشخص می‌شود. از جمله این‌که برآوردگرهای معرفی‌شده مانند روش‌های قبلی رفع مسئله مرزی به‌خوبی در برخورد با داده‌های مثبت برآوردهایی بدون اریبی مرزی ارائه می‌دهند بنابراین می‌توانیم از عملکرد برآوردگرهای هسته نامتقارن و برآوردگرهای MBC

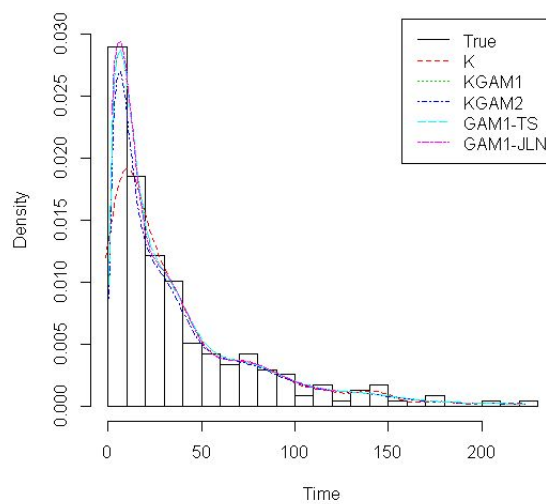
قرار می‌گیرد مربوط به طول عمر قرنیه‌های پیوند داده‌شده به ۱۱۹ بیمار است که در طول سال‌های ۱۳۶۵ تا ۱۳۸۰ در مراکز درمانی لبافی‌نژاد و حضرت رسول اکرم (ص) تهران تحت عمل جراحی پیوند قوز قرنیه چشم قرار گرفته‌اند. این بیماران پس از پیوند تا مدت‌زمان نسبتاً طولانی تحت پیگیری قرار گرفته‌اند و منظور از طول عمر قرنیه پیوندی، زمان از لحظه پیوند تا پس داده شدن قرنیه یا زمان آخرین معاینه قرنیه است. به این منظور اطلاعات بیماران شامل جنس و سن در زمان پیوند اول و اطلاعات هر یک از پیوندها از قبیل تاریخ عمل پیوند، اندازه قطر قرنیه دهنده برحسب میلی‌متر، دفع یا عدم دفع قرنیه پیوندی و تاریخ دفع یا تاریخ آخرین زمان معاینه چشم و برخی اطلاعات دیگر ثبت شده است. این داده‌ها توسط خیری [۱] موردبررسی قرار گرفتند. از آنجاکه این داده‌ها دارای یک مرز طبیعی در مبدأ هستند و توزیع آن‌ها به‌طور تجربی با تمرکز مشاهدات در نزدیکی مرز و یک دم بلند با داده‌های پراکنده مشخص می‌شوند، بنابراین انگیزه استفاده از هسته‌های نامتقارن برای برآورد را فراهم می‌کند و با توجه به این داده‌ها دارای مقادیری مثبت هستند، بنابراین حضور برآورد چگالی در ناحیه منفی، سبب مسئله مرزی می‌شود. در شکل ۴ نمودار چگالی این داده‌ها را با استفاده از برآورد چگالی هسته استاندارد و سه روش رفع مسئله مرزی تبدیل لگاریتمی، روش انعکاسی و روش هسته مرزی آورده شده است. شکل ۵ برآورد چگالی با استفاده از هسته‌های استاندارد، $GAM1$ ، $GAM2$ ، $TS-GAM1$ و $JLN-GAM1$ را برای این داده‌ها نشان می‌دهد.

در نمودار اول شکل ۴ از روش تبدیل در داده‌ها که برای اصلاح داده‌های چوله موفق عمل می‌کند، برای اصلاح مرزی به‌کاررفته است. استفاده از تبدیل لگاریتمی هنگامی که با مشاهدات مثبت و توزیع‌هایی با دم کشیده سروکار داریم رایج‌تر است. یک ویژگی تبدیل لگاریتم این است که دم راست توزیع را فشرده می‌کند. از آنجایی‌که اغلب توزیع طول عمر مطابقت خوبی با توزیع‌هایی از نوع پارتو یا لگ نرمال دارند، تبدیل لگاریتمی می‌تواند مورد استفاده قرار گرفته و سبب رفع ویژگی‌هایی در داده‌ها گردد که در برآوردگر چگالی هسته استاندارد مشکل‌ساز بوده و در نتیجه کیفیت برازش را در این داده‌ها بهبود دهد و از آنجایی‌که چگالی روی دامنه $(0, +\infty)$ تعریف شده است (در نتیجه مشاهدات مثبت‌اند)، حالت خاص تبدیل تابع لگاریتمی $Y = G(X) = \log X$ را در نظر می‌گیریم. اگر چگالی داده‌های تبدیل‌شده f_Y با برآوردگر چگالی هسته گاوسی برآورد شده باشد، آنگاه برآورد چگالی هسته تبدیل لگاریتمی

در ترکیب با هسته‌های نامتقارن در رفع مسئله مرزی مطمئن باشیم. نکته دیگر تطابق دو رده برآوردگرهای MBC و عملکرد نزدیک به هم این دو رده است که توانسته‌اند بدون اینکه پراکندگی قابل‌ملاحظه‌ای را سبب شوند، اریبی مرزی را کاهش داده و عملکردی بهینه به‌ویژه در ناحیه مرزی داشته باشند.



شکل ۴. نمودار هیستوگرام و مقایسه برآورد تابع چگالی طول عمر قرنیه‌های پیوندی با استفاده از تابع هسته استاندارد، روش تبدیل لگاریتمی، روش انعکاسی و روش هسته مرزی



شکل ۵. نمودار هیستوگرام و مقایسه برآورد تابع چگالی طول عمر قرنیه‌های پیوندی با استفاده از تابع هسته‌های استاندارد (K),

$GAM \setminus - JLN$, $GAM \setminus - TS$, $KGAM \setminus 2$, $KGAM \setminus 1$

۵ بحث و نتیجه‌گیری

مورد بررسی قرار گرفت. نتایج شبیه‌سازی‌ها دو رده MBC با استفاده از هسته‌های نامتقارن به‌منظور برآورد چگالی با دامنه $(0, \infty)$ و بررسی داده‌های واقعی طول عمر قرنی‌های پیوندی کاهش اریبی را با دو روش MBC با استفاده از هسته‌های نامتقارن تأیید می‌کند. لازم است یادآوری شود که اگرچه عملکرد برآوردگرهای MBC با استفاده از هسته‌های نامتقارن به‌صورت نظری تا حدود زیادی قانع‌کننده است و خواص بسیار خوبی از ترکیب این دو روش نشان داده شد، اما در کارهای عملی ضرورت دارد از نهایت امکانات و اطلاعات بهره برد و برای تحلیل یک مجموعه داده پیشنهاد می‌شود در صورت امکان از دیگر روش‌های برآورد چگالی نیز استفاده شود.

در این مقاله دو رده از روش‌های MBC با استفاده از هسته‌های نامتقارن جهت رفع مسئله مرزی در برآورد چگالی احتمال داده‌های نامنفی معرفی شد. ابتدا به‌صورت نظری مشخص شد با فرض اینکه چگالی واقعی به‌اندازه کافی هموار است، هر دو روش کاهش اریبی مرتبه اریبی را از $O(b)$ به $O(b^2)$ بهبود می‌بخشند، درحالی‌که مرتبه بزرگی واریانس بدون تغییر می‌ماند. MSE این دو رده با ساختار نامنفی، نرخ همگرایی از مرتبه $O(n^{-\frac{4}{3}})$ برای قسمت مرزی دارد. سپس عملکرد این روش با برآوردگرهای هسته استاندارد و دیگر روش‌های اصلاح مسئله مرزی

مراجع

- [۱] خیري، سلیمان، فقیه زاده سقراط، جوادی محمدعلی، جعفری نسب محمدرضا، حسینی سید محمدباقر و انیسیان آر.ش. (۱۳۸۲). بررسی زمان بقای پیوند دوطرفه قرنی در قوز قرنی با استفاده از مدل شکنندگی کاکس. مجله چشم‌پزشکی بینا. ۱: ۴۰-۴۵.
- [۲] رجبی، لیللا. (۱۳۹۶). بررسی روش تصحیح اریبی ضربی با استفاده از هسته‌های نامتقارن برای برآورد چگالی ناپارامتری داده‌های نامنفی. (پایان‌نامه کارشناسی ارشد)، دانشگاه شهید چمران اهواز.
- [3] Abramson, I.S., (1982). On bandwidth variation in kernel estimates a square root law. *Annals of Statistics*. **10**, 1217-1223
- [4] Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics and Data Analysis*, **31(2)**, 131-145.
- [5] Gasser, T., and Müller, H. G. (1979). Kernel estimation of regression functions. *Smoothing techniques for curve estimation Conference*, **757**, 23-68.
- [6] Gospodinov, N., 141 Hirukawa, M. (2012). Nonparametric estimation of scalar diffusion models of interest rates using asymmetric kernels. *Journal of Empirical Finance*, **19(4)**, 595-609.
- [7] Hirukawa, M. (2010). Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval. *Computational Statistics and Data Analysis*, **54(2)**, 473-495.
- [8] Hirukawa, M. (2018). *Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance*. Springer.
- [9] Hirukawa, M (2018). Bias Correction in Density Estimation. in *Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance*, 41-58.
- [10] Hirukawa, M., and Sakudo, M. (2012). Nonparametric Multiplicative Bias Correction for Kernel-Type Density Estimation Using Positive Data. *Working paper*, Faculty of Economics, Setunan University, Japan.
- [11] Hirukawa, M., and Sakudo, M. (2014). Nonnegative bias reduction methods for density estimation using asymmetric kernels. *Computational Statistics and Data Analysis*, **75**, 112-123.

- [12] Hirukawa, M., and Sakudo, M. (2019). Another bias correction for asymmetric kernel density estimation with a parametric start. *Statistics and Probability Letters*, 145, 158-165.
- [13] Igarashi, G. (2016). Bias reductions for beta kernel estimation. *Journal of Nonparametric Statistics*, **28**(1), 1-30.
- [14] Igarashi, G., and Kakizawa, Y. (2015). Bias corrections for some asymmetric kernel estimators. *Journal of Statistical Planning and Inference*, **159**, 37-63.
- [15] Igarashi, G., and Kakizawa, Y. (2020). Multiplicative bias correction for asymmetric kernel density estimators revisited. *Computational Statistics and Data Analysis*, **141**, 40-61.
- [16] Jin, X., and Kawczak, J. (2003). Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, **4**, 103-124.
- [17] Jones, M. C., Linton, O., and Nielsen, J. P. (1995). A simple bias reduction method for density estimation. *Biometrika*, **82**(2), 327-338.
- [18] Marron, J. S. and D. Ruppert (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B*, **56**, 653-671.1.
- [19] Müller, H. G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78**, 521-530.
- [20] Müller, H. G. (1993). On the boundary kernel method for non-parametric curve estimation near endpoints. *Scandinavian Journal of Statistics*, **20**, 313-328.
- [21] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. *Nonparametric statistics*, **16**(1-2), 217-226.
- [22] Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics-Theory and methods*, **14**, 1123-1136.
- [23] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- [24] Terrell, G. R., and Scott, D. W. (1980). On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, **8**(5), 1160-1163.
- [25] Zhang, S., Karunamuni, R. J., and Jones, M. C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, **94**(448), 1231-1240.

Investigation of multiplicative bias correction method using asymmetric kernels

Leila Rajabi ¹ and Behzad Mansouri ²

Abstract:

Kernel density estimation is a standard method for estimating the probability density function, which in many cases works well. However, it has been found that it does not work well for negative, sloping, and wide-tail distributions, which are common features of the distribution of longevity, income, and so on. The purpose of this paper is to evaluate the performance of multiplicative bias correction (MBC) methods using asymmetric kernel estimators and compare this estimator with other boundary problem solving methods. In this paper, in addition to introducing MBC methods in combination with asymmetric kernel estimators, a simulation study shows that this estimator can, in some cases, provide a much better fit for density estimation than the standard kernel estimator. MBC methods using asymmetric kernel estimators were also used to estimate the lifetime density of transplanted corneas in 119 patients.

Keywords: Boundary problem, Asymmetric kernel, Multiplicative bias correction, lifetime density.

¹ Master of Statistics, Shahid Chamran University of Ahvaz

² Associate Professor of Statistics, Shahid Chamran University of Ahvaz