

بهبود نسبی کاهش اریبی در برآوردگر چگالی با استفاده از هسته برون‌یابی شده هندسی

رضا صالحی*، علی شادرخ، مسعود یارمحمدی؛

دانشگاه پیام نور، گروه آمار، تهران

پذیرش ۹۶/۱۱/۲۸

دریافت ۹۶/۰۵/۳۰

چکیده

یکی از روش‌های ناپارامتری برآورد چگالی احتمال روش هسته‌ای است. در این مقاله، به منظور کاهش اریبی برآورد چگالی هسته‌ای، روش‌هایی مانند هسته متداول، هسته متداول برون‌یابی شده هندسی، هسته کاهش اریبی و هسته کاهش اریبی برون‌یابی شده هندسی معرفی و مورد بحث و بررسی قرار می‌گیرد. همچنین خواص نظری از جمله نحوه انتخاب پارامتر همواری و میزان دقت برآوردگرهای حاصل بررسی می‌شود. هرگاه هسته‌ها متقارن باشند، میانگین انتگرال مربعات خطای روش هسته‌ای کاهش اریبی برون‌یابی شده هندسی نسبت به سایر روش‌ها به نرخ همگرایی سریع‌تری دست می‌یابد. به منظور بررسی عملکرد این برآوردگرها، بررسی شبیه‌سازی مونت‌کارلو انجام شده است. همچنین نتایج به دست آمده به کمک تحلیل داده‌های واقعی نشان داده شده است. نتایج نشان می‌دهند که میزان اریبی در روش‌های هسته‌ای کاهش اریبی و هسته‌ای کاهش اریبی برون‌یابی شده هندسی به طور چشم‌گیری کاهش می‌یابد.

واژه‌های کلیدی: اریبی، برآورد چگالی، برون‌یابی هندسی، پارامتر همواری، هسته متقارن.

رده‌بندی ریاضی (۲۰۱۰): ۶۲G۰۷، ۶۲G۰۵

۱. مقدمه

یکی از تکنیک‌های برآورد چگالی، برآورد چگالی هسته‌ای است [1]. پژوهش‌های نظری این روش و کاربردهای آن به کمک رایانه‌ها برای پردازش داده‌ها، انقلابی در برآورد چگالی احتمال به وجود آورده است [2]. سیلورمن [3]، هاردل [4] و بسیاری دیگر از افراد این روش را بسط داده و آن را به ابعاد بالاتر تعمیم داده‌اند. از آن جاکه برآورد چگالی هسته‌ای با نمونه‌های کوچک نیز برآورد خوبی ارائه می‌کند از این رو، دارای کاربردهای گوناگونی در زمینه‌های بانک‌داری [5]، اقلیم‌شناسی [6]، گراف‌های رایانه‌ای [7] اقتصاد [8]، جغرافیا [9]، آب‌شناسی [10]، پردازش تصاویر [11]، علوم اعصاب [12]، نظارت [13]، باستان‌شناسی [14]، پزشکی [15] و غیره است. تکنیک برآورد چگالی هسته‌ای در اواخر دهه ۱۹۵۰ برای برآورد چگالی بافت‌نگار به‌عنوان نمایش گرافی یک توزیع فراوانی بیان شد. سیلورمن [1] و پارزن [16] این روش را بیان و توسعه دادند. فارل [17] بهترین نرخ هم‌گرایی مجانبی $O(n^{\frac{2r}{2r+1}})$ را به کمک میانگین مربعات خطا^۱ (MSE) برای برآوردگرهای هسته‌ای متعامد به دست آورد. سامیودین و ال سید [18]

هسته‌هایی را که باعث کاهش اریبی می‌شدند معرفی کردند. ال‌سید و همکاران [19] برآوردگری با استفاده از بعضی شناسه‌های احتمالاتی برای کاهش اریبی پیشنهاد دادند. مارون و روپرت [20] الگوریتمی مناسب برای چندجمله‌ای‌های درجه دوم و تابع توزیع تجمعی بتا در کاهش اریبی مرزی پیشنهاد کردند. کیم و همکاران [21] یک روش چاوله برای کاهش اریبی توسعه دادند که واریانس تنها از طریق فاکتور ثابت مناسب افزایش می‌یافت. مانبف و مارتینز-فیو [22] روی اریبی کاهش یافته نسبی هسته در برآوردگر هسته کلاسیک از طریق شرط لایپ‌شیتز کار کردند. این روش به‌طور طبیعی انتخاب سازوار پارامترهای هموارسازی را رهبری کرده و می‌تواند باعث کاهش اریبی شود. کایرات و همکاران [23] بهبودی اریبی را با هسته‌های اپانچ‌نیکوف و گرام-شارلیه در برآورد چگالی هسته‌ای بررسی کردند. کیم و کیم [24] یک نسخه از برآوردگر چگالی هسته‌ای که MSE را کاهش می‌داد ارائه کردند. آن‌ها نشان دادند که کاهش MSE ناشی از کاهش اریبی است نه واریانس. ایگراشی و کاکیزاوا [25] برآوردگر هسته متقارن با تکیه‌گاه $(0, \infty)$ را به‌کمک تکنیک‌های اصلاح اریبی نامنفی با هسته نرمال استاندارد بررسی کردند و نشان دادند که بهترین میانگین انتگرال مربعات خطا با $O(n^{-\frac{8}{9}})$ رخ می‌دهد. شارون و جینگ جینگ [26] رویکرد جدیدی برای کاهش اریبی برآوردگر چگالی هسته‌ای ارائه کردند.

چنان‌چه از پژوهش‌های مذکور پیداست در دهه‌های گذشته کوشش بسیاری شده که برآورد چگالی احتمال با دقت بالاتری ارائه شود. از این رو، در این مقاله به کاهش اریبی برآوردگرهای هسته‌ای متقارن با استفاده از برون‌یابی هندسی می‌پردازیم. در بخش ۲، برآوردگر هسته متداول را معرفی می‌شود در بخش ۳، برآوردگر هسته‌ای کاهش اریبی بررسی می‌شود. در بخش ۴، روش برون‌یابی هندسی برای کاهش اریبی را بررسی می‌کنیم. روش‌های شبیه‌سازی در بخش ۵، را ارائه کرده، نتایج را به‌کمک هسته نرمال استاندارد بررسی و مقایسه می‌کنیم. در بخش ۶، سه مثال عددی از داده‌های واقعی را بررسی کرده و سرانجام خلاصه نتایج را در بخش ۷، بیان می‌شود.

۲. برآوردگر هسته‌ای متداول

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی n تایی از یک توزیع با تابع چگالی احتمال مجهول f باشد.

برآورد چگالی هسته‌ای متداول¹ (UK) در نقطه x را با \hat{f} نشان داده و بدین‌صورت تعریف می‌شود:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (1)$$

که در آن h یک عدد حقیقی مثبت، پهنای پنجره یا پارامتر همواری نامیده می‌شود و درجه‌ای از همواری را کنترل می‌کند و $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$ تابع هسته‌ای نامیده می‌شود. تابع چگالی نامعلوم f ، تابع هسته K و پارامتر همواری h معمولاً در این روابط صدق می‌کنند:

$$a1) \int_{-\infty}^{\infty} K(x) dx = 1$$

$$a2) \sup_{-\infty < x < \infty} |K(x)| < \infty$$

$$a3) \lim_{x \rightarrow \infty} |xK(x)| = 0$$

$$\begin{aligned} a4) \quad & K(x) = K(-x) \\ a5) \quad & h \rightarrow 0, nh \rightarrow \infty \text{ if } n \rightarrow \infty \\ a6) \quad & \|f''\|_2^2 = \int_{-\infty}^{\infty} f''^2(x) dx < \infty \\ a7) \quad & \|K\|_2^2 = \int_{-\infty}^{\infty} K^2(x) dx < \infty \\ a8) \quad & \int_{-\infty}^{\infty} xK(x) dx = 0 \\ a9) \quad & \mu_2(K) = \int_{-\infty}^{\infty} x^2 K(x) dx < \infty \end{aligned}$$

پارزن [۱۶] نشان داد تحت شرایط $a1$ تا $a5$ برآوردگر \hat{f}_h سازگار در میانگین درجه دو است. چنان که ملاحظه می شود برآوردگر هسته ای (۱) به پارامتر همواری h بستگی دارد. اگر مقدار h خیلی کوچک اختیار شود برآوردگر \hat{f}_h بسیار موج خواهد بود. از این رو، لازم است روشی مناسب برای تعیین مقدار بهینه h ارائه شود. فرض کنید تابع چگالی نامعلوم f و تابع هسته K در شرایط $a6$ تا $a9$ صدق کنند و هم چنین f'' در همسایگی x مطلقاً پیوسته باشد. بر اساس سیلورمن [۳] و واند و جونس [۲۷] داریم:

$$Bias(\hat{f}_h(x)) = E(\hat{f}_h(x)) - f(x) = \frac{1}{2} h^2 f''(x) \int y^2 K(y) dy + O(h^3) \quad (۲)$$

و

$$Var(\hat{f}_h(x)) = E(\hat{f}_h(x) - E(\hat{f}_h(x)))^2 = \frac{1}{nh} f(x) \int K^2(y) dy + O((nh)^{-1}) \quad (۳)$$

بنابراین از (۲) و (۳) نتیجه می گیریم که:

$$\begin{aligned} MSE(\hat{f}_h(x)) &= Bias^2(\hat{f}_h(x)) + var(\hat{f}_h(x)) \\ &\sim \frac{h^4}{4} (f''(x))^2 \left(\int y^2 K(y) dy \right)^2 + \frac{1}{nh} f(x) \int K^2(y) dy + O\left(\frac{1}{n} + h^5\right). \end{aligned}$$

در نتیجه MSE بهینه از مرتبه $O\left(n^{-\frac{4}{5}}\right)$ با $h_{opt} = O\left(n^{-\frac{1}{5}}\right)$ خواهد بود.

۳. برآوردگر هسته ای کاهش اربیبی

به منظور کاهش اربیبی برآوردگر چگالی هسته ای متداول UK ، برآوردگر چگالی هسته ای متداول با پارامتر همواری ah در نظر بگیرید:

$$\hat{f}_{ah}(x) = \frac{1}{anh} \sum_{i=1}^n K\left(\frac{x - X_i}{ah}\right), \quad a > 0$$

تحت شرایط $a1$ تا $a5$ داریم:

$$\begin{aligned} Bias(\hat{f}_{ah}(x)) &= E(\hat{f}_{ah}(x)) - f(x) = ahf'(x) \int xK(x) dx + \frac{(ah)^2}{2!} f''(x) \int x^2 K(x) dx \\ &+ \frac{(ah)^3}{3!} f'''(x) \int x^3 K(x) dx + O(h^4). \end{aligned}$$

با کم کردن مقدار برآورد اربیبی از برآوردگر هسته ای متداول داریم:

$$\begin{aligned} \tilde{f}_{ah}(x) &= \hat{f}_{ah}(x) - Bias(\hat{f}_{ah}(x)) \\ &= \frac{1}{anh} \sum_{i=1}^n K\left(\frac{x-X_i}{ah}\right) - ahf'(x) \int xK(x)dx - \frac{(ah)^2}{2!} f''(x) \int x^2K(x)dx \\ &\quad - \frac{(ah)^3}{3!} f'''(x) \int x^3K(x)dx + O(h^4). \end{aligned}$$

اگر هسته K متقارن و تمام گشتاورهای فرد آن صفر باشد آن‌گاه برآوردگر پیشنهادی هسته‌ای کاهش اریبی (BRK)^۱ در نقطه x را با \tilde{f} نشان داده و به صورت زیر خواهد بود:

$$\begin{aligned} \tilde{f}_{ah}(x) &= \frac{1}{anh} \sum_{i=1}^n K\left(\frac{x-X_i}{ah}\right) - \frac{(ah)^2}{2!} f''(x) \int x^2K(x)dx \\ &= \frac{1}{anh} \sum_{i=1}^n K\left(\frac{x-X_i}{ah}\right) - \frac{ah}{2n} \int x^2K(x)dx \sum_{i=1}^n K''\left(\frac{x-X_i}{h}\right). \end{aligned} \tag{۴}$$

برای بررسی این که این برآوردگر جدید از نظر ساختاری قادر به کاهش اریبی و در نتیجه کاهش MSE هست یا خیر، اریبی و واریانس آن را محاسبه می‌کنیم.

قضیه ۱. فرض کنید هسته K متقارن و تمام گشتاورهای فرد آن صفر باشند و f در همسایگی x دارای مشتق مرتبه چهار باشد. همچنین فرض کنید شرایط $a4$ ، $a5$ و $a8$ برقرار باشند. بنابراین داریم:

$$Bias(\tilde{f}_{ah}(x)) = -\frac{(ah)^4}{2!2!} f'''(x) \left[\int x^2K(x)dx \right]^2 + O(h^4)$$

و

$$Var(\tilde{f}_{ah}(x)) = \frac{f(x)}{2anh} \left[\int x^2K(x)dx \right]^2 \int (K'')^2(x)dx + O(n^{-1})$$

و در نتیجه MSE بهینه از مرتبه $O\left(n^{-\frac{8}{9}}\right)$ با $O\left(n^{-\frac{1}{9}}\right)$ است. h_{opt}

اثبات. با استفاده از بسط تیلور $f(x+hs)$ حول x داریم:

$$\begin{aligned} E\tilde{f}_{ah}(x) &= E\hat{f}_{ah}(x) - \frac{(ah)^2}{2!} E(\hat{f}_{ah}''(x)) \int x^2K(x)dx \\ &= f(x) + \frac{(ah)^2}{2!} f''(x) \int x^2K(x)dx + O(h^4) - \frac{(ah)^2}{2!} E(\hat{f}_{ah}''(x)) \int x^2K(x)dx \end{aligned}$$

که

$$\begin{aligned} E(\hat{f}_{ah}''(x)) &= E\left[\frac{1}{anh(ah)^2} \sum_{i=1}^n K''\left(\frac{x-X_i}{ah}\right) \right] \\ &= \frac{1}{(ah)^3} \int K''\left(\frac{x-t}{ah}\right) f(t)dt \\ &= \frac{1}{ah} \int K\left(\frac{x-t}{ah}\right) f''(t)dt = \int K(s) f''(x-ahs)ds \\ &= \int K(s) \left[f''(x) - ah sf'''(x) + \frac{(ahs)^2}{2!} f''''(x) - \frac{(ahs)^3}{3!} f''''''(x) \right] ds + O(h^4) \end{aligned}$$

1. Bias Reduction Kernel

$$= f''(x) + \frac{(ah)^2}{2} f'''(x) \int s^2 K(s) ds + O(h^4).$$

بنابراین داریم:

$$\begin{aligned} E\tilde{f}_{ah}(x) &= f(x) + \frac{(ah)^2}{2!} f''(x) \int x^2 K(x) dx + O(h^4) \\ &\quad - \frac{(ah)^2}{2!} \int x^2 K(x) dx \left[f''(x) + \frac{(ah)^2}{2!} f'''(x) \int s^2 K(s) ds \right] + O(h^4) \\ &= f(x) - \frac{(ah)^2 (ah)^2}{2! 2!} f'''(x) \left[\int s^2 K(s) ds \right]^2 + O(h^4). \end{aligned}$$

در نتیجه میزان اریبی برابر است با:

$$\text{Bias}(\tilde{f}_{ah}(x)) = -\frac{(ah)^4}{2! 2!} f'''(x) \left[\int s^2 K(s) ds \right]^2 + O(h^4).$$

از طرف دیگر واریانس برآوردگر برابر است با

$$\begin{aligned} \text{Var}(\tilde{f}_{ah}(x)) &= \text{Var} \left[\hat{f}_{ah}(x) - \frac{(ah)^2}{2!} \hat{f}_{ah}''(x) \int s^2 K(s) ds \right] \\ &\leq 2\text{Var}(\hat{f}_{ah}(x)) + 2\text{Var} \left(\frac{(ah)^2}{2!} \hat{f}_{ah}''(x) \int s^2 K(s) ds \right) \quad (5) \\ &= 2\text{Var}(\hat{f}_{ah}(x)) + \frac{(ah)^4}{2} \left(\int s^2 K(s) ds \right)^2 \text{Var}(\hat{f}_{ah}''(x)) \end{aligned}$$

که

$$\begin{aligned} \text{Var}(\hat{f}_{ah}''(x)) &= \text{Var} \left[\frac{1}{n(ah)^3} \sum_{i=1}^n K'' \left(\frac{x - X_i}{ah} \right) \right] \\ &= \frac{1}{n(ah)^6} \text{Var} \left(K'' \left(\frac{x - X_1}{ah} \right) \right) \\ &= \frac{1}{n(ah)^6} \left\{ \int \left(K'' \left(\frac{x-t}{ah} \right) \right)^2 f(t) dt - \left[\int K'' \left(\frac{x-t}{ah} \right) f(t) dt \right]^2 \right\} \quad (6) \\ &= \frac{1}{n(ah)^5} \int (K''(s))^2 f(x - ahs) ds - \frac{1}{n} (f''(x) + O(h^2))^2 \\ &= \frac{1}{n(ah)^5} \int (K''(s))^2 (f(x) - ahsf'(x) + O(h^2)) ds - \frac{1}{n} (f''(x))^2 + O\left(\frac{h^2}{n}\right) \\ &= \frac{f(x)}{n(ah)^5} \int (K''(s))^2 ds + O\left(\frac{1}{nh^4}\right). \end{aligned}$$

با جای گذاری (۶) در (۵) داریم:

$$\begin{aligned} \text{Var}(\tilde{f}_{ah}(x)) &\leq 2 \left(\frac{f(x)}{nah} \int K^2(s) ds + O\left(\frac{1}{nh}\right) \right) + \\ &\frac{(ah)^4}{2} \left[\int s^2 K(s) ds \right]^2 \frac{f(x)}{n(ah)^5} \int (K''(s))^2 ds + O\left(\frac{1}{nh^4}\right) \\ &= \frac{f(x)}{2anh} \left[\int s^2 K(s) ds \right]^2 \int (K''(s))^2 ds + O\left(\frac{1}{n}\right). \end{aligned}$$

بنابراین به‌سادگی می‌توان دید که پارامتر همواری برابر $h = O(n^{-\frac{1}{9}})$ است. هم‌چنین MSE بهینه از مرتبه $O(n^{-\frac{8}{9}})$ خواهد بود، از این‌رو، اثبات کامل است.

تذکر ۱،۳. برآوردگر BRK در رابطه (۴) ممکن است در بعضی نقاط مرزی x منفی باشد از این‌رو، برای این‌که برآوردگر چگالی مثبتی داشته باشیم، آن را به‌صورت (۷) باز تعریف می‌کنیم:

$$\tilde{f}_{ah}(x) = \frac{\tilde{f}_{ah}(x) I_{(\tilde{f}_{ah}(x) > 0)}}{\int \tilde{f}_{ah}(x) I_{(\tilde{f}_{ah}(x) > 0)} dx} \quad (7)$$

که در آن I_A تابع نشان‌گر روی مجموعه A و \tilde{f} برآوردگر BEK باز تعریف شده است.

۴. برآوردگر هسته‌ای کاهش اریبی برون‌یابی هندسی

برون‌یابی هندسی برای برآورد چگالی هسته‌ای را ترل و اسکات معرفی کردند [28]. برآوردگر چگالی هسته‌ای متداول با دو پارامتر همواری مختلف h و $2h$ را در نظر بگیرید:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

و

$$\hat{f}_{2h}(x) = \frac{1}{2nh} \sum_{i=1}^n K\left(\frac{x - X_i}{2h}\right).$$

فرض کنید هسته K متقارن و تمام گشتاورهای فرد آن صفر باشند. ترل و اسکات [28] برآوردگر هسته‌ای متداول برون‌یابی هندسی^۱ (GEUK) در نقطه x را با f^* نشان داده و بدین‌صورت پیشنهاد دادند:

$$f_h^*(x) = \left(\hat{f}_h(x)\right)^{\frac{4}{3}} \left(\hat{f}_{2h}(x)\right)^{-\frac{1}{3}}. \quad (8)$$

توجه کنید که انتگرال برآوردگر $f_h^*(x)$ برابر یک نمی‌شود. ترل و اسکات [28] مقدار MSE بهینه از مرتبه $O(n^{-\frac{8}{9}})$ و پارامتر همواری برابر $h_{opt} = O(n^{-\frac{1}{9}})$ به‌دست آوردند.

حال برآوردگر چگالی هسته‌ای متداول با دو پارامتر همواری مختلف h و ah را در نظر بگیرید. بنابراین با کم کردن اریبی برآوردگر هسته‌ای متداول از آن داریم:

$$\tilde{f}_h(x) = \hat{f}_h(x) - \text{Bias}(\hat{f}_h(x)) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) - \frac{h^2}{2!} f''(x) \int s^2 K(s) ds$$

و

$$\tilde{f}_{ah}(x) = \hat{f}_{ah}(x) - \text{Bias}(\hat{f}_{ah}(x)) = \frac{1}{anh} \sum_{i=1}^n K\left(\frac{x-X_i}{ah}\right) - \frac{(ah)^2}{2!} f''(x) \int s^2 K(s) ds.$$

فرض کنید $I_h(x) = E(\tilde{f}_h(x))$ باشد. بنابراین برای هسته‌های متقارن K که تمام گشتاورهای فرد آن صفر باشند داریم:

$$\begin{aligned} I_h(x) &= f(x) - \frac{h^4}{2!2!} f'''(x) \left(\int s^2 K(s) ds \right)^2 + O(h^4) \\ &= f(x) \left[1 - \frac{h^4}{f(x)2!2!} f'''(x) \left(\int s^2 K(s) ds \right)^2 \right] + O(h^4). \end{aligned}$$

حال با لگاریتم‌گیری از رابطه مذکور داریم:

$$\log(I_h(x)) = \log(f(x)) + \log \left[1 - \frac{h^4}{f(x)2!2!} f'''(x) \left(\int s^2 K(s) ds \right)^2 \right].$$

با استفاده از بسط تیلور $\log(1-x)$ داریم:

$$\begin{aligned} \log(I_h(x)) &= \log(f(x)) - \frac{h^4}{f(x)} \left(\frac{f'''(x)}{2!2!} \left(\int s^2 K(s) ds \right)^2 \right) - \frac{1}{2} \frac{h^8}{f^2(x)} \left(\frac{f'''(x)}{2!2!} \left(\int s^2 K(s) ds \right)^2 \right)^2 \\ &\quad - \frac{1}{3} \frac{h^{12}}{f^3(x)} \left(\frac{f'''(x)}{2!2!} \left(\int s^2 K(s) ds \right)^2 \right)^3 + \dots \\ &= \log(f(x)) - \frac{m}{f(x)} h^4 - \frac{1}{2} \frac{m^2}{f^2(x)} h^8 - \frac{1}{3} \frac{m^3}{f^3(x)} h^{12} + \dots \end{aligned}$$

که

$$m = \frac{f'''(x)}{2!2!} \left(\int s^2 K(s) ds \right)^2.$$

به‌طور مشابه برای پارامتر ah داریم:

$$\log(I_{ah}(x)) = \log(f(x)) - \frac{m}{f(x)} (ah)^4 - \frac{1}{2} \frac{m^2}{f^2(x)} (ah)^8 - \frac{1}{3} \frac{m^3}{f^3(x)} (ah)^{12} + \dots.$$

بنابراین به‌منظور کاهش اریبی، برآوردگر هسته‌ای برون‌یابی شده هندسی در نقطه x را با \tilde{f}^* نشان داده و بدین صورت پیشنهاد می‌دهیم:

$$\tilde{f}_h^*(x) = \left(\tilde{f}_h(x) \right)^{w_1} \left(\tilde{f}_{ah}(x) \right)^{w_2}.$$

حال با لگاریتم‌گیری از برآوردگر بالا داریم:

$$\begin{aligned} \log(\tilde{f}_h^*(x)) &= w_1 \log(\tilde{f}_h(x)) + w_2 \log(\tilde{f}_{ah}(x)) = w_1 \log(I_h(x)) + w_2 \log(I_{ah}(x)) \\ &= w_1 \log(f(x)) - w_1 \frac{m}{f(x)} h^4 - \frac{1}{2} w_1 \frac{m^2}{f^2(x)} h^8 - \frac{1}{3} w_1 \frac{m^3}{f^3(x)} h^{12} + \dots \\ &\quad + w_2 \log(f(x)) - w_2 \frac{m}{f(x)} (ah)^4 - \frac{1}{2} w_2 \frac{m^2}{f^2(x)} (ah)^8 - \frac{1}{3} w_2 \frac{m^3}{f^3(x)} (ah)^{12} + \dots \\ &= (w_1 + w_2) \log(f(x)) - (w_1 + a^4 w_2) \frac{m}{f(x)} h^4 - \frac{1}{2} (w_1 + a^8 w_2) \frac{m^2}{f^2(x)} h^8 + \dots. \end{aligned}$$

از این‌رو، برای ارائه برآوردگری مطلوب که قادر به کاهش اریبی باشد لازم است عبارت

$$w_1 \log(I_h(x)) + w_2 \log(I_{ah}(x))$$

بر اساس $\log(f(x))$ باشد و h^4 حذف شود. بنابراین باید w_1 و w_2 در روابط زیر صدق کنند:

$$w_1 + w_2 = 1 \Rightarrow \begin{cases} w_1 = \frac{a^4}{a^4 - 1} \\ w_2 = \frac{-1}{a^4 - 1} \end{cases}, \quad a > 0, a \neq 1.$$

در نتیجه:

$$\frac{a^4}{a^4 - 1} \log(I_h(x)) + \frac{-1}{a^4 - 1} \log(I_{ah}(x)) = \log(f(x)) + \frac{a^4}{2} \frac{m^2}{f^2(x)} h^8 + \frac{a^4(a^4 + 1)}{3} \frac{m^3}{f^3(x)} h^{12} + \dots$$

اکنون با استفاده از بسط سری توابع نمایی داریم:

$$(I_h(x))^{a^4} (I_{ah}(x))^{-1} = f(x) + \frac{a^4}{2} \frac{m^2}{f(x)} h^8 + O(h^{12}).$$

با بازنویسی \tilde{f}_h و \tilde{f}_{ah} ، به ترتیب داریم:

$$\tilde{f}_h(x) = I_h(x) + Z$$

و

$$\tilde{f}_{ah}(x) = I_{ah}(x) + W.$$

که در آن متغیرهای Z و W هر دو دارای امید ریاضی برابر صفر و واریانس آن‌ها از مرتبه $O((nh)^{-1})$ است.

بنابراین برآوردگر هسته‌ای کاهش اریبی برون‌یابی هندسی^۱ (GEBRK) پیشنهادی نهایی بدین صورت است:

$$\begin{aligned} \tilde{f}_h^*(x) &= (\tilde{f}_h(x))^{a^4} (\tilde{f}_{ah}(x))^{-1} \\ &= (I_h(x))^{a^4} \left[1 + \frac{Z}{I_h(x)} \right]^{a^4} (I_{ah}(x))^{-1} \left[1 + \frac{W}{I_{ah}(x)} \right]^{-1}. \end{aligned} \tag{9}$$

حال با استفاده از بسط $(1+k)^x = 1+kx + O(k^2)$ داریم:

$$\begin{aligned} \tilde{f}_h^*(x) &= (I_h(x))^{a^4} (I_{ah}(x))^{-1} \left[1 + \frac{a^4}{a^4 - 1} \frac{Z}{I_h(x)} + O\left(\left(\frac{Z}{I_h(x)}\right)^2\right) \right] \\ &\quad \times \left[1 + \frac{-1}{a^4 - 1} \frac{W}{I_{ah}(x)} + O\left(\left(\frac{W}{I_{ah}(x)}\right)^2\right) \right] \\ &= (I_h(x))^{a^4} (I_{ah}(x))^{-1} + \frac{a^4}{a^4 - 1} Z \left(\frac{I_h(x)}{I_{ah}(x)}\right)^{\frac{1}{a^4 - 1}} + \frac{-1}{a^4 - 1} W \left(\frac{I_h(x)}{I_{ah}(x)}\right)^{\frac{a^4}{a^4 - 1}} + O((Z+W)^2). \end{aligned}$$

هم‌چنین با استفاده از تعریف $I_h(x)$ ، داریم:

$$\frac{I_h(x)}{I_{ah}(x)} = 1 + O(h^4).$$

بنابراین

$$\begin{aligned} Bias(\tilde{f}_h^*(x)) &= E\left[\left(\tilde{f}_h(x)\right)^{\frac{a^4}{a^4-1}}\left(\tilde{f}_{ah}(x)\right)^{\frac{-1}{a^4-1}}\right] - f(x) \\ &= E\left[\left(I_h(x)\right)^{\frac{a^4}{a^4-1}}\left(I_{ah}(x)\right)^{\frac{-1}{a^4-1}} + \frac{a^4}{a^4-1}Z\left(1+O(h^4)\right)^{\frac{1}{a^4-1}}\right. \\ &\quad \left. + \frac{-1}{a^4-1}W\left(1+O(h^4)\right)^{\frac{a^4}{a^4-1}} + O\left((Z+W)^2\right)\right] - f(x) \\ &= f(x) + \frac{1}{2}a^4m^2h^8 + O(h^{12}) + 0 + 0 + O\left(\frac{1}{nh}\right) - f(x) = \frac{1}{2}a^4m^2h^8 + O\left(h^{12} + \frac{1}{nh}\right) \end{aligned}$$

9

$$\begin{aligned} Var(\tilde{f}_h^*(x)) &= Var\left[\left(\tilde{f}_h(x)\right)^{\frac{a^4}{a^4-1}}\left(\tilde{f}_{ah}(x)\right)^{\frac{-1}{a^4-1}}\right] \\ &= E\left[\frac{a^4}{a^4-1}Z + \frac{-1}{a^4-1}W + O\left((Z+W)^2\right)\right]^2 \\ &= Var\left[\frac{a^4}{a^4-1}Z + \frac{-1}{a^4-1}W\right] + O\left(n^{-1}h^3 + (nh)^{-2}\right) = O\left(\frac{1}{nh}\right). \end{aligned}$$

در نتیجه MISE برآوردگر $\tilde{f}_h^*(x)$ از مرتبه $O\left(h^{16} + \frac{1}{nh}\right)$ است. بنابراین به سادگی می‌توان

دید که پارامتر همواری برابر $h_{opt} = O\left(n^{\frac{-1}{17}}\right)$ است. هم‌چنین MSE بهینه از مرتبه $O\left(n^{\frac{-16}{17}}\right)$ خواهد بود. نتایج

بالا در قضیه ۲ خلاصه شده است.

قضیه ۲. فرض کنید هسته K متقارن و تمام گشتاورهای فرد آن صفر باشند و f در همسایگی x دارای مشتق مرتبه چهار باشد. هم‌چنین فرض کنید شرایط $a5$ تا $a9$ برقرار باشند. در این صورت داریم:

$$\begin{aligned} Bias(\tilde{f}_h^*(x)) &= \frac{a^4}{32}h^8(f'''(x))^2\left(\int x^2K(x)dx\right)^2 + O\left(h^{12} + \frac{1}{nh}\right), \\ Var(\tilde{f}_h^*(x)) &= O\left(\frac{1}{nh}\right). \end{aligned}$$

در نتیجه

$$\begin{aligned} MSE(\tilde{f}_h^*(x)) &= O\left(h^{16} + \frac{1}{nh}\right) \\ \text{و MSE بهینه از مرتبه } O\left(n^{\frac{-16}{17}}\right) \text{ با } h_{opt} &= O\left(n^{\frac{-1}{17}}\right) \text{ است.} \end{aligned}$$

۵. پژوهش‌های شبیه‌سازی

در این بخش با استفاده از شبیه‌سازی، ابتدا مقدار بهینه برای پارامتر a را پیدا می‌کنیم، سپس به کمک شبیه‌سازی مونت کارلو برآوردگرهای حاصل را بررسی و مقایسه می‌کنیم.

۱.۵. انتخاب بهینه پارامتر a

در این بخش با استفاده از بررسی شبیه‌سازی نشان می‌دهیم که کمینه کردن $MSE(\tilde{f}_h(x))$ حاصل یک انتخاب قابل اطمینانی برای پارامتر a است. بنابراین هر مقداری از a را که MSE برآوردگر پیشنهادی را کمینه کند به‌عنوان برآوردگر بهینه برای کاهش اریبی معرفی می‌کنیم. چون کمینه کردن MSE به‌کمک مشتق‌گیری امکان‌پذیر نیست بنابراین با استفاده از روش‌های عددی سعی می‌کنیم MSE را به حداقل برسانیم. از آن‌جاکه میزان اریبی برابر ناحیه اشغال شده بین چگالی واقعی و چگالی برآورد شده است، با به‌کارگیری روش‌های عددی مانند روش مستطیلی و روش سیمپسون (بوردن و فیروز [29]) مقدار بهینه a را با کمینه کردن MSE به‌دست می‌آوریم. داده‌های به‌کار رفته در شبیه‌سازی از توزیع نرمال استاندارد تولید شده‌اند. همچنین هسته K ، برابر چگالی نرمال استاندارد در نظر گرفته شده است. با توجه به شکل ۵ مقادیر بهینه برای پارامتر a به‌ترتیب در دو روش مستطیلی و سیمپسون در برآوردگر BRK برابر ۰/۵۸ و ۰/۶۹ و در برآوردگر GEBRK برابر ۰/۶۱ و ۰/۶۰ به‌دست آمد. اما از آن‌جاکه دقت روش سیمپسون بیش‌تر است بنابراین مقدار ۰/۶۹ در برآوردگر BRK و مقدار ۰/۶۰ در برآوردگر GEBRK را برای مقدار بهینه a پیشنهاد می‌دهیم. در نتیجه برآوردگرهای نهایی پیشنهادی بدین‌صورت هستند.

برآوردگر BRK پیشنهادی برابر با:

$$\tilde{f}_{ah}(x) = \frac{1}{0.69nh} \sum_{i=1}^n K\left(\frac{x-X_i}{0.69h}\right) - \frac{0.69h}{2n} \int s^2 K(s) dx \sum_{i=1}^n K^n\left(\frac{x-X_i}{0.69h}\right) \quad (10)$$

و برآوردگر GEBRK پیشنهادی برابر با:

$$\tilde{f}_h^*(x) = (\hat{f}_h(x))^{-0.15} (\hat{f}_{0.60h}(x))^{1.15} \quad (11)$$

است. این برآوردگرها مقدار اریبی را نسبت به برآوردگر GEUK به‌طور چشم‌گیری کاهش می‌دهند.

۲.۵. بررسی شبیه‌سازی مونت کارلو

در این بخش مطالعه شبیه‌سازی برای کاهش اریبی با استفاده از شبیه‌سازی مونت کارلو برای چهار برآوردگر: برآوردگر UK داده شده در رابطه (۱)، برآوردگر GEUK داده شده در رابطه (۸)، برآوردگر BRK داده شده در رابطه (۱۰) و برآوردگر GEBRK داده شده در رابطه (۱۱) را انجام می‌دهیم. به‌صورت تصادفی ۱۰۰۰ نمونه مستقل به حجم‌های ۲۰، ۵۰، ۲۰۰ و ۵۰۰ انتخاب می‌کنیم. با توجه به این‌که توزیع نرمال استاندارد حول صفر متقارن است از این‌رو، مقدار x را برابر ۰، ۰/۵، ۱، ۱/۵، ۲، ۲/۵، ۳، ۳/۵ در نظر گرفته و برآوردگرهای هسته‌ای در این نقاط محاسبه و با هم مقایسه می‌شوند. همچنین نرمال استاندارد را به‌عنوان تابع هسته K انتخاب می‌کنیم. از آن‌جاکه هسته K متقارن است بنابراین پارامتر همواری را برای برآوردگر UK برابر $h = n^{-\frac{1}{5}}$ ، برای برآوردگرهای GEUK و BRK برابر $h = n^{-\frac{1}{9}}$ و برای برآوردگر GEBRK برابر $h = n^{-\frac{1}{17}}$ انتخاب می‌کنیم. اریبی، واریانس و MSE برآوردگرها بدین صورت برآورد می‌شوند:

$$Bias(\hat{f}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{f}_i - f),$$

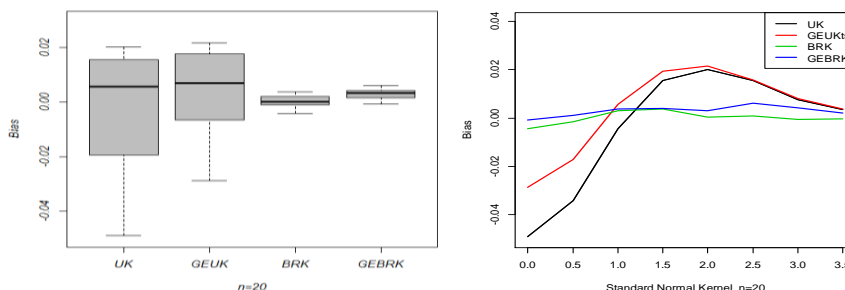
$$Var(\hat{f}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{f}_i - \bar{f})^2 \quad \text{with} \quad \bar{f} = \frac{1}{1000} \sum_{i=1}^{1000} \hat{f}_i$$

$$MSE(\hat{f}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{f}_i - f)^2$$

که f چگالی واقعی $f(x)$ و \hat{f} هر کدام از برآوردگرهای $\hat{f}_h(x)$ ، $f_h^*(x)$ ، $\tilde{f}_h(x)$ و $\tilde{f}_h^*(x)$ است. بدون از دست دادن کلیت، فرض می‌کنیم که f برابر چگالی نرمال استاندارد است. نتایج شبیه‌سازی در جدول‌های ۱ تا ۴ و شکل‌های ۱ تا ۵ ارائه شده است. با توجه به جدول‌های ۱ تا ۴ و شکل‌های ۲ و ۳ دو برآوردگر UK و GEUK در کاهش اریبی عملکرد مشابهی دارند. همچنین دو برآوردگر BRK و GEBRK نیز عملکرد مشابهی در کاهش اریبی دارند. در حالی که برآوردگرهای BRK و GEBRK دارای اریبی کم‌تری نسبت به برآوردگرهای UK و GEUK هستند. به نظر می‌رسد که دو برآوردگر BRK و GEBRK نسبت به دو برآوردگر UK و GEUK تمایل به واریانس و MSE بزرگ‌تر دارند، البته این اتفاق برای اندازه نمونه‌های کوچک رخ می‌دهد. به‌طور کلی به نظر می‌رسد که برآوردگرهای BRK و GEBRK در کاهش اریبی عملکرد بهتری نسبت به دو برآوردگر دیگر داشته باشند. در شکل ۵ الف و ب به کمک روش‌های عددی مستطیلی و سیمپسون، مقدار بهینه برای پارامتر a به‌دست آمد.

جدول ۱. اریبی، واریانس و MSE برآوردگرهای چگالی هسته‌ای مختلف با $n=20$

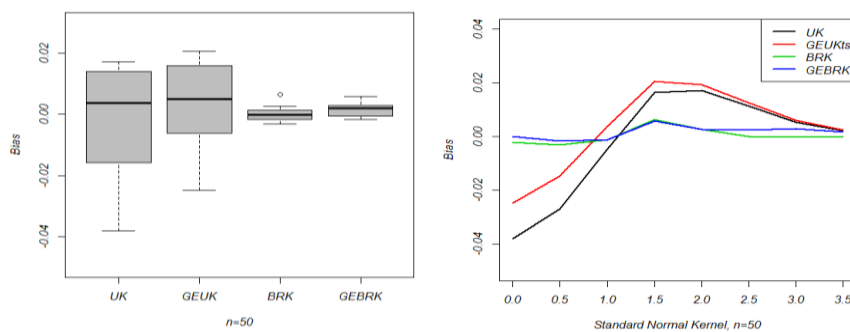
برآوردگرهای هسته‌ای				معیار	x
GEURK	BRK	GEUK	UK		
-0.0008744	-0.0043145	-0.0287971	-0.0490673	Bias	0
0.0131361	0.0119197	0.0049240	0.0046605	Variance	
0.0131368	0.0119384	0.0057533	0.0070636	MSE	
0.0009871	-0.0014806	-0.0170679	-0.0343345	Bias	0.5
0.0127337	0.0115978	0.0046782	0.0044488	Variance	
0.0127347	0.0116001	0.0049695	0.0056277	MSE	
0.0037656	0.0030493	0.0057541	-0.0044515	Bias	1
0.0097261	0.0089766	0.0036465	0.0034274	Variance	
0.0097403	0.0089859	0.0036796	0.0034473	MSE	
0.0039859	0.0037757	0.0194854	0.0155765	Bias	1.5
0.0058686	0.0056479	0.0025974	0.0024670	Variance	
0.0058845	0.0056622	0.0029771	0.0027096	MSE	
0.0029150	0.0003296	0.0216026	0.0200788	Bias	2
0.0023339	0.0025575	0.0015503	0.0015348	Variance	
0.0023424	0.0025576	0.0020170	0.0019380	MSE	
0.0061031	0.0009452	0.0157371	0.0154379	Bias	2.5
0.0007744	0.0009277	0.0005984	0.0006480	Variance	
0.0008117	0.0009286	0.0008460	0.0008863	MSE	
0.0043093	-0.0006175	0.0080055	0.0076499	Bias	3
0.0002209	0.0002514	0.0001744	0.0001997	Variance	
0.0002394	0.0002518	0.0002385	0.0002582	MSE	
0.0021371	-0.0003987	0.0037673	0.0034971	Bias	3.5
0.0000609	0.0000611	0.0000519	0.0000595	Variance	
0.0000655	0.0000612	0.0000661	0.0000717	MSE	



شکل ۱. نمودار خطی و جعبه‌ای از مقادیر اریبی برآوردگرهای چگالی هسته‌ای مختلف با $n=20$

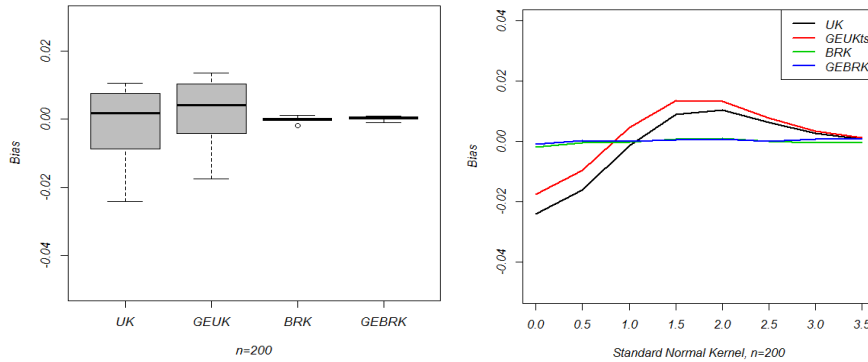
جدول ۲. اریبی، واریانس و MSE برآوردگرهای چگالی هسته‌ای مختلف با $n=50$

برآوردگرهای هسته‌ای				معیار	χ
GEBRK	BRK	GEUK	UK		
0.0001028	-0.0020778	-0.0247974	-0.0380254	Bias	0
0.0055016	0.0053727	0.0021368	0.0022895	Variance	
0.0055016	0.0053771	0.0027517	0.0037355	MSE	
-0.0017575	-0.0031931	-0.0147192	-0.0269493	Bias	0.5
0.0051192	0.0049926	0.0019592	0.0021085	Variance	
0.0051223	0.0050028	0.0021759	0.0028347	MSE	
-0.0011724	-0.0012606	0.0039369	-0.0044831	Bias	1
0.0041423	0.0040459	0.0016291	0.0017487	Variance	
0.0041437	0.0040475	0.0016446	0.0017688	MSE	
0.0058714	0.0064131	0.0205309	0.0163337	Bias	1.5
0.0027298	0.0026958	0.0012435	0.0013168	Variance	
0.0027643	0.0027369	0.0016650	0.0015836	MSE	
0.0025983	0.0026198	0.0193139	0.0172091	Bias	2
0.0011351	0.0011859	0.0006430	0.0006947	Variance	
0.0011418	0.0011928	0.0010160	0.0009094	MSE	
0.0023454	0.0000744	0.0125671	0.0114241	Bias	2.5
0.0003056	0.0003800	0.0002309	0.0002584	Variance	
0.0003111	0.0003800	0.0003888	0.0003889	MSE	
0.0030277	-0.0002167	0.0060476	0.0053992	Bias	3
0.0000884	0.0001143	0.0000726	0.0000893	Variance	
0.0000976	0.0001143	0.0001092	0.0001185	MSE	
0.0017904	-0.0000821	0.0023964	0.0020862	Bias	3.5
0.0000252	0.0000252	0.0000161	0.0000215	Variance	
0.0000284	0.0000271	0.0000218	0.0000258	MSE	

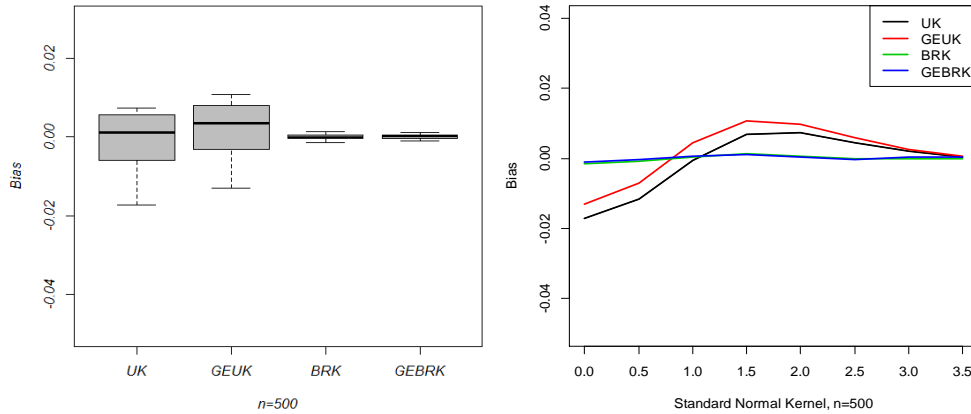
شکل ۲. نمودار خطی و جعبه‌ای از مقادیر اریبی برآوردگرهای چگالی هسته‌ای مختلف با $n=50$ جدول ۳. اریبی، واریانس و MSE برآوردگرهای چگالی هسته‌ای مختلف با $n=200$

برآوردگرهای هسته‌ای				معیار	χ
GEBRK	BRK	GEUK	UK		
-0.0010077	-0.0019429	-0.0175200	-0.0241311	Bias	0
0.0015771	0.0017283	0.0006608	0.0008785	Variance	
0.0015782	0.0017321	0.0009677	0.0014608	MSE	
0.0002561	-0.0003355	0.0096319	-0.0160644	Bias	0.5
0.0014281	0.0015437	0.0005986	0.0007851	Variance	
0.0014282	0.0015438	0.0006913	0.0010431	MSE	

-0.0000172	-0.0002388	0.0046548	-0.0014281	Bias	1
0.0010852	0.0011776	0.0004703	0.0006081	Variance	
0.0010852	0.0011777	0.0004919	0.0006103	MSE	
0.0004057	0.0006546	0.0135346	0.0088776	Bias	1.5
0.0007316	0.0007654	0.0003729	0.0004531	Variance	
0.0007318	0.0007658	0.0005561	0.0005319	MSE	
0.0006592	0.0010200	0.0131997	0.0104526	Bias	2
0.0003131	0.0003292	0.0001803	0.0002141	Variance	
0.0003135	0.0003303	0.0003545	0.0003233	MSE	
-0.0000297	-0.0000465	0.0076281	0.0062307	Bias	2.5
0.0001076	0.0001196	0.0000621	0.0000812	Variance	
0.0001076	0.0001196	0.0001203	0.0001201	MSE	
0.0007227	-0.0003725	0.0034179	0.0026957	Bias	3
0.0000223	0.0000297	0.0000153	0.0000213	Variance	
0.0000229	0.0000299	0.0000270	0.0000286	MSE	
0.0008959	-0.0002176	0.0011451	0.0008463	Bias	3.5
0.0000059	0.0000063	0.0000031	0.0000047	Variance	
0.0000067	0.0000063	0.0000044	0.0000054	MSE	



شکل ۳. نمودار خطی و جعبه‌ای از مقادیر اریبی برآوردگرهای چگالی هسته‌ای مختلف با $n=200$

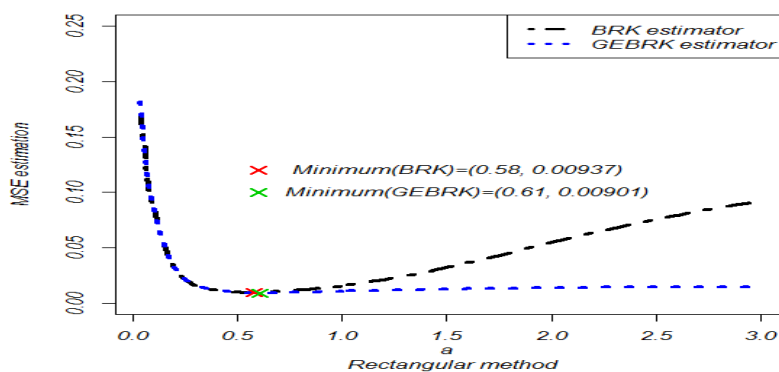


شکل ۴. نمودار خطی و جعبه‌ای از مقادیر اریبی برآوردگرهای چگالی هسته‌ای مختلف با $n=500$

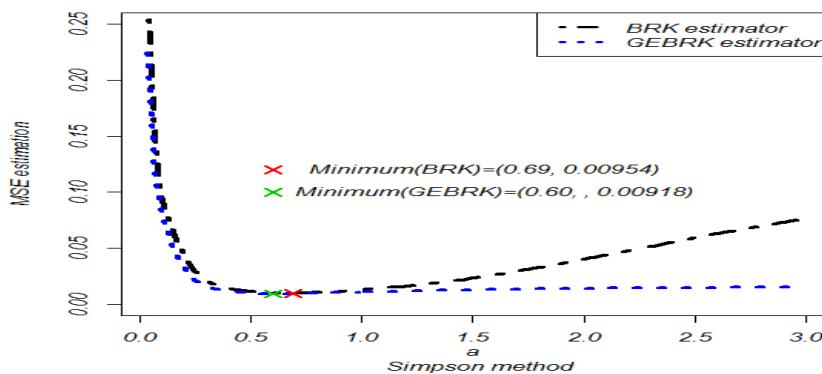
جدول ۴. اریبی، واریانس و MSE برآوردگرهای چگالی هسته‌ای مختلف با $n=500$

برآوردگرهای هسته‌ای				معیار	χ
GEURK	BRK	GEUK	UK		
-0.0010009	-0.0015517	-0.0130368	-0.0172833	Bias	۰
0.0006946	0.0008075	0.0003160	0.0004708	Variance	
0.0006956	0.0008099	0.0004860	0.0007695	MSE	

-0.0003381	-0.0007380	-0.0069776	-0.0115680	Bias	۰.۵
0.0006499	0.0007498	0.0003018	0.0004404	Variance	
0.0006500	0.0007504	0.0003505	0.0005742	MSE	
0.0006237	0.0003521	0.0043709	-0.0006894	Bias	۱
0.0004385	0.0005043	0.0002077	0.0002974	Variance	
0.0004389	0.0005044	0.0002268	0.0002977	MSE	
0.0010621	0.0012794	0.0106814	0.0068941	Bias	۱.۵
0.0003019	0.0003365	0.0001648	0.0002179	Variance	
0.0003031	0.0003381	0.0002790	0.0002654	MSE	
0.0003591	0.0006728	0.0098648	0.0073176	Bias	۲
0.0001351	0.0001512	0.0000770	0.0001025	Variance	
0.0001353	0.0001517	0.0001743	0.0001560	MSE	
-0.0002716	-0.0001899	0.0058700	0.0043550	Bias	۲.۵
0.0000448	0.0000489	0.0000252	0.0000347	Variance	
0.0000449	0.0000490	0.0000597	0.0000537	MSE	
0.0003488	-0.0000117	0.0025127	0.0019690	Bias	۳
0.0000118	0.0000139	0.0000063	0.0000101	Variance	
0.0000120	0.0000139	0.0000127	0.0000139	MSE	
0.0003943	-0.0002014	0.0006978	0.0004636	Bias	۳.۵
0.0000019	0.0000024	0.0000009	0.0000017	Variance	
0.0000020	0.0000024	0.0000014	0.0000019	MSE	



شکل ۵. آ) نمودار برآورد MSE به عنوان تابعی از پارامتر a با استفاده از روش مستطیلی.

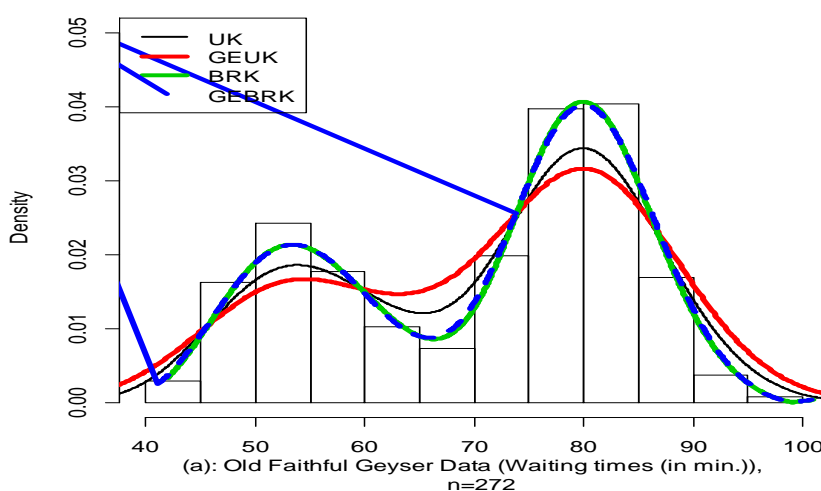


شکل ۵. ب) نمودار برآورد MSE به عنوان تابعی از پارامتر a با استفاده از روش سیمپسون

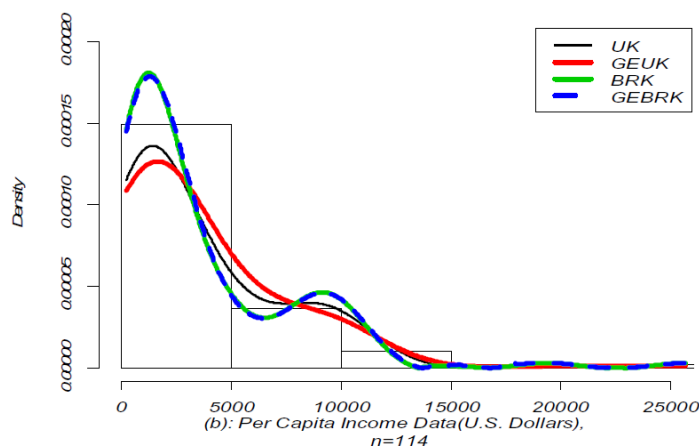
۶. داده‌های واقعی

در این بخش سه مجموعه از داده‌های واقعی را در نظر گرفته و به کمک برآوردهای ارائه شده، تابع چگالی احتمال این داده‌ها را برآورد می‌کنیم. مجموعه اول شامل ۲۷۲ داده مربوط به چشمه آب گرم است که فارل و آزالینی و باومن ذکر کرده‌اند [17]، [30]. دومین مجموعه شامل ۱۱۴ مقدار درآمد سرانه واقعی است که رومر بیان کرده است

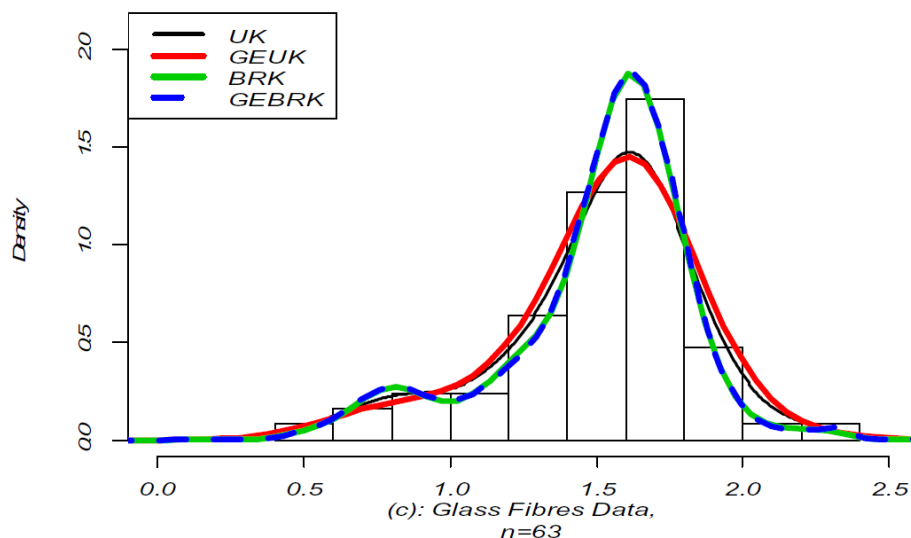
[31]. سومین مجموعه شامل ۶۳ مقدار اندازه‌گیری شده در یک آزمایشگاه فیزیک در شهر لندن در باره مقاومت الیاف شیشه‌های ۱/۵ سانتی‌متری است که یعقوب‌زاده و شادرخ بیان کرده‌اند [32]. داده‌های چشمه آب گرم برای دو متغیر زمان انتظار بین فوران‌ها و مدت زمان فوران بر حسب دقیقه در پارک ملی یلوستون اندازه‌گیری شده‌اند. بر اساس چهار برآوردگر UK، GEUK، BRK و GEBRK چگالی زمان انتظار فوران بعدی را بر اساس این داده‌ها برآورد می‌کنیم. نتایج در شکل ۶ نشان داده شده است. چنانچه از شکل ۶ الف ملاحظه می‌شود دو روش UK و GEUK مشابه هم عمل می‌کنند در حالی که دو روش BRK و GEBRK نیز مشابه هم و نسبت به دو روش قبلی دقت بیش‌تر و آریبی کم‌تری دارند. به‌طور مشابه نتایج دو مجموعه داده دیگر نیز در شکل ۶ ب و ج آمده است.



شکل ۶. آ نمودار بافت‌نگار داده‌های واقعی چشمه آب گرم و چگالی‌های برازش داده شده



شکل ۶. ب نمودار بافت‌نگار داده‌های واقعی درآمد سرانه واقعی و چگالی‌های برازش داده شده



شکل ۶. پ نمودار بافت‌نگار داده‌های واقعی مقاومت الیاف شیشه و چگالی‌های برازش داده شده

۷. نتایج

هدف اصلی این مقاله بهبود کاهش اریبی برآورد چگالی با معرفی برآوردگر چگالی هسته‌ای جدید است. در بخش ۲ برآوردگر چگالی متداول معرفی شد که اریبی آن از مرتبه $O(h^3)$ و MSE بهینه از مرتبه $O(n^{-4})$ است. در بخش ۳ برآوردگر هسته‌ای کاهش اریبی را پیشنهاد دادیم که اریبی آن از مرتبه $O(h^4)$ و MSE بهینه از مرتبه $O(n^{-8})$ است. در بخش ۴ نخست برآوردگری که ترل و اسکات ارائه کردند [28]، متذکر شدیم که اریبی آن از مرتبه $O(h^4)$ و MSE بهینه از مرتبه $O(n^{-8})$ است سپس برآوردگری که به‌طور چشم‌گیری باعث کاهش اریبی برآورد چگالی می‌شود یعنی برآوردگر هسته‌ای کاهش اریبی برونمایی هندسی را پیشنهاد دادیم. خواص نظری از جمله اریبی، واریانس و MSE این برآوردگر بررسی و ثابت شد که اریبی آن از مرتبه $O(h^{13})$ و MSE بهینه از مرتبه $O(n^{-16})$ است. به‌نظر می‌رسد که برآوردگر BRK نسبت به برآوردگر GEUK اریبی را بهبود نمی‌بخشد. اما نتایج جدول‌های ۱ تا ۴ و شکل‌های ۱ تا ۴ نشان می‌دهد که اریبی بهینه شده در برآوردگر BRK نسبت به برآوردگرهای UK و GEUK کمتر است. در حالی که برآوردگر GEBRK نسبت به برآوردگر GEUK دارای نرخ همگرایی سریع‌تری در اریبی و MSE است. به‌طور کلی نتایج جدول‌های ۱ تا ۴ و شکل‌های ۱ تا ۴ و ۶ نشان می‌دهند که دو برآوردگر BRK و GEBRK نسبت به برآوردگرهای UK و GEUK دقت بیش‌تری دارند و در کاهش اریبی مؤثر بوده‌اند.

تقدیر و تشکر

از زحمات سردبیر و هم‌چنین داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود مقاله شد، تشکر و قدردانی می‌کنیم.

منابع

1. Rosenblatt M., "Remarks on some nonparametric estimates of a density function", Ann. Math. Stat., 27(3) (1956) 832-837.

2. Scott D. W., "Density estimation theory, practice and visualization", first ed., Wiley, USA (1992).
3. Silverman B. W., "Density estimation for statistics and data analysis", first ed., Chapman and Hall, USA (1986).
4. Hardle W., "Smoothing Techniques: with Implementation in S", Springer-Verlag Newyork, (1991).
5. Hwang J.-N, Lay S.-R., Lippman A., "Nonparametric multivariate density estimation: a comparative", (1994).
6. Joyner T. A., Rohli R., "Kernel density estimation of tropical cyclone frequencies in the North Atlantic basin", *Int. J. Geosci.*, 1(3) (2010) 121-129.
7. Hurter C., Ersoy O., Telea A., "Graph bundling by kernel density estimation", *Comp. Graph. Forum*, 31(3) (2012) 865-874.
8. Joshi N., Brady M., "Simplified computations for nonparametric windows method for probability density function estimation", *IEEE Trans. Pattern Anal. Mach. Intell.*, 33 (8) (2011) 1673-1680.
9. Cai Q., Rushton G., Bhaduri B., "Tests of an improved kernel density estimation method for identifying disease clusters", *J. Geograph. Syst.*, 14(3) (2011) 243-264.
10. Faucher D., Rasmussen P. F., Bobee B., "Distribution based bandwidth selection method for kernel quantile estimation", *J. Hydrol.*, 250 (1-4) (2001) 1-11.
11. Bors A. G., Nasios N., "Kernel bandwidth estimation for nonparametric modeling", *IEEE Trans. Syst., Man, Cybernet.-Part B: Cybernet.*, 39(6) (2009) 1543-1555.
12. Shimazaki H., Shinomoto S., "Kernel bandwidth optimization in spike rate estimation", *J. Comput. Neurosci.*, 29(1-2) (2010) 171-182.
13. Wang H., Mirota D., Hager G.D., "A generalized Kernel consensusbased robust estimator", *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1) (2010) 178-184.
14. Baxter M. J., Beardah C. C., Wright R. V. S., "Some archaeological applications of kernel density estimates", *J. Archaeol. Appl. Kernel Density Est.*, 24 (4) (1997) 347-354.
15. Chen T.-B., Lu H. H.-S, Lee Y.-S, Lan H.-J., "Segmentation of cDNA microarray images by kernel density estimation", *J. Biomed. Inform.* 41(6) (2008) 1021-1027.
16. Parzen E., "On estimation of a probability density function and mode", *Ann. Math. Stat.*, 33(3) (1962) 1065-1076.
17. Farrell R. H., "On the best obtainable asymptotic rates of convergence in estimation of a density function at a point", *The Annals of Mathematics and Statistics*, 43 (1972) 170-180.

18. Samiuddin M., El-Sayyad G. M., "On nonparametric kernel density estimates", *Biometrika*, 77 (1990) 865-874.
 19. El-Sayyad G. M., Samiuddin M., Abdel-Ghaly A. A., "A new kernel density estimate", *Journal of Nonparametric Statistics*, 3 (1992) 1-11.
 20. Marron J. S., Ruppert D., "Transformations to reduce boundary bias in kernel density estimation", *Journal of the Royal Statistical Society: Series B*, 4 (1992) 653-671.
 21. Kim C., Kim W., Park B. U., "Skewing and generalized jackknifing in kernel density estimation", *Communications in Statistics: Theory and Methods*, 32 (2003) 2153-2162.
 22. Mynbaev K., Martins-Filho, C., "Bias reduction in kernel density estimation via Lipschitz Condition", *Journal of Nonparametric Statistics*, 22 (2010) 219-235.
 23. Kairat T. Mynbaev, Saralees Nadarajah, Christopher S. Withers, Aziza S. Aipenova, "Improving bias in kernel density estimation", *Statistics and Probability Letters*, 94 (2014) 106-112.
 24. Kim J., Kim C., "Reducing the mean squared error in kernel density estimation", *Journal of the Korean Statistical Society*, 42 (2013) 387-397.
 25. Igarashi G., Kakizawa Y., "Bias corrections for some asymmetric kernel estimators", *Journal of Statistical Planning and Inference*, 159 (2015) 37-63.
 26. Xiaoran X., Jingjing W., "Some Improvement on Convergence Rates of Kernel Density Estimator", *Applied Mathematics*, 5 (2014) 1684-1696.
 27. Wand M. P., Jones M. C., "Kernel smoothing", Chapman Hall, London, (1995).
 28. Terrell G. R., Scott D. W., "On improving convergence rates for nonnegative kernel density estimators", *The Annals of Statistics*, 8 (1980) 1160-1163.
 29. Burden, Richard L., Faires, J. Douglas, "Numerical analysis", 9th ed., Brooks/Cole, (2010).
 30. Azzalini A., Bowman A. W., "A look at some data on the Old Faithful geyser", *Applied Statistics*, 39 (1990) 357-365.
 31. Romer D., "Openness and ination: theory and evidence", *Quart. J. Econ.*, 108 (1993) 869-903.
۳۲. یعقوب‌زاده شهرستان ش.، شادرخ ع.، "مقایسه دو زیر مجموعه از خانواده توزیع‌های بتا G با دو زیر مجموعه از خانواده توزیع‌های زگرافوس-بالاکریشنان G به کمک روش شبیه‌سازی مونت‌کارلو"، مدل‌سازی پیشرفته ریاضی، دوره ۶، شماره ۱ (۱۳۹۵) ۴۱-۵۹.