[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]

مقایسهٔ کوچک نمونه‌ای برآورد تابع چگالی با روش‌های هستهٔ گاما و سری‌های متعامد

محی‌الدین ایزدی*، عبداله جلیلیان
دانشگاه رازی، گروه آمار

دریافت ۹۶/۰۷/۲۲ پذیرش ۹۷/۱۲/۱۱

چکیده

برآوردگر هسته‌ای استاندارد دچار مشکل اریبی مرزی برای تابع‌های چگالی احتمال توزیع‌هایی با تکیه‌گاه اعداد حقیقی مثبت است. برآوردگرهای هسته‌ای گاما و سری‌های متعامد دو جای‌گزین برای برآوردگر هسته‌ای استاندارد هستند که مبرا از اریبی مرزی‌اند. در این مقاله یک بررسی شبیه‌سازی به‌منظور مقایسهٔ عملکرد کوچک‌نمونه‌ای برآوردگرهای هسته‌ای گاما و برآوردگر سری‌های متعامد با پایه‌های لاگِر و اِرمیت اجرا شده است. بر اساس این مطالعهٔ شبیه‌سازی دریافتیم که اگر انتخاب پایه برای برآوردگر سری‌های متعامد بر اساس شناخت جزئی از شکل تابع چگالی هدف باشد، برآوردگر سری‌های متعامد می‌تواند عملکردی بهتر از برآوردگر هسته‌ای گاما داشته باشد.

واژه‌های کلیدی: اریبی مرزی، میانگین جمع‌بستهٔ توان دوم خطا، پارامتر هموارساز.

کد موضوع‌بندی ریاضی (۲۰۱۰): **62G07، 62Gxx، 93E14**

مقدمه

برآورد ناپارامتری تابع چگالی احتمال بر اساس یک نمونهٔ تصادفی از آن یکی از پویاترین زمینه‌های پژوهشی آمار در چند دههٔ اخیر بوده است. مقاله‌های مروری و کتاب‌های متعددی در مورد روش‌های موجود و سیر تکوینی آن‌ها بحث کرده‌اند (تاپیا و تامسون [۱]، سیلورمن [۲]، آیزن‌من [۳]، واند و جونز [۴] و دوروی و لوگوسی [۵]). فرض کنید X_1، ...، X_n نمونه‌ای تصادفی از یک توزیع احتمال با تابع چگالی نامعلوم $f: R \to [0,\infty)$ باشد. یکی از نام‌آشناترین برآوردگرهای تابع چگالی f در نقطهٔ $x \in R$، طبق رابطهٔ (۱) است:

$$\hat{f}(x;\delta) = \frac{1}{n\delta} \sum_{i=1}^{n} K\left(\frac{x-X_i}{\delta}\right) \qquad (۱)$$

که در آن K تابع هسته و $\delta > 0$ پارامتر هموارساز یا پهنای نوار[1] نامیده می‌شود. تابع هسته اغلب یک تابع چگالی احتمال متقارن $K(-u) = K(u)$ با ویژگی‌های

$$\int_R K(u)\, du = 1, \quad \int_R uK(u)\, du = 0, \quad \int_R u^2 K(u)\, du < \infty,$$

است. هم‌چنین فرض می‌شود که پارامتر هموارساز $\delta = \delta_n$ نیز هنگامی که $n \to \infty$، در شرط‌های $\delta \to 0$ و $n\delta \to \infty$ صدق می‌کند. برآوردگر $\hat{f}(x;\delta)$، به برآوردگر هسته‌ای استاندارد و یا برآوردگر پارزن–روزنبلت معروف است. نشان داده شده است که با انتخاب مناسب پارامتر هموارساز، برآوردگر هسته‌ای استاندارد سازگار بوده است و میانگین توان دوم خطای آن و هم‌چنین میانگین جمع‌بستهٔ توان دوم خطا[2]،

*نویسنده مسئول m.izadi@razi.ac.ir

1. Bandwidth
2. Mean Integrated Squared Error

$$MISE(\hat{f}) = E\int_R \left[f - \hat{f}(x;\delta)\right]^2 \, dx, \qquad (۲)$$

با نرخ رشد $O(n^{-\frac{4}{5}})$ به صفر همگرا است (پارزن [6]). سیلورمن [۲] و واند و جونز [۴] مروری جامع بر عملکرد و ویژگی‌های برآوردگر هسته‌ای استاندارد تابع چگالی داشته‌اند که خوانندگان علاقه‌مند می‌توانند به این مراجع رجوع کنند.

با وجود این هنگامی که تکیه‌گاه تابع چگالی بازه‌ای کران‌دار مانند $[a,b]$ باشد، برآوردگر (۱) عملکرد مناسبی در نقاط نزدیک به مرزها که ناحیهٔ مرزی گفته می‌شود، ندارد. در واقع این برآوردگر در نقاط مرزی دارای اریبی است که به مسئلهٔ اریبی مرزی[۱] معروف است. از آن‌جاکه در بسیاری از کاربردها، کمیت‌های تصادفی مورد بررسی نامنفی هستند، استفاده از برآوردگر هسته‌ای استاندارد مناسب نیست، زیرا این برآوردگر در نقاط نزدیک به صفر دارای مسئلهٔ اریبی مرزی است. به‌طور مثال، فرض کنید f یک تابع چگالی حتمال با تکیه‌گاه $[0,\infty)$ باشد که مشتق دوم آن موجود و در یک همسایگی $x=0$ پیوسته است. در این صورت برای تابع هستهٔ متقارن K با تکیه‌گاه $[-1,1]$ و هر $x \in [0,\delta)$ (مارون و راپرت [۷])،

$$\begin{aligned} Bias\,\{\hat{f}(x;\delta) &= E\hat{f}(x;\delta) - f(x) \\ &= -f(x)\int_{\frac{x}{\delta}}^{1} K(u)du - \delta\, f'(x)\int_{-1}^{\frac{x}{\delta}} uK(u)du \\ &\quad + \frac{\delta^2}{2} f''(x)\int_{-1}^{\frac{x}{\delta}} u^2K(u)\,du + o(\delta^2)\,. \end{aligned}$$

بنابراین واضح است که برای $x \in [0,\delta)$ برآورگر هسته‌ای استاندارد دارای اریبی است. به‌عنوان مثال،

$$Bias\,\hat{f}(0;\delta) \simeq \frac{1}{2}\, f(0).$$

برای رهایی از این مشکل راه‌کارهای متعددی پیشنهاد شده است که از جمله می‌توان به‌روش بازتاب داده‌ها (شوستر [۸]، سیلورمن [۲] و کلاین و هارت [۹])، روش هسته‌های کران‌دار (گسر و همکاران [۱۰]، مولر [۱۱]، [۱۲] و جونز [۱۳]) و روش تبدیل داده‌ها (وند و همکاران [۱۴] و مارون و راپرت [۷]) اشاره کرد. یکی از روش‌های بسیار موثر در این زمینه استفاده از هسته‌های نامتقارن است. چن [۱۵] استفاده از هستهٔ بتا را برای برآورد توابع چگالی با تکیه‌گاه $[0,1]$ پیشنهاد و بررسی کرده است. چن [۱۶] برای برآورد تابع‌های چگالی با تکیه‌گاه $[0,\infty)$، دو برآوردگر با هستهٔ گاما معرفی کرده و نشان داده است که این برآوردگرها مبرا از مسئلهٔ اریبی مرزی است. هم‌چنین وی با استفاده از بررسی‌های شبیه‌سازی نشان داده است که این برآوردگرها دارای عملکرد بهتری نسبت به برآوردگرهای معرفی شده به‌وسیلهٔ جونز [۱۳]، جونز و فوستر [۱۷] و مولر و وانگ [۱۸] است. مالک و شاینلی [۱۹] نیز به بررسی مقایسهٔ عملکرد برآوردگرهای گاما با برآوردگرهای ارائه شده شوستر [۸]، کارونامونی و آلبرت [۲۰]، گسر و مولر [۲۱] و جونز و فوستر [۱۷] پرداخته‌اند. آن‌ها با استفاده از شبیه‌سازی، برای حالت‌های مختلفی از تابع چگالی، نشان داده‌اند که در حالت کلی برآوردگرهای گاما به‌ویژه برآوردگر نوع اول نسبت به برآوردگرهای دیگر دارای عملکرد مناسب‌تر و یا رقابت‌پذیرند.

اما برآوردگرهای هسته‌ای تنها برآودگرهای پیشنهاد شده برای برآورد ناپارامتری تابع چگالی نیستند. برآودگرهای سری‌های متعامد[۲] دستهٔ دیگری از برآوردگرها هستند که نخستین بار چنکوف [۲۰] معرفی کرد و پس از آن سایر

1. Boundary bias
2. Orthogonal series estimators

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]

پژوهش‌گران بررسی کردند. در این برآوردگرها، ضرایب هموارسازی در نظر گرفته می‌شوند که همانند پهنای نوار در روش برآورد هسته‌ای بر عملکرد برآوردگردها تأثیرگذارند و در عمل باید به‌صورت بهینه انتخاب شوند [۲۳]. بررسی‌های بزرگ نمونه‌ای نشان داده‌اند که برآودگرهای سری‌های متعامد نیز مانند برآوردگرهای هسته‌ای خواص مجانبی مطلوبی دارند [۲۴]. مروری جامع بر این نوع برآوردگرها را اِفروموویچ [۲۵] ارائه کرده است.

در این مقاله، با در نظر گرفتن چندین تابع چگالی خاص، رفتار کوچک نمونه‌ای برآوردگرهای هسته‌ای مبتنی بر هستهٔ گاما و برآوردگرهای سری‌های متعامد را بررسی و عملکرد آنها را در قالب یک بررسی شبیه‌سازی با یک‌دیگر مقایسه می‌کنیم. برای این منظور، برآوردگرهای هسته‌ای با هستهٔ گاما در بخش بعد و برآوردگرهای سری‌های متعامد در بخش سوم معرفی می‌شوند. سپس در بخش پایانی عملکرد این برآوردگرها برای چند تابع چگالی معین بررسی و مقایسه می‌شود.

برآورد با هستهٔ گاما

فرض کنید f یک تابع چگالی احتمال با تکیه‌گاه $[0,\infty)$ و X_1، ... ، X_n نمونه‌ای تصادفی از f باشد. چنان‌که گفته شد، استفاده از برآوردگر هسته‌ای استاندارد در همسایگی نقطهٔ صفر دارای اریبی است. برای رفع این مشکل چِن[16] دو برآوردگر هسته‌ای که در آن‌ها از تابع چگالی گاما به‌عنوان تابع هسته استفاده شده است را معرفی کرد. فرض کنید

$$K(u;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\,u^{\alpha-1}\,exp\left(-\frac{u}{\beta}\right),\ u\geq 0,$$

تابع چگالی توزیع گاما با پارامتر شکل $\alpha>۰$ و پارامتر مقیاس $\beta>۰$ باشد. برآوردگر

$$\hat{f}_{g_1}(x;\omega)=\frac{1}{n}\ \sum_{i=1}^{n}\ K\left(X_i;\frac{x}{\omega}+1,\omega\right),\ x\geq 0, \tag{۳}$$

نخستین کلاس از برآوردگرهای گاما است که چِن [۱۶] معرفی کرده است. در برابری (۳)، $\omega=\omega_n$ دنباله‌ای از اعداد مثبت است که پارامتر هموارساز نامیده می‌شود و در شرط‌های $\omega\to 0$ و $n\omega\to\infty$ هنگامی که $n\to\infty$ صدق می‌کند. چِن نشان داد

$$E\left[\hat{f}_{g_1}(x;\omega)\right]=\ f(x)\ +\ \omega\left\{f'(x)+\frac{x}{2}f''(x)\right\}\ +o(\omega) \tag{۴}$$

و

$$Var\left[\hat{f}_{g_1}(x;\omega)\right]\sim\begin{cases}\dfrac{1}{2\sqrt{\pi}}n^{-1}\,\omega^{-\frac{1}{2}}\,x^{-1/2}\,f(x), & \dfrac{x}{\omega}\to\ \infty\\ \dfrac{\Gamma(2\kappa+1)}{2^{1+2\kappa}}\Gamma^2(\kappa+1)n^{-1}\omega^{-1}f(x), & \dfrac{x}{\omega}\ \to\kappa,\end{cases}$$

که در آن f' و f'' به‌ترتیب مشتق‌های اول و دوم f است. چون اریبی برآوردگر $\hat{f}_{g_1}(.;\omega)$ از مرتبهٔ $O(\omega)$ است، بنابراین این برآوردگر مبرا از مسئلهٔ اریبی مرزی برآوردگرهای استاندارد است. اما وجود f' در (۴) مطلوب نیست، از این رو چِن [۱۶] کلاس دیگری از برآوردگرهای هسته‌ای گاما را به‌صورت (۵) معرفی کرد:

$$\hat{f}_{g_2}(x;\omega)=\frac{1}{n}\sum_{i=1}^{n}K(X_i;\rho_\omega(x),\omega),\ x\geq 0 \tag{۵}$$

که در آن

$$\rho_\omega(x)=\begin{cases}\dfrac{x}{\omega}, & x\geq\ 2\omega,\\ \left(\dfrac{x}{2\omega}\right)^2+1, & x\in\ [0,2\omega).\end{cases}$$

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]

می‌توان نشان داد

$$Bias\{\hat{f}_{g_2}(x;\omega)\} = \begin{cases} \frac{1}{2}xf''(x)\omega + o(\omega), & x \geq 2\omega \\ \xi_\omega(x)\omega f'(x) + o(\omega), & x \in [0, 2\omega) \end{cases} \qquad (۶)$$

که در آن

$$\xi_\omega(x) = (1-x)\frac{\rho_\omega(x) - \frac{x}{\omega}}{1 + \omega\rho_\omega(x) - x}.$$

با توجه به برابری (۶)، مشاهده می‌شود که f' تنها به‌ازای xهایی که در بازهٔ کوچکی در نزدیکی صفراند در میزان اریبی برآوردگر $\hat{f}_{g_2}(.;\omega)$ تأثیرگذار است.

برآورد با سری‌های متعامد

فرض کنید $\mathrm{L}^2 = \mathrm{L}^2(0,\infty)$ مجموعهٔ همهٔ تابع‌های $h: [0,\infty) \to R$ باشد به گونه‌ای که

$$\int_0^\infty h^2(x)\,d\,x < \infty.$$

در این صورت L^2 یک فضای هیلبرت با ضرب داخلی

$$\langle h, g\rangle = \int_0^\infty h(x)g(x)d\,x, \quad h, g \in L^2$$

و نرم

$$||h|| = \sqrt{\langle h, h\rangle} = \left(\int_0^\infty h^2(x)d\,x\right)^{\frac{1}{2}}, \qquad h \in L^2$$

است [۲۶]. اگر گردایهٔ $\{\phi_j: j \in \mathbb{N}_0\} \subset L^2$ یک پایهٔ یکامتعامد برای L^2 باشد، آن‌گاه هر $h \in L^2$ دارای بسط یکتای

$$h(x) = \sum_{i=0}^\infty c_j\,\phi_j(x)$$

برحسب ϕ_jها است که در آن

$$c_j = \langle h, \phi_j\rangle = \int_0^\infty h(x)\phi_j(x)dx.$$

با فرض این که $f \in L^2$، تابع چگالی f بر حسب پایهٔ $\{\phi_j: j \in \mathbb{N}_0\}$ دارای بسط

$$f(x) = \sum_{j=0}^\infty \theta_j\,\phi_j(x)$$

با ضرایب

$$\theta_j = \langle h, \phi_j\rangle = \int_0^\infty f(x)\phi_j(x)d\,x = E[\phi_j(X)]$$

است. با توجه به این که $\theta_j = E[\phi_j(X)]$،

$$\hat{\theta}_j = \frac{1}{n}\sum_{i=1}^n \phi_j(X_i)$$

یک برآوردگر نااریب و سازگار برای θ_j و در نتیجه

$$\hat{f}_\infty(x) = \sum_{j=0}^\infty \hat{\theta}_j\,\phi_j(x),$$

یک برآوردگر ناریب برای $f(x)$ است (واتسون [۲۷]). با این حال $\hat{f}_{\infty}(x)$ واریانسی نامتناهی دارد و محاسبهٔ آن در عمل امکان‌پذیر نیست (وَهبا [۲۳]). برای رفع مشکل نامتناهی بودن واریانس $\hat{f}_{\infty}$ می‌توان ضرایب $\hat{\theta}_j$ را با وزن‌های b_j منقبض و از این برآوردگر استفاده کرد [۲۴]:

$$\hat{f}(x;\boldsymbol{b}) = \sum_{j=0}^{\infty} b_j \ \hat{\theta}_j \ \phi_j(x).$$

در برآوردگر $\hat{f}(.;\boldsymbol{b})$ بردار $\boldsymbol{b} = (b_0, b_1, b_2, \dots)$ را بردار هموارساز گویند و فرض می‌شود $0 \le \mathrm{b_j} \le 1$ و $\sum_{j=0}^{\infty} b_j < \infty$. یک انتخاب رایج برای $\mathbf{b}$،

$$b_j = \begin{cases} 1, & j \le J \\ 0, & j > J \end{cases} \quad (۷)$$

است که منجر به برآودگر بریده شدهٔ[1] (۸) می‌شود:

$$\hat{f}_o(x;J) = \sum_{j=0}^{J} \hat{\theta}_j \ \phi_j(x). \quad (۸)$$

وَهبا [۲۳] استفاده از ضرایب هموارساز پارامتری

$$b_j = b_j(c_1, c_2) = \frac{1}{\{1 + c_1 j^{c_2}\}}$$

را پیشنهاد کرد که در آن $c_1 > ۰$ و $c_۲ > ۱$ مجهول هستند.

بدیهی است که بردار هموارساز $\mathbf{b}$ بر عملکرد برآوردگر $\hat{f}_o(.;\boldsymbol{b})$ تأثیرگذار است و باید به‌روش بهینه‌ای انتخاب شود. در حالت کلی میانگین جمع‌بستهٔ توان دوم خطا (۲) برای $\hat{f}_o(.;\boldsymbol{b})$ برابر است با:

$$MISE\left(\hat{f}_o(.;\boldsymbol{b})\right) = \sum_{j=0}^{\infty} b_j^2 Var(\hat{\theta}_j) + \sum_{j=0}^{\infty} (b_j - 1)^2 \theta_j^2.$$

اما

$$E\left[(\hat{\theta}_j)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^{n} \phi_j(X_i)\right)^2\right] = = \frac{1}{n} E[\phi_j^2(X)] + \theta_j^2 \left(1 - \frac{1}{n}\right)$$

و

$$Var(\hat{\theta}_j) = E\left[(\hat{\theta}_j)^2\right] - \left[E(\theta_j)\right]^2 = \frac{1}{n}\left(E[\phi_j^2(X)] - \theta_j^2\right).$$

بنابراین

$$\mathrm{MISE}(\hat{f}_0(.;\mathbf{b})) = \sum_{j=0}^{\infty} \left[\frac{\mathrm{b_j^2}}{n}\left(E[\phi_j^2(X)] - \theta_j^2\right) + \theta_j^2 \left(b_j^2 - 2b_j\right)\right] + \sum_{j=0}^{\infty} \theta_j^2.$$

با توجه به این که (برابری پارسوال[2])

$$\sum_{j=0}^{\infty} \theta_j^2 = \int_R f^2(x) dx$$

به بردار هموارساز $\boldsymbol{b}$ بستگی ندارد، کمینه‌سازی $MISE(\hat{f}_0(.;\mathbf{b}))$ نسبت به $\boldsymbol{b}$ با کمینه‌سازی

$$I(\boldsymbol{b}) = MISE(\hat{f}_o(.;\boldsymbol{b})) - \int_R f(x)\, d\, x = \sum_{j=0}^{\infty} \left[\frac{b_j^2}{n}\left(E[\phi_j^2(X)] - \theta_j^2\right) + \theta_j^2 \left(b_j^2 - 2b_j\right)\right]$$

1. Truncated estimator
2. Parseval's identity

معادل است [۲۸]. بنابراین انتخاب بهینهٔ $\boldsymbol{b}$ با کمینه‌سازی $I(\boldsymbol{b})$ تعیین می‌شود. به‌عنوان مثال، انتخاب بهینهٔ مقدار برینشی[۱] $J = J_n$ در بردار هموارساز (۷) با کمینه‌سازی

$$I(J) = \sum_{j=0}^{J} \left[\frac{1}{n}\left(E\left[\phi_j^2(X)\right] - \theta_j^2\right) - \theta_j^2\right] \quad (۹)$$

نسبت به $J = 0,1,2, \dots$ به‌دست می‌آید. در عمل مقدار $\theta j2$ مجهول است که می‌توان آن را با برآوردگر نااریب

$$\widehat{\theta^2}_j = \frac{1}{n(n-1)} \sum_{1\le i<l\le n} \phi_j(X_i)\phi_j(X_l)$$

برآورد کرد. هم‌چنین با استفاده از

$$D_j = \frac{1}{n}\sum_{i=1}^{n} \phi_j^2(X_i)$$

به‌عنوان برآورد $E\left[\,\phi_j^2(X)\,\right]$، با کمینه‌سازی

$$\hat{I}(\boldsymbol{b}) = \sum_{j=0}^{\infty} \left[\frac{b_j^2}{n}\left(D_j - \widehat{\theta^2}_j\right) + \widehat{\theta^2}_j\left(b_j^2 - 2b_j\right)\right]$$

برآوردی برای بردار هموارساز $\boldsymbol{b}_n$ به‌دست می‌آید [۲۴].

ویژگی‌های بزرگ نمونه‌ای $\boldsymbol{b}_n$ و $\hat{f}_o(.\,;\boldsymbol{b}_n)$ با در نظر گرفتن برخی شرایط برای تابع چگالی f بررسی شده‌اند [۲۸]. به‌عنوان مثال، برای بردار هموارساز (۷)، اگر برای مقدار برینشی $J = J_n$ داشته باشیم $J \to \infty$ و $\frac{J}{n} \to 0$ هرگاه $n \to \infty$ و به‌علاوه J از مرتبهٔ $O(n^{\frac{1}{q}})$، $q \ge 2$، باشد، آن‌گاه میانگین جمع‌بستهٔ توان دوم خطای $\hat{f}_o(.\,;\boldsymbol{b})$، با نرخ رشد $O(n^{-\left(1-\frac{1}{q}\right)})$ به صفر همگرا است [۲۹].

انتخاب پایهٔ یکامتعامد $\{\phi_j : j \in \mathbb{N}_0\}$ به تکیه‌گاه f بستگی دارد. برای تابع‌های چگالی با تکیه‌گاه فشرده مانند بازهٔ $[0,a]$ اغلب از پایهٔ مثلثاتی

$$\phi_j(x) = \begin{cases} \frac{1}{\sqrt{a}}, & j = 0 \\ \sqrt{\frac{2}{a}}\cos\left(\frac{j\pi x}{a}\right), & j \ge 1 \end{cases}$$

استفاده می‌شود، در حالی که برای تابع‌های چگالی با تکیه‌گاه R، پایهٔ ارمیت نرمال شدهٔ[۲]

$$\phi_j(x) = \frac{1}{\sqrt{2^j j!\sqrt{\pi}}} e^{\frac{x^2}{2}} H_j(x)$$

که در آن

$$H_j(x) = (-1)^j e^{x^2}\left(\frac{d^j}{dx^j} e^{-x^2}\right), \; j \ge 0,$$

چندجمله‌ای ارمیت مرتبهٔ j-ام است و برای تابع‌های چگالی با تکیه‌گاه $[0,\infty)$، پایهٔ لاگِر نرمال شدهٔ[۳]

$$\phi_j(x) = e^{-\frac{x}{2}} L_j(x)$$

که در آن

$$L_j(x) = \frac{1}{j!} e^x \left[\frac{d^j}{dx^j}\left(e^{-x}x^j\right)\right] = \sum_{k=0}^{j} \binom{j}{k} \frac{(-1)^k}{k!} x^k, \; j \ge 0,$$

1. Cutoff value
2. Normalized Hermite basis
3. Normalized Laguerre basis

چندجمله‌ای لاگِر مرتبهٔ j-ام است، در نظر گرفته می‌شود [۳]. شکل ۱ نمودار ۱۰ تابع نخست دو پایهٔ لاگِر و اِرمیت نرمال شده را نشان می‌دهد. علاوه بر پایه‌های یکامتعامد رایج، از موجک‌ها[۱] نیز می‌توان استفاده کرد. در این حالت برآوردگر حاصل اغلب برآوردگر موجکی تابع چگالی احتمال نامیده می‌شود (ر.ک دِوروی و لوگوسی [۵] فصل ۱۳).

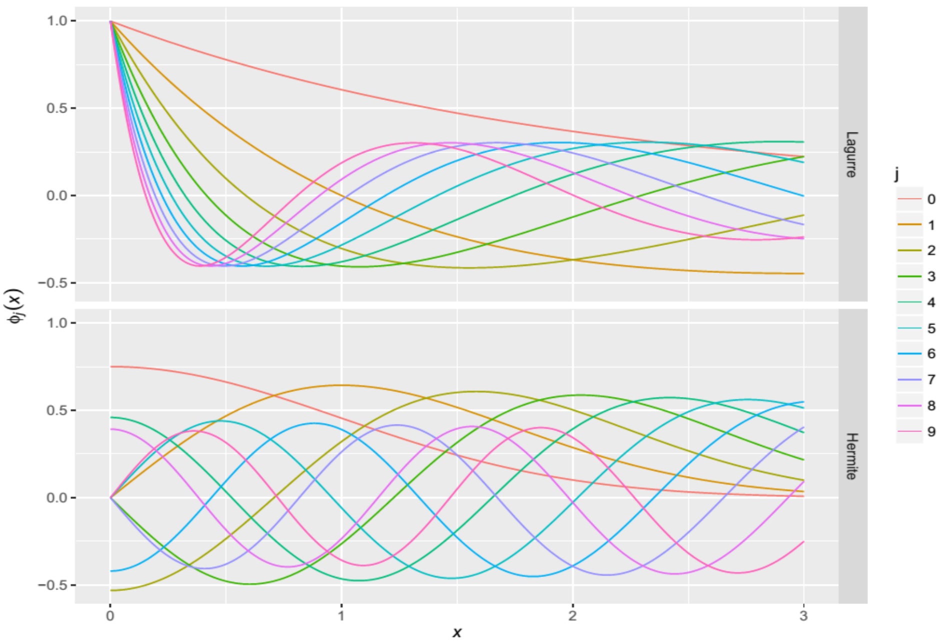
چنان‌که در بخش ۱ بیان شد، برآوردگرهای هسته‌ای با هسته‌های استاندارد مانند نرمال، اپانِچنیکوف و غیره، به‌عنوان برآورد توابع چگالی با تکیه‌گاه $[0,\infty)$، دارای مشکل اریبی (مجانبی) مرزی هستند و استفاده از برآوردگرهای هسته‌ای گاما و سری‌های متعامد دو راه برای رهایی از این مشکل می‌باشند. برای تبیین بهتر این موضوع، در شکل ۲، تابع چگالی توزیع نمایی با پارامتر نرخ $\lambda = 0.2$ و برآوردهای هسته‌ای با هسته‌های اپانِچنیکوف و گامای نوع اول و هم‌چنین برآورد با استفاده از سری‌های متعامد در بازهٔ $[0,1]$ رسم شده است. برای به‌دست آوردن برآوردها، از یک نمونه ۱۰۰۰ تایی از توزیع نمایی با پارامتر $\lambda = 0.2$ استفاده شده است. با توجه به شکل، مشکل اریبی مرزی برای برآوردگر با هستهٔ اپانِچنیکوف و برطرف کردن این مشکل با استفاده از دو برآوردگر دیگر مشهود است.

مقایسهٔ برآوردگرها

در این بخش، با استفاده از بررسی شبیه‌سازی، برآوردگرهای هسته‌ای گامای (۳) و (۵) با برآوردگر سری‌های متعامد (۸) برای انواع مختلفی از توابع چگالی با تکیه‌گاه $[0,\infty)$ و با شکل‌های گوناگون در همسایگی صفر مقایسه می‌شوند.

مالک و شاینلی [۱۹] از خانوادهٔ توزیع‌های تعمیم یافته‌ی F برای مقایسهٔ برآوردگرهای گاما با برآوردگرهای ذکرشده در بخش مقدمه استفاده کردند. این خانوادهٔ چهار پارامتری از توزیع‌ها دارای تابع چگالی احتمال (۱۰) است:

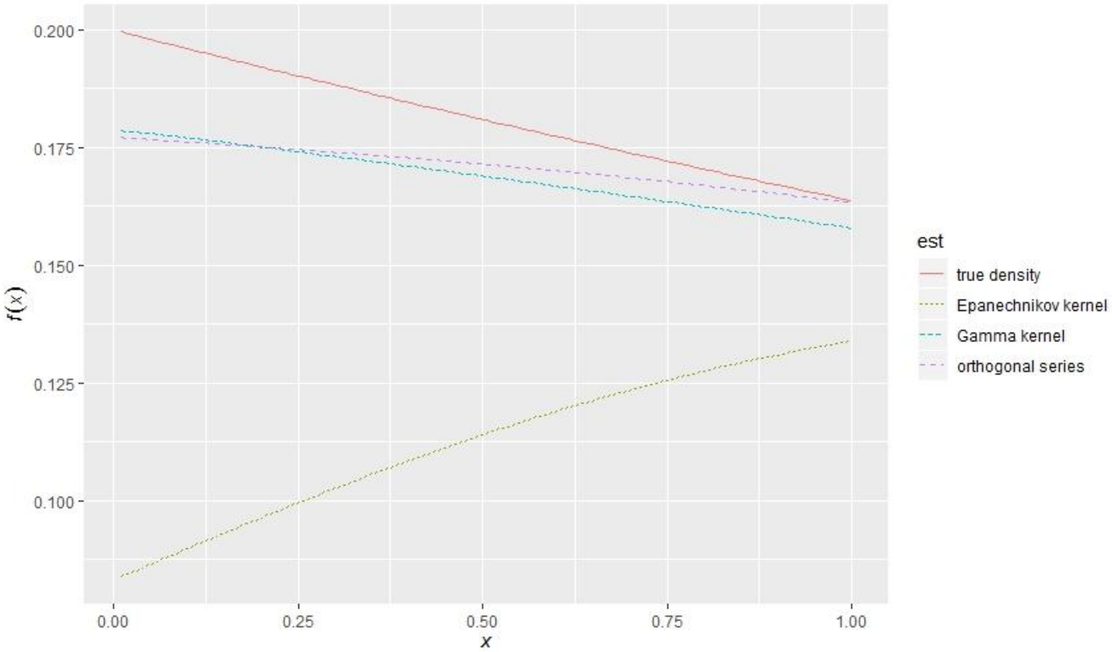
$$f(x) = \frac{a x^{am-\left[\eta+\left(\frac{x}{\lambda}\right)^{a}\right]^{-\eta-m}} \eta^{\eta}}{\lambda^{am} B(m,\eta)}, \quad x \geq 0, a, m, \eta, \lambda > 0 \qquad (۱۰)$$



شکل ۱. منحنی ۱۰ تابع نخست پایه‌های لاگِر و اِرمیت نرمال شده به‌ازای مقادیر $0 \leq x \leq 3$

1

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]



شکل ۲. تابع چگالی توزیع نمایی و برآورد آن به سه روش هسته اپانِچنیکوف و هسته گاما نوع اول و سری‌های متعامد

که در آن $B(m,\eta)=\frac{\Gamma(\eta)\Gamma(m)}{\Gamma(m+\eta)}$. تابع چگالی احتمال (۱۰) از نظر شکل در یک همسایگی نزدیک به صفر بسیار انعطاف‌پذیر است [۳۰]، و از این رو، این خانواده می‌تواند گزینه‌ای مناسب برای مقایسهٔ عملکرد برآوردگرهای مختلف برای تابع‌های چگالی با تکیه‌گاه $[0,\infty)$ باشد.

مالک و شاینلی [۱۹] با انتخاب ۷ مقدار مختلف برای پارامترهای (a,m,η,λ) به تابع‌های چگالی احتمال مختلفی از خانوادهٔ توزیع‌های F تعمیم یافته با رفتارهای گوناگون در همسایگی نزدیک به صفر دست یافتند. مقدارهای در نظر گرفته شده به‌وسیلهٔ مالک و شاینلی [۱۹] برای پارامترهای (a,m,η,λ)، که منجر به شکل‌های گوناگون تابع چگالی (۱۰) در همسایگی نزدیک به صفر می‌شوند، در جدول ۱ فهرست شده‌اند. مقدار پارامترها طوری انتخاب شده‌اند که این تابع‌های چگالی احتمال دارای امید ریاضی برابر با یک باشند. با این حال تابع چگالی احتمال در حالت D_2 دارای واریانس نامتناهی است زیرا گشتاور مرتبهٔ k-ام توزیع تعمیم یافتهٔ F برای مقادیر $k<a\eta$ متناهی و برابر با

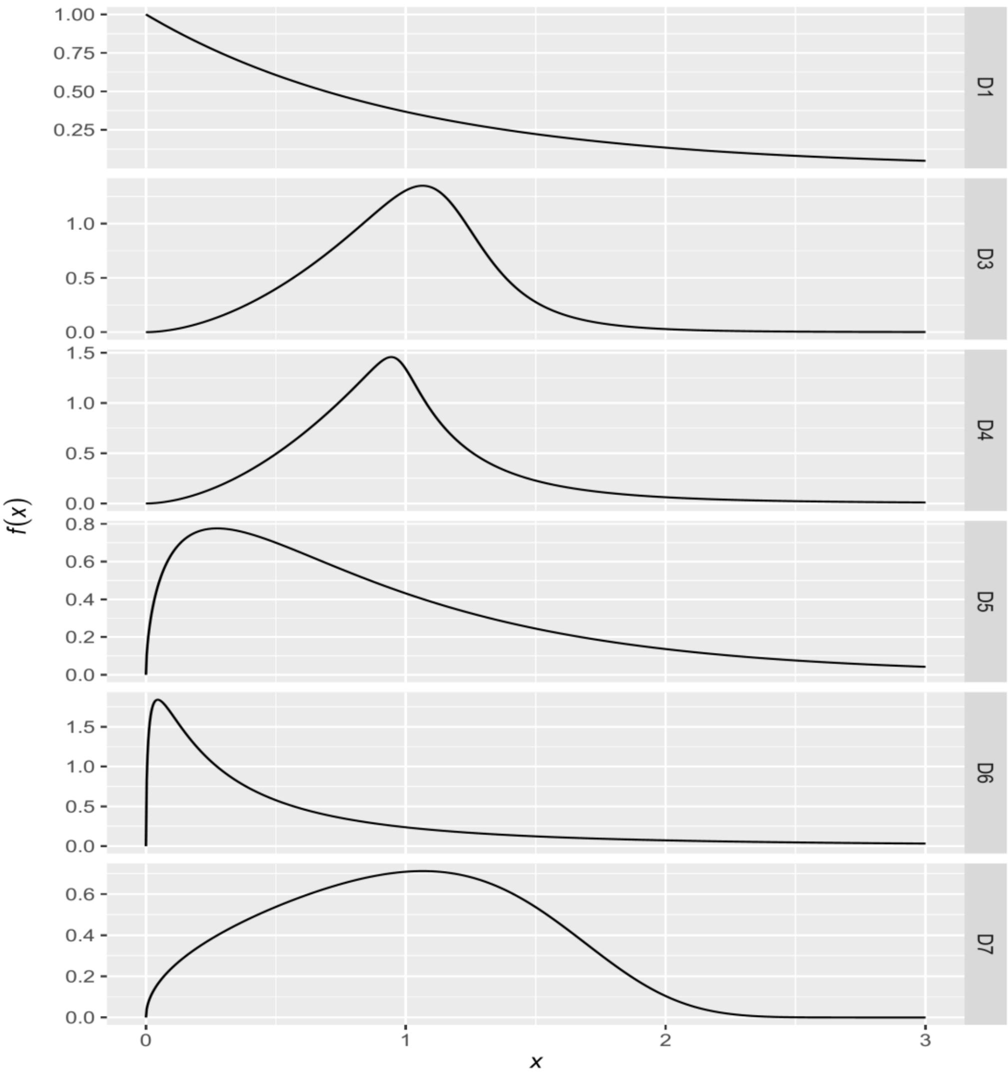
$$\int_0^\infty x^k\ f(x)dx\ =\frac{\lambda\eta^{\frac{1}{a}}\Gamma\left(m+\frac{k}{a}\right)\Gamma\left(\eta-\frac{k}{a}\right)}{\Gamma(m)\Gamma(\eta)}$$

و در غیر این‌صورت نامتناهی است.

جدول ۱. مقادیر در نظر گرفته شده به‌وسیلهٔ مالک و شاینلی [۱۹] برای پارامترهای توزیع تعمیم‌یافتهٔ F

λ	η	m	a	
۱	∞	۱	۱	D_1
۰/۰۹۷	۱/۲	۰/۷	۰/۹	D_2
۱/۲٤	۰/۵	۰/۲	۱٤	D_3
۱/۰۳۵	۰/۱	۰/۰۸	۳۵	D_4
۰/۰۵۹	۵	۳	۰/۵۵	D_6
۱/٦۸	∞	۰/۳	۰/۵	D_7

با صرف نظر کردن از حالت D_2، در این بررسی شبیه‌سازی از تابع‌های چگالی احتمال D_1 و D_3 تا D_7 برای مقایسهٔ برآوردگرهای سری‌های متعامد با برآوردگرهای هسته‌ای گاما استفاده می‌شود. نمودار این تابع چگالی‌ها در شکل ۳ رسم شده است.



شکل ۳. نمودار تابع چگالی توزیع تعمیم‌یافتهٔ F در ۶ حالت در نظر گرفته شده برای شبیه‌سازی

پیش از انجام این مهم، لازم به ذکر است که پارامتر هموارساز ω در $\hat{f}_{g_1}$ و $\hat{f}_{g_2}$ و مقدار برینشی J در $\hat{f}_O$ بر عملکرد این برآوردگرها بسیار تأثیر گذارند. مقادیر کوچک آنها اریبی کم اما واریانس بزرگ برآوردگرهای متناظر را به همراه دارند در حالی که مقادیر بزرگ آنها منجر به نتیجهٔ عکس می‌شوند. بنابراین پارامتر هموارساز ω و مقدار برینشی J باید به‌گونه‌ای بهینه انتخاب شوند. یک معیار بهینگی رایج، استفاده از میانگین جمع‌بستهٔ توان‌دوم خطا است. بر این اساس، پارامتر هموارساز ω و مقدار برینشی J بهینه با کمینه‌سازی $MISE$ به‌دست می‌آیند. این انتخاب، مقدار بهینهٔ فراموضعی[1] نامیده می‌شود. مقدار بهینهٔ فراموضعی پارامتر هموارساز ω در برآوردگرهای گاما به‌وسیلهٔ چن [۱۶] به‌دست آمده است؛ اما شکل بسته‌ای برای انتخاب بهینهٔ فراموضعی مقدار برینشی J در برآوردگرهای سری‌های متعامد وجود ندارد.

1. Global optimal value

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]

برای تعیین مقادیر بهینهٔ ω و J در برآوردگرهای $\hat{f}_{g_1}$، $\hat{f}_{g_2}$ و $\hat{f}_o$ بر اساس یک نمونهٔ تصادفی نوعی X_1، ...، X_n از تابع چگالی هدف f به تعداد $n_{sim} = 5000$ بار نمونه‌هایی با حجم n تولید می‌شود. برای هر نمونهٔ تولید شدهٔ X_{j1}، ...، X_{jn}، $j = 1, \dots, n_{sim}$، برآورد $\hat{f}^{(j)}$ برای تابع چگالی احتمال هدف f با برآوردگرهای هستهٔ گامای (۳) و (۵) و برآوردگر سری‌های متعامد (۸) با پایه‌های لاگِر و اِرمیت نرمال شده به‌دست می‌آید. سپس مقدار میانگین جمع‌بستهٔ توان دوم خطا برای برآوردگرهای هستهٔ گاما به‌صورت (۱۱)

$$\widehat{MISE}\left(\hat{f}_g(.\,;\omega)\right) = \frac{1}{n_{sim}}\sum_{j=1}^{n_{sim}}\sum_{l=1}^{m}\left[f(x_l) - \hat{f}_g^{(j)}(x_l,\omega)\right]^2 \Delta x_l \qquad (۱۱)$$

و برای برآوردگرهای سری‌های متعامد به‌صورت

$$\widehat{MISE}\left(\hat{f}_o(.\,;J)\right) = \frac{1}{n_{sim}}\sum_{j=1}^{n_{sim}}\sum_{l=1}^{m}\left[f(x_l) - \hat{f}_o^{(j)}(x_l)\right]^2 \Delta x_l \qquad (۱۲)$$

تقریب زده می‌شود، که در آن $\Delta\, x_i = x_i - x_{i-1}$، $x_0 = 0$ و x_i چندک $\frac{i}{m}$-ام چگالی f است. در این‌جا $m = 1000$ تعداد نقاطی است که برای تقریب انتگرال موجود در میانگین جمع‌بستهٔ توان دوم خطا (۲) با یک مجموع ریمانی به‌کار گرفته می‌شوند. مقدارهای بهینهٔ ω و J به‌ترتیب با کمینه‌سازی (۱۱) و (۱۲) به‌دست می‌آیند.

به‌منظور مقایسهٔ برآوردگرها، از تابع چگالی احتمال هر یک از حالت‌های D_1 و D_3 تا D_7 به تعداد $n_{sim} = 5000$ بار نمونه‌هایی به حجم $n = 50, 100, 400$ تولید می‌شوند. مقدارهای بهینهٔ ω و J برای هر یک از نمونه‌های شبیه‌سازی شدهٔ به‌روش بالا به‌دست می‌آیند و مقدار تقریبی میانگین جمع‌بستهٔ توان دوم خطای هر کدام از برآوردگرهای متناظر با جای‌گذاری مقدارهای بهینهٔ ω و J در (۱۱) و (۱۲) محاسبه می‌شوند.

برای پیاده‌سازی محاسبات این بررسی شبیه‌سازی از نرم‌افزار و زبان برنامه‌نویسی R و بستهٔ نرم‌افزاری Evmix [۳۱]، برای برآودگرهای هسته‌ای گاما، بستهٔ نرم‌افزاری Orthopolynom [۳۲]، برای محاسبات پایه‌های برآوردگرهای سری‌های متعامد و بستهٔ نرم‌افزاری Flexsurv [۳۳]، برای شبیه‌سازی از خانوادهٔ توزیع‌های تعمیم‌یافتهٔ F استفاده شده است.

مقدار بهینهٔ پارامتر هموارساز ω برای برآوردگرهای هسته‌ای گامای $\hat{f}_{g_1}$ و $\hat{f}_{g_2}$ و مقدار بهینهٔ برینشی J برای برآورگرهای سری‌های متعامد با پایه‌های لاگِر $(\hat{f}_{o_1})$ و اِرمیت نرمال شده $(\hat{f}_{o_2})$، که به‌ترتیب با نمادهای ω_1^*، ω_2^*، J_1^* و J_2^* نمایش داده شده‌اند، در جدول ۲ آمده است. هم‌چنین مقدار تقریبی $MISE$ حاصل برای این برآوردگرها نیز در این جدول گزارش شده‌اند.

در حالت D_1 که تابع چگالی هدف رفتاری مشابه با تابع‌های پایهٔ لاگِر نرمال شده دارد (شکل ۱ و شکل ۲ را مقایسه کنید)، برآوردگر سری‌های متعامد با پایهٔ لاگِر نرمال شده با تعداد بسیار کمی جمله $(J_1^* = 2, 3)$ برای همهٔ حجم‌های نمونه مقدار MISE کم‌تری نسبت به سایر برآوردگرها دارد در حالی که برآوردگر سری‌های متعامد با پایهٔ اِرمیت نرمال شده به تعداد نسبتاً زیادی جمله $(J_2^* \geq 23)$ نیاز دارد و مقدار $MISE$ آن دست کم سه برابر سایر برآوردگرها است. به عکس در حالت D_3 و برای همهٔ حجم‌های نمونه، این برآوردگر سری‌های متعامد با پایهٔ اِرمیت نرمال شده است که کارایی بهتری ($MISE$ کم‌تری) نسبت به سایر رقبا دارد، در حالی که برآوردگر سری‌های متعامد با پایهٔ لاگِر نرمال شده دارای بدترین عملکرد (بیش‌ترین $MISE$) است. هم‌چنین تابع چگالی احتمال در حالت D_7 (شکل ۲) به شکل عمومی تابع‌های پایهٔ اِرمیت نرمال شده (شکل ۱) نزدیک‌تر است و به‌همین دلیل در برآوردگر سری‌های متعامد به تعداد کم‌تری از تابع‌های پایهٔ اِرمیت نرمال شده نسبت به پایهٔ لاگِر نرمال شده نیاز است

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]

$(\frac{J_1^*}{J_2^*} \geq 1.75)$. در این حالت برآوردگر سری‌های متعامد با پایهٔ اِرمیت نرمال شده کارایی بیش‌تری نسبت به همین برآوردگر با پایهٔ لاگر نرمال شده دارد.

جدول ۲. پارامتر هموارساز ω و مقدار برینشی J بهینه و میانگین جمع‌بستهٔ توان دوم خطا برای برآوردگرهای هستهٔ گاما $\hat{f}_{g_1}$ و $\hat{f}_{g_2}$ برآوردگر سری‌های متعامد با پایایهٔ لاگِر $\hat{f}_{o_1}$ و پایهٔ اِرمیت $\hat{f}_{o_2}$ در برآورد تابع‌های چگالی‌های در نظر گرفته شده از خانوادهٔ توزیع‌های تعمیم‌یافتهٔ F (ر.ک. جدول ۱) با حجم نمونه‌های متفاوت

		مقدار برینشی		پارامتر هموارساز		*MISE*			
	n	J_1^*	J_2^*	ω_1^*	ω_1^*	$\hat{f}_{o_1}$	$\hat{f}_{o_2}$	$\hat{f}_{g_1}$	$\hat{f}_{g_2}$
	۵۰	۲	۲۳	۰/۱۵۹۷	۰/۲۲۶۴	۰/۰۰۹۲	۰/۰۵۵۱	۰/۰۱۷۳	۰/۰۱۷۷
D_1	۱۰۰	۲	۴۷	۰/۱۱۹۰	۰/۱۷۲۵	۰/۰۰۴۹	۰/۰۴۱۷	۰/۰۱۱۰	۰/۰۱۰۹
	۴۰۰	۳	۷۵	۰/۰۶۶۸	۰/۱۰۲۳	۰/۰۰۱۷	۰/۰۲۲۸	۰/۰۰۴۳	۰/۰۰۳۹
	۵۰	۴۷	۲۳	۰/۰۲۱۶	۰/۰۲۲۴	۰/۰۴۳۶	۰/۰۳۳۱	۰/۰۳۷۰	۰/۰۳۶۵
D_3	۱۰۰	۴۷	۳۴	۰/۰۱۵۲	۰/۰۱۵۸	۰/۰۳۰۳	۰/۰۲۲۵	۰/۰۲۳۰	۰/۰۲۲۷
	۴۰۰	۴۷	۵۵	۰/۰۷۸	۰/۰۰۸۱	۰/۰۱۷۶	۰/۰۰۷۶	۰/۰۰۸	۰/۰۰۸۶
	۵۰	۳۴	۲۴	۰/۰۲۲۰	۰/۰۲۲۹	۰/۰۴۲۸	۰/۰۳۸۴	۰/۰۳۷۶	۰/۰۳۷۷
D_4	۱۰۰	۳۵	۳۷	۰/۰۱۵۱	۰/۰۱۵۵	۰/۰۳۳۸	۰/۰۲۶۱	۰/۰۲۴۲	۰/۰۲۳۹
	۴۰۰	۴۳	۷۷	۰/۰۰۷۳	۰/۰۰۷۳	۰/۰۳۳۷	۰/۰۴۱۸	۰/۰۰۹۳	۰/۰۰۹۳
	۵۰	۸	۱۵	۰/۱۴۲۷	۰/۱۸۰۶	۰/۰۲۱۷	۰/۰۳۲۱	۰/۰۱۸۶	۰/۰۲۰۱
D_5	۱۰۰	۱۱	۲۷	۰/۰۹۱۷	۰/۰۹۸۰	۰/۰۱۳۱	۰/۰۲۲۵	۰/۰۱۲۳	۰/۰۱۲۹
	۴۰۰	۱۸	۱۵	۰/۳۸۵	۰/۰۴۱۵	۰/۰۰۴۴	۰/۰۱۰۳	۰/۰۰۵۱	۰/۰۰۴۸
	۵۰	۶	۷۱	۰/۰۷۷۸	۰/۱۰۲۰	۰/۰۳۲۶	۰/۱۰۲	۰/۰۳۹۲	۰/۰۴۲۲
D_6	۱۰۰	۷	۷۵	۰/۰۵۳۶	۰/۰۷۴۷	۰/۰۲۱۹	۰/۰۸۰۹	۰/۰۲۵۴	۰/۰۲۶۷
	۴۰۰	۴۰	۷۹	۰/۰۲۲۶	۰/۰۲۶۵	۰/۱۲۳۸	۱/۲۳۱۵	۰/۰۱۱۰	۰/۰۱۱۵
	۵۰	۱۳	۴	۰/۰۶۸۳	۰/۰۷۴۱	۰/۰۳۴۲	۰/۰۲۷۰	۰/۰۲۲۰	۰/۰۲۱۳
D_7	۱۰۰	۲۳	۶	۰/۰۴۸۶	۰/۰۵۲۸	۰/۰۲۷۳	۰/۰۱۸۰	۰/۰۱۳۹	۰/۰۱۳۳
	۴۰۰	۳۵	۲۰	۰/۰۲۵۴	۰/۰۲۷۳	۰/۰۱۰۶	۰/۰۰۷۶	۰/۰۰۵۲	۰/۰۰۵۰

برای همهٔ حجم‌های نمونه در حالت‌های D_4 و D_7 برآوردگرهای هسته‌ای گاما کارایی بهتری نسبت به برآوردگرهای سری‌های متعامد دارند. در همهٔ حالت‌ها جز D_6 کارایی برآودگر هسته‌ای گامای $\hat{f}_{g_2}$ نسبت به $\hat{f}_{g_1}$ برای حجم نمونهٔ بزرگ $(n = 400)$ اندکی بیش‌تر از یک است. این امر گواهی دیگری برای پیشنهاد چِن [۱۶] در استفاده از برآوردگر $\hat{f}_{g_2}$ به جای $\hat{f}_{g_1}$ برای رسیدن به نرخ هم‌گرایی سریع‌تر *MISE* برآوردگر هسته‌ای گاما فراهم می‌کند. در حالت D_5 برای حجم نمونه‌های $n = 50{,}100$ کارایی برآوردگر هسته‌ای گامای $\hat{f}_{g_1}$ و برای حجم نمونهٔ $n = 400$ کارایی برآوردگر سری‌های متعامد با پایهٔ لاگِر نسبت به سایر برآوردگرها بیش‌تر است. در حالت D_6 برآوردگر سری‌های متعامد با پایهٔ لاگِر برای حجم‌های نمونهٔ $n = 50, 100$ کارایی بیش‌تری نسبت به سایر برآوردگرها دارد، اما در

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.3.333]

حجم نمونهٔ $n = 400$ مقدار $MISE$ هر دو برآوردگر سری‌های متعامد افزایش می‌یابد و برآوردگر هسته‌ای گامای $\hat{f}_{g_1}$ نسبت به سایر برآوردگرها کاراتر است. چنان‌که در بخش برآورد سری‌های متعامد اشاره شد، برای اطمینان از هم‌گرایی $MISE$ برآوردگر $\hat{f}_o(.;J)$ به صفر لازم است تابع چگالی هدف f در شرایطی (از جمله متناهی بودن چند گشتاور اول) صدق کند و $J = J_n$ با نرخی سریع‌تر از n^2 رشد کند. با توجه به مقادیر a و η در حالت D_6 (ر.ک. جدول ۱)، گشتاورهای مرتبهٔ سوم و بالاتر تابع چگالی احتمال در این حالت نامتناهی‌اند و به‌علاوه مقادیر J_2^* در جدول ۲ با افزایش n به نسبت کند افزایش می‌یابند. به همین دلیل شرایط هم‌گرایی $MISE$ برآوردگرهای سری‌های متعامد در این حالت برقرار نیست و مقدار $MISE$ برای این برآوردگرها با افزایش n به ۴۰۰ به‌جای کاهش، افزایش داشته است.

بحث و نتیجه‌گیری

در این مقاله عملکرد کوچک‌نمونه‌ای برآوردگرهای هسته‌ای گاما و سری‌های متعامد با دو پایهٔ لاگِر و اِرمیت نرمال شده برای برآورد ناپارامتری تابع چگالی احتمال یک توزیع نامنفی مقایسه شد. در قالب یک بررسی شبیه‌سازی ملاحظه شد که انتخاب برآودگر مناسب از بین این دو دسته از برآوردگرها منوط به داشتن میزانی از اطلاع در مورد تابع چگالی هدف است. در صورتی که پایهٔ برآوردگر سری‌های متعامد طوری انتخاب شده باشد که تابع‌های آن پایه رفتاری مشابه تابع چگالی هدف داشته باشند، برآوردگر سری‌های متعامد کارایی بیش‌تری نسبت به برآوردگر هسته‌ای گاما دارد. اما اگر تابع چگالی هدف در شرایطی که برای سازگاری برآودگرهای سری‌های متعامد نیاز است صدق نکند، برآوردگرهای سری‌های متعامد ممکن است عملکرد نامطلوبی نسبت به برآودگرهای هسته‌ای گاما داشته باشند.

تقدیر و تشکر

از داوران محترم که با پیشنهادها و راهنمایی‌های خود مقاله را پربار کردند، کمال تقدیر و تشکر را داریم.

منابع

1. Tapia R. A., Thompson J. R., "Nonparametric Probability Density Estimation", Johns Hopkins University Press, Baltimore (1978).
2. Silverman B. W., "Density estimation for statistics and data analysis", Chapman & Hall/CRC press, London (1986).
3. Izenman A. J., "Review Papers: Recent Developments in Nonparametric Density Estimation", Journal of the American Statistical Association, 86(413) (1991) 205-224.
4. Wand M. P., Jones M. C., "Kernel Smoothing", Chapman and Hall/CRC, London (1995).
5. Devroye L., Lugosi G., "Combinatorial Methods in Density Estimation", Springer Science & Business Media, New York, (2012).
6. Parzen E., "On Estimation of a Probability Density Function and Mode", The Annals of Mathematical Statistics, 33 (3) (1962) 1065-1076.

7. Marron J., Ruppert D., "Transformations to Reduce Boundary Bias in Kernel Density Estimation", Journal of the Royal Statistical Society. Series B (Methodological), 56 (4) (1994) 653-671.
8. Schuster E. F., "Incorporating support constraints into nonparametric estimators of densities", Communications in Statistics - Theory and Methods, 14(5) (1985) 1123-1136.
9. Cline D. B. H., Hart J. D., "Kernel Estimation of Densities with Discontinuities or Discontinuous Derivatives", Statistics, 22(1) (1991) 69-84.
10. Gasser T., Muller H. G., Mammitzsch V., "Kernels for Nonparametric Curve Estimation", Journal of the Royal Statistical Society. Series B (Methodological), 47 (2) (1985) 238-252.
11. Muller H. G., "Smooth Optimum Kernel Estimators Near Endpoints", Biometrika, 78 (3) (1991) 521-530.
12. Muller H. G., "On the boundary kernel method for non-parametric curve estimation near endpoints", Scand. J. Stat., 20 (1993) 313-328.
13. Jones M. C., "Simple boundary correction for kernel density estimation", Statistics and Computing, 3(3) (1993) 135-146.
14. Wand M. P., Marron J. S., Ruppert D., "Transformations in Density Estimation", Journal of the American Statistical Association, 86 (414) (1991) 343-353.
15. Chen S. X., "Beta kernel estimators for density functions", Computational Statistics & Data Analysis, 31 (2) (1999) 131-145.
16. Chen S. X., "Probability Density Function Estimation Using Gamma Kernels", Annals of the Institute of Statistical Mathematics, 52 (3) (2000) 471-480.
17. Jones M., Foster P., "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation", Statistica Sinica, 6(4) (1996) 1005-1013.
18. Muller H. G., Wang J. L., "Hazard Rate Estimation under Random Censoring with Varying Kernels and Bandwidths", Biometrics, 50(1) (1994) 61-76.
19. Malec P., Schienle M., "Nonparametric Kernel Density Estimation Near the Boundary", Computational Statistics & Data Analysis, 72 (2014) 57-76.
20. Karunamuni R. J., Alberts T., "On boundary correction in kernel density estimation", Statistical Methodology, 2 (3) (2005) 191-212.
21. Gasser T., Muller H. G., "Kernel estimation of regression functions. In Gasser, T., Rosenblatt, M. (Eds.): Smoothing Techniques for Curve Estimation", Lecture Notes in Mathematics, vol. 757. Springer, Heidelberg (1979) 23-68.
22. Cencov N. N., "Estimation of an unknown density function from observations", In Dokl. Akad. Nauk, SSSR Vol. 147 (1964) 45-48.

[DOR: 20.1001.1.25882546.1399.6.3.3.6]
[DOI: 10.52547/mmr.6.3.333]

23. Wahba G., "Data-Based Optimal Smoothing of Orthogonal Series Density Estimates", The Annals of Statistics, 9 (1) (1981) 146-156.

24. Hall P., "Cross-validation and the smoothing of orthogonal series density estimators", Journal of Multivariate Analysis, 21 (2) (1987) 189-206.

25. Efromovich S., "Orthogonal series density estimation. Wiley Interdisciplinary Reviews: Computational Statistics", 2 (4) (2010) 467-476.

26. Kreyszig E., "Introductory functional analysis with applications", Wiley, New York (1989).

27. Watson G. S., "Density Estimation by Orthogonal Series", The Annals of Mathematical Statistics, 40(4) (1969) 1496-1498.

28. Efromovich S., "Nonparametric curve estimation: methods, theory, and applications", Springer Science & Business Media, (2008).

29. Walter G. G., "Properties of Hermite Series Estimation of Probability Density", The Annals of Statistics, 5 (6) (1977) 1258-1264.

30. Cox C., "The generalized F distribution: An umbrella for parametric survival analysis", Statistics in Medicine, 27 (21) (2008) 4301-4312.

31. Scarrott C. J., Hu Y., "evmix 0.2.7: Extreme Value Mixture Modelling, Threshold Estimation and Boundary Corrected Kernel Density Estimation", Availableon CRAN: http://www.math.canterbury.ac.nz/~c.scarrott/evmix (2017).

32. Novomestky F., "orthopolynom 1.0-5: Collection of functions for orthogonal and orthonormal polynomials", Available on CRAN: https://CRAN.R-project org/package=orthopolynom (2013).

33. Jackson C., "flexsurv: A Platform for Parametric Survival Modeling in R", Journal of Statistical Software, 70 (8) (2016) 1-33.

[DOR: 20.1001.1.25882546.1399.6.3.3.6]

[DOI: 10.52547/mmr.6.3.333]