مجله علوم دانشگاه تهران
جلد سی‌ام (۱۳۸۳) شماره ۱ (۷۷-۶۷)

برآورد تابع چگالی احتمال به روشهای هسته‌ای و B-اسپلاین

محسن محمدزاده، رضا صالحی
گروه آمار، دانشکده علوم پایه، دانشگاه تربیت مدرس
E-mail: mohsen_m@modares.ac.ir
(دریافت: ۸۱/۱۱/۱ ؛ پذیرش: ۸۲/۲/۲۴)

چکیده

یکی از روشهای ناپارامتری متداول برای برآورد تابع چگالی احتمال روش هسته‌ای است و در سالهای اخیر استفاده از B-اسپلاین ها برای برآورد تابع چگالی احتمال مطرح شده است. هر یک از این دو روش بنوعی به پارامتر نامعلوم همواری بستگی دارند که مقدار آن در میزان هموازی و دقت برآوردگرهای حاصل تاثیر بسزایی دارد. در این مقاله روشهای برآورد تابع جگالی احتمال با استفاده از توابع هسته‌ای و B-اسپلاین‌ها ارایه و نحوه انتخاب پارامترهای هموازی برای هر دو روش مطرح و میزان دقت برآوردگرهای حاصل بر اساس معیار میانگین انتگرال مربعات خطا و همچنین تاثیر تعداد و میزان پراکندگی داده ها در دقت دو روش مورد مطالعه قرار گرفته است. نتایج نشانگر آن هستند که افزایش پراکندگی داده هایی که از توزیع متقارن پیروی می‌کنند موجب کاهش دقت هر دو برآوردگر می‌شود. در حالیکه این امر برای داده‌هایی که از توزیعی غیر متقارن پیروی می‌کنند، برعکس موجب بهبود دقت برآوردگرهای هر دو روش می‌گردد.

واژه‌های کلیدی: *B- اسپلاین، هسته، پارامتر هموازی، برآورد چگالی.*

مقدمه

معمولا برای برآورد تابع چگالی احتمال از دو روش پارامتری یا ناپارامتری استفاده می‌شود. هرگاه فرم توزیع جامعه را بتوان از طریق نمونه یا اطلاعات قبلی، مسائل مشابه، سنخیت صفت مورد بررسی با توزیع خاص و ... مشخص نمود، از روش پارامتری استفاده می‌شود. اما اگر هیچ اطلاعی در مورد فرم توزیع جامعه در اختیار نباشد، تابع چگالی به روش ناپارامتری بر اساس اطلاعات نمونه برآورد می‌شود. این روش شرایطی را فراهم می‌سازد تا داده‌ها فرم توزیع خود را تعیین نمایند. روشهای هسته‌ای و B- اسپلاین دو روش ناپارامتری هستند که در این مقاله مورد مطالعه قرار میگیرند. ایده برآوردگر هسته‌ای توسط (Rosenblatt,1956) و (Parzen,1962) مطرح شد. بعداً (Watson,1963)، (Epanechnikov, 1969) و (Nadaraya, 1974) مقاله‌هایی در این زمینه ارائه کردند و تا به امروز مطالعه برای بسط و گسترش برآوردگرهای هسته‌ای همچنان ادامه دارد. نخستین تئوری برایB- اسپلاین‌ها توسط (Schoenberg, 1946) پیشنهاد شد و (Cox, 1972) و (DeBoor,1972) فرمول بازگشتی برای محاسبه عددی آن بیان کردند. با اینکه B-اسپلاین‌ها در ریاضیات بطور قابل ملاحظه‌ای بسط و گسترش یافته و تعاریف متفاوتی بنا به کاربرد، به آن داده‌اند و حتی از آن در گرافهای کامپیوتری استفاده‌های چشم‌گیری شده است، اما کاربرد و استفاده آن در آمار به سالهای اخیر بر می‌گردد. (Ciesielski,1990) برآورد چگالی اسپلاین یک متغیره و چند متغیره را مورد بررسی قرار داد. (Mohammadzadeh,1998) الگوریتمی برای بدست آوردن پارامتر همواری اسپلاین‌ها معرفی کرد و (Kent *et al.*, 2000) نحوه بهینه‌سازی معیار اعتبار متقابل تعمیم‌یافته را برای تعیین پارامتر همواری ارائه نمودند. (Krzykowski,2001) مقداری تقریبی برای تعیین پارامتر همواری B-اسپلاین‌ها ارائه داد که محاسبه آن نسبت به روشهای دیگر ساده‌تر می‌باشد. در این مقاله روش برآورد تابع چگالی احتمال با استفاده از هسته‌ها و تعیین پارامتر همواری آن ارائه می‌شود. همچنیین روش برآورد تابع چگالی احتمال با استفاده از B-اسپلاین‌ها و نحوه تعیین پارامتر همواری آن معرفی خواهد شد. در پایان این دو روش به کمک شبیه‌سازی مورد مقایسه عددی قرار می‌گیرند و نهایتاً بحث و نتیجه‌گیری در مورد نتایج حاصل ارائه خواهد شد.

برآورد چگالی هسته‌ای

فرض کنید $X_1,\dots,X_n$ نمونه‌ای تصادفی از یک توزیع با تابع چگالی احتمال f باشد. در اینصورت برآوردگر هسته‌ای $f(x)$ بصورت:

$$\hat{f}_h(x) = \hat{f}_h(X_1,\dots,X_n;x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h}) \qquad (۱)$$

تعریف می‌شود، که در آن مقدار حقیقی و مثبت h میزان همواری برآوردگر هسته‌ای را کنترل می‌کند و $K(.)$ تابعی است که لازم است در شرایط:

$$\int_{-\infty}^{\infty} K(y)dy = 1 \text{ ،} \quad \sup_{-\infty<y<\infty} |K(y)| < \infty \quad \text{و} \quad \lim_{y\to\infty} |yK(y)| = 0$$

صدق کند و تابع هسته‌ای نامیده می‌شود. (Parzen,1962) نشان داد اگر h به صفر میل کند و $nh \to \infty$، آنگاه برآوردگر $\hat{f}_h(x)$ سازگار در میانگین است. بعلاوه $\hat{f}_h$ یک تابع چگالی احتمال است، یعنی $\int_{-\infty}^{\infty} \hat{f}_h(x)dx = 1$ ، و تمام خواص پیوستگی و مشتق‌پذیری K را به ارث می برد. همانطور که ملاحظه می‌شود برآوردگر هسته‌ای (۱) به پارامتر هموار h بستگی دارد. اگر مقدار h خیلی کوچک اختیار شود برآوردگر $\hat{f}_h$ بسیار مواج خواهد بود در حالیکه برای مقادیر بزرگ h این برآوردگر بیش از حد هموار و بیانگر رفتار تابع چگالی نخواهد بود. لذا لازم است روشی مناسب برای تعیین مقدار بهینه آن ارائه شود. برای این منظور اگر هدف برآورد چگالی مجهول f با استفاده از برآوردگر هسته‌ای و بر اساس معیار انتگرال مربعات خطای

$$ISE(h) = \int_{-\infty}^{\infty} (\hat{f}_h(x) - f(x))^2\, dx \qquad (۲)$$

باشد، (Hall, 1982) ثابت کرد معیار ISE و میانگین آن یعنی:

$$\begin{aligned} MISE(h) &= E\int_{-\infty}^{\infty} (\hat{f}_h(x) - f(x))^2\, dx \\ &= \int_{-\infty}^{\infty} Var(\hat{f}_h(x))dx + \int_{-\infty}^{\infty} bias^2(\hat{f}_h(x)) \end{aligned}$$

که درآن امید نسبت به $X_n,\dots,X_1$ است، بطور مجانبی معادل هستند. بنابراین h به گونه‌ای انتخاب می‌شود که $MISE$ مینیمم شود. برای این منظور یک تابع هسته‌ای متقارن با شرایط $\int_{-\infty}^{\infty} t^2K(t)dt > 0$ و $\int_{-\infty}^{\infty} tK(t)dt = 0$ درنظر گرفته می‌شود.

(Hardle,1991) نشان داد اگر f'' مطلقاً پیوسته و $\| f'' \|_2^2 = \int_{-\infty}^{\infty} f''^2(x)dx < \infty$ باشد، بعلاوه تابع هسته‌ای K درشرایط $\| K \|_2^2 = \int_{-\infty}^{\infty} K^2(x)dx < \infty$ و $\mu_2(K) = \int u^2K(u)du < \infty$ صدق کند، آنگاه:

$$\int_{-\infty}^{\infty} bias^2(\hat{f}_h(x))dx = \frac{1}{4}h^2\mu_2^2(K)\|f''\|_2^2 + O(h^5)$$

و:

$$\int_{-\infty}^{\infty} Var(\hat{f}_h(x))dx = \frac{1}{nh}\|K\|_2^2 + O(\frac{1}{n})$$

بنابراین:

$$MISE(h) = \frac{\|K\|_2^2}{nh} + \frac{1}{4}h^4\mu_2^2(K)\|f''\|_2^2 + O(\frac{1}{n} + h^5)$$
$$= AMISE(h) + O(\frac{1}{n} + h^5)$$

اگر از $AMISE(h)$ نسبت به h مشتق گرفته و مساوی صفر قرار داده شود، آنگاه h بهینه بصورت:

$$\hat{h}_K = \left[\frac{\|K\|_2^2}{\mu_2^2(K)\|f''\|_2^2}\right]^{\frac{1}{5}} n^{-\frac{1}{5}} \qquad (۳)$$

حاصل می‌شود و از آنجا که مشتق مرتبه دوم $AMISE(h_K)$ به ازای $\hat{h}_K$ مثبت است، $\hat{h}_K$ مقدار $AMISE(h)$ را مینیمم می‌کند. ولی به دلیل وابستگی $\hat{h}_K$ به مشتق مرتبه دوم تابع چگالی مجهول f، محاسبه مستقیم آن مقدور نمی‌باشد. (Parzen, 1962) نشان داد که $\hat{h}_K$ از مرتبه $n^{-\frac{1}{2r+1}}$ است و $MISE$ با نرخ $n^{-\frac{2r}{2r+1}}$ به صفر همگراست، که در آن r عددی مثبت است بطوریکه $\lim_{u\to 0}\frac{1-k(u)}{|u|^r} > 0$ و k تبدیل فوریه تابع هسته‌ای K است. لذا وقتی حجم نمونه n افزایش یابد $\hat{h}_K$ با آهنگ آهسته‌ای به صفر همگراست. از آنجا که $\|f''\|_2^2$ میزان ناهمواری چگالی f را نشان می‌دهد، هر چه تابع چگالی هموارتر باشد $\hat{h}_K$ کوچکتر خواهد بود. بعنوان مثال اگر f تابع چگالی توزیع نرمال با میانگین μ و واریانس σ^2 باشد، برای تابع هسته‌ای نرمال استاندارد داریم: $\hat{h}_K = 1.06\sigma n^{-\frac{1}{5}}$ و برای هسته دو وزنی $\hat{h}_K = 2.78\sigma n^{-\frac{1}{5}}$ است، که در صورت نامعلوم بودن σ انحراف معیار نمونهٔ S جایگزین آن می‌گردد. چون در حالت کلی تابع f و در نتیج مشتق مرتبه دوم آن نامعلوم

است، (Park *et al*,1989) برآورد ناپارامتری f'' را به روشهای جایگذاری تصحیح شده ارائه نمود که با استفاده از مشاهدات می‌توان $\hat{h}_K$ را محاسبه کرد.

برآورد چگالی B-اسپلاین

اگر $U_1,\ldots,U_r$ نمونه‌ای تصادفی از توزیع یکنواخت بر فاصله $\left[-\frac{1}{2},\frac{1}{2}\right]$ باشد، چگالی مجموع این متغیرهای تصادفی بصورت:

$$B^{(r)}(x)=\frac{\partial}{\partial x}P(U_1+U_2+\ldots+U_r\leq x),\quad x\in R$$

است. برای هر $r\geq 1$، تابع $B^{(r)}(.)$ دارای مشتق پیوسته مرتبه $r-2$ با تکیه‌گاه $I_r=\left[-\frac{r}{2},\frac{r}{2}\right]$ است، که B-اسپلاین مرتبه r نامیده می‌شود. بعلاوه اگر برای هر عدد صحیح i قرار دهیم $J_i=(i+\frac{r}{2},i+1+\frac{r}{2})$ آنگاه برای هر $x\in J_i$ برحسب اینکه J_i زیر مجموعه ای از I_r یا خارج آن باشد، تابع $B^{(r)}(x)$ یک چند جمله‌ای مرتبه r یا صفر می‌باشد. بعنوان مثال B-اسپلاین‌های مرتبه‌های یک تا سه بترتیب بصورت:

$$B^{(2)}(x)=\begin{cases}1+x & -1\leq x<0\\ 1-x & 0\leq x\leq 1\\ 0 & o.w.\end{cases}\quad,\quad B^{(1)}(x)=\begin{cases}1 & -\frac{1}{2}\leq x<\frac{1}{2}\\ 0 & o.w.\end{cases}$$

و

$$B^{(3)}(x)=\begin{cases}\frac{1}{2}(x+\frac{3}{2})^2 & -\frac{3}{2}\leq x<-\frac{1}{2}\\ -x^2+\frac{3}{4} & -\frac{1}{2}\leq x\leq\frac{1}{2}\\ \frac{1}{2}(x-\frac{3}{2})^2 & \frac{1}{2}<x\leq\frac{3}{2}\\ 0 & o.w.\end{cases}$$

هستند، که $B^{(2)}$ حاصلضرب پیچشی دو B-اسپلاین مرتبه یک بصورت $B^{(2)}=B^{(1)}*B^{(1)}$ است و بطور مشابه B-اسپلاین مرتبه سه بصورت

$B^{(3)} = B^{(2)} * B^{(1)}$ است. همانطور که ملاحظه می‌شود $B^{(1)}$ یک تابع پله ای ثابت، $B^{(2)}$ تابع خطی قطعه‌ای و $B^{(3)}$ تابع درجه دو قطعه‌ای هستند.

می‌توان با انتقال و تغییر مقیاس $B^{(r)}$، B-اسپلاین را برای مجموعه‌ای از نقاط روی R که بطور یکسان تقسیم شده اند، بصورت:

$$B_{s,h}^{(r)}(x) = B^{(r)}(\frac{x}{h} - s), \qquad x \in R, s \in Z$$

تعریف نمود. (Ciesielski, 1987) برآوردگر B-اسپلاین مرتبه r تابع چگالی f را براساس نمونه تصادفی $X_1, ..., X_n$ بصورت:

$$\hat{f}_{n,h,r}(x) = \frac{1}{n}\sum_{j=1}^{n} Q_h^{(r)}(X_j, x), \qquad h \in R_+, x \in R, r \geq 1 \qquad (۴)$$

تعریف نمود، که در آن h پارامتر هموار‌ی، $Q_h^{(r)}(x,y) = \frac{1}{h}Q^{(r)}(\frac{x}{h},\frac{y}{h}), h > 0$ هسته مقیاس ساز و $Q^{(r)}(x,y) = \sum B^{(r)}(x-s)B^{(r)}(y-s), (x,y) \in R^2$ هسته اسپلاین نامیده می‌شود. (Krzykowski, 2001) نشان داد این تابع دارای خواص:

$$Q^{(r)}(x,y) = Q^{(r)}(y,x) = 0 \qquad |x-y| > r$$

$$\int Q^{(r)}(x,y)dx = 1 \qquad y \in R$$

$$\int xQ^{(r)}(x,y)dx = y \qquad y \in R, r > 1$$

$$\int x^2 Q^{(r)}(x,y)dx = y^2 + \frac{r}{6} \qquad y \in R, r > 2$$

است. برای مطالعه بیشتر درخصوص این برآوردگر و حالت چند متغیره آن می‌توان به (Ciesielski, 1990) مراجعه نمود. (Krzykowski, 2001) نشان داد برآوردگر چگالی $\hat{f}_{n,h,r}(x)$ در فضای نرم دار L_1 سازگار است، یعنی اگر $EX_1^2 < \infty$ باشد آنگاه برای هر $x \in R$، $\| \hat{f}_{n,h,r}(x) - f(x) \|_1$ در احتمال به صفر همگرا است. با توجه به اهمیت مقدار پارامتر هموار‌ی h، در اینجا راه حلی برای تعیین مقدار بهینه آن ارائه می‌شود. برای این منظور میانگین نمونه‌ای $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$، یعنی برآوردگر ناریب با واریانس کمینه برای میانگین $\mu = \int xf(x)dx$ و $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$ برآوردگر ناریب با

واریانس کمینه برای واریانس $\sigma^2 = \int (x-\mu)^2 f(x)dx$ را در نظر می‌گیریم. با توجه به خواص هسته اسپلاین داریم:

$$\int \hat{f}_{n,h,r}(x)dx = \frac{1}{n}\sum_{j=1}^{n} \int x \frac{1}{h} Q^{(r)}(\frac{X_j}{h}, \frac{x}{h})dx = 1$$

درنتیجه (۴) برآوردگری از یک چگالی احتمال است. بعلاوه میانگین آن برابر میانگین نمونـه‌ای $\overline{X}_n$ است، زیرا:

$$\begin{aligned}\int x\hat{f}_{n,h,r}(x)dx &= \frac{1}{n}\sum_{j=1}^{n} \int x \frac{1}{h} Q^{(r)}(\frac{X_j}{h}, \frac{x}{h})dx \\ &= \frac{1}{n}\sum_{j=1}^{n} \int uhQ^{(r)}(\frac{X_j}{h}, u)du \\ &= h\frac{1}{n}\sum_{j=1}^{n}\frac{X_j}{h} \\ &= \overline{X}_n\end{aligned}$$

و واریانس آن عبارتست از:

$$\begin{aligned}\int (x-\overline{X}_n)^2 \hat{f}_{n,h,r}(x)dx &= \frac{1}{n}\sum_{j=1}^{n} \int x^2 \frac{1}{h} Q^{(r)}(\frac{X_j}{h}, \frac{x}{h})dx - (\overline{X}_n)^2 \\ &= \frac{1}{n}\sum_{j=1}^{n} h^2(\frac{X_j^2}{h^2} + \frac{r}{6}) - (\overline{X}_n)^2 \\ &= \frac{1}{n}\sum_{j=1}^{n} X_j^2 + \frac{rh^2}{6} - (\overline{X}_n)^2 \\ &= \frac{rh^2}{6} + \frac{n-1}{n}S_n^2\end{aligned}$$

اکنون از مساوی قرار دادن واریانس $\hat{f}_{n,h,r}(x)$ و واریانس نمونه‌ای S_n^2 داریم:

$$S_n^2 = \frac{rh^2}{6} + \frac{n-1}{n}S_n^2$$

بنابراین برآورد گشتاوری h بصورت

$$h_B = h_{B,n}(X_1, \ldots, X_n) = \sqrt{\frac{6}{nr}}S_n \qquad (۵)$$

قابل ارائه است که بر اساس مشاهدات محاسبه می‌شود.

مقایسه برآوردگرها

در این بخش دقت برآوردگرهای حاصل از دو روش هسته‌ای و B-اسپلاین مورد مطالعه قرار می‌گیرد. برای بررسی تاثیر عواملی از جمله حجم نمونه، واریانس داده‌ها، تقارن توزیع، انواع توابع هسته‌ای و B-اسپلاین‌های متفاوت در میزان دقت برآوردگر توابع چگالی یک مطالعه شبیه‌سازی بصورت زیر انجام گرفته است. از توزیعهای متقارن نرمال با میانگین صفر و واریانس‌های ۳۲ ، ۱۶ ، ۴ ، ۱ = σ^2 و همچنین توزیعهای نامتقارن گاما با پارامترهای
$(\alpha, \beta) =$ (۱ و ۱)، (۱ و ۲)، (۱ و ۴)، (۲ و ۴)
نمونه‌هایی تصادفی به اندازه‌های ۱۰، ۱۵ و ۳۰ تولید شده‌اند. برای برآورد تابع چگالی به روش هسته‌ای، سه تابع هسته‌ای نرمال استاندارد $K(y) = \frac{1}{\sqrt{2\pi}} Exp(-\frac{y^2}{2})$ ، اپانچنیکوف $K(y) = \frac{3}{4}(1 - y^2) I(|y| \leq 1)$ و دو وزنی $K(y) = \frac{15}{16}(1 - y^2)^2 I(|y| \leq 1)$ را درنظر گرفته پارامتر همواری را با استفاده از (۳) محاسبه و برآوردگر هسته‌ای نیز از رابطه (۱) تعیین شده‌اند. همچنین بازای ۳ و ۲ =r پارامتر همواری (۵) تعیین و برآورد B-اسپلاین مرتبه r تابع چگالی از رابطه (۴) محاسبه گردیده است. برای هر ترکیب از پارامترهای (r, σ^2, K, n) ، عملیات را ۱۰۰ بار تکرار نموده و میانگین ISE های متناظر با هر برآوردگر، یعنی $MISE$ محاسبه و نتایج در جداول ۱ و ۲ خلاصه شده‌اند.

همانطور که در جدول ۱ ملاحظه می‌شود، مقدار $MISE$ برای برآورد تابع چگالی با افزایش واریانس داده‌هایی که از توزیع نرمال تولید شده‌اند، افزایش پیدا می‌کند، درحالی که در جدول۲ با افزایش واریانس داده‌هایی که از توزیع نامتقارن گاما تولید شده‌اند مقدار $MISE$ کاهش می‌یابد. مقایسه $MISE$ در دو جدول نشانگر آنستکه هر دو روش با نمونه‌های بزرگتر برآوردهای دقیقتری ارایه می‌نمایند و برای نمونه‌های کوچک تفاوت قابل ملاحظه‌ای بین دقت دو نوع برآوردگر وجود ندارد.

جدول ۱- مقادیر $MISE$ برای توزیع نرمال.

برآوردگر B-اسپلاین		برآوردگر هسته‌ای			σ^2	n
درجه دو	خطی	دو وزنی	اپانچنیکوف	نرمال		
۰/۰۵۵۰	۰/۰۵۳۹	۰/۰۳۱۳	۰/۰۵۶۰	۰/۰۳۲۳	۱	۱۰
۰/۱۶۴۶	۰/۱۷۲۵	۰/۱۶۲۰	۰/۱۵۷۹	۰/۱۶۲۰	۴	
۰/۲۵۲۰	۰/۲۵۳۶	۰/۲۴۳۲	۰/۲۴۰۱	۰/۲۴۳۲	۱۶	
۰/۲۶۷۰	۰/۲۶۷۸	۰/۲۶۱۸	۰/۲۶۰۰	۰/۲۶۱۷	۳۲	
۰/۰۳۴۸	۰/۰۲۹۸	۰/۰۱۶۶	۰/۰۲۷۲	۰/۰۱۷۰	۱	۱۵
۰/۱۰۴۷	۰/۱۰۹۶	۰/۱۰۷۲	۰/۱۰۴۵	۰/۱۰۷۳	۴	
۰/۱۶۶۹	۰/۱۶۷۹	۰/۱۶۱۹	۰/۱۶۰۲	۰/۱۶۱۹	۱۶	
۰/۱۷۷۴	۰/۱۷۷۹	۰/۱۷۴۴	۰/۱۷۳۴	۰/۱۷۴۴	۳۲	
۰/۰۱۷۳	۰/۰۱۲۷	۰/۰۰۴۷	۰/۰۰۷۶	۰/۰۰۴۹	۱	۳۰
۰/۰۴۷۶	۰/۰۴۸۹	۰/۰۵۲۶	۰/۰۵۰۷	۰/۰۵۲۶	۴	
۰/۰۸۱۸	۰/۰۸۲۱	۰/۰۸۰۷	۰/۰۷۹۷	۰/۰۸۰۶	۱۶	
۰/۰۸۷۸	۰/۰۸۷۹	۰/۰۸۷۱	۰/۰۸۶۱	۰/۰۸۷۱	۳۲	

جدول ۲- مقادیر $MISE$ برای توزیع گاما.

برآوردگر B-اسپلاین		برآوردگر هسته‌ای			σ^2	(α,β)	n
درجه دو	خطی	دو وزنی	اپانچنیکوف	نرمال			
۰/۱۶۷۹	۰/۱۵۶۸	۰/۱۵۰۶	۰/۱۳۶۲	۰/۱۴۵۴	۱	(۱و۱)	۱۰
۰/۰۷۸۰	۰/۰۷۲۵	۰/۰۶۸۲	۰/۰۶۰۵	۰/۰۶۵۶	۴	(۱و۲)	
۰/۰۳۰۴	۰/۰۲۸۶	۰/۰۳۰۸	۰/۰۲۵۱	۰/۰۲۹۵	۱۶	(۱و۴)	
۰/۰۰۷۰	۰/۰۰۶۳	۰/۰۰۵۵	۰/۰۰۵۸	۰/۰۰۵۳	۳۲	(۲و۴)	
۰/۱۰۷۲	۰/۰۹۶۷	۰/۰۸۹۹	۰/۰۷۰۵	۰/۰۸۵۸	۱	(۱و۱)	۱۵
۰/۰۴۹۹	۰/۰۴۴۸	۰/۰۴۰۷	۰/۰۳۰۸	۰/۰۳۸۷	۴	(۱و۲)	
۰/۰۱۸۵	۰/۰۱۶۶	۰/۰۱۸۷	۰/۰۱۳۱	۰/۰۱۷۷	۱۶	(۱و۴)	
۰/۰۰۴۱	۰/۰۰۳۱	۰/۰۰۲۷	۰/۰۰۲۲	۰/۰۰۲۵	۳۲	(۲و۴)	
۰/۰۶۲۰	۰/۰۵۴۲	۰/۰۳۷۹	۰/۰۲۷۱	۰/۰۳۶۰	۱	(۱و۱)	۳۰
۰/۰۲۹۲	۰/۰۲۵۳	۰/۰۱۷۱	۰/۰۱۱۸	۰/۰۱۶۲	۴	(۱و۲)	
۰/۰۱۱۱	۰/۰۰۹۲	۰/۰۰۸۰	۰/۰۰۵۱	۰/۰۰۷۵	۱۶	(۱و۴)	
۰/۰۰۳۸	۰/۰۰۲۳	۰/۰۰۱۰	۰/۰۰۰۷	۰/۰۰۰۹	۳۲	(۲و۴)	

بحث و نتیجه‌گیری

تابع چگالی احتمال به دو روش هسته‌ای و B-اسپلاین برآورد شده و بر اساس میانگین انتگرال مربعات خطا مورد مقایسه عددی قرار گرفته‌اند. همانطور که انتظار می‌رود افزایش حجم نمونه موجب کاهش میانگین انتگرال مربعات خطای متناظر با برآوردگرهای حاصل از هر دو روش می‌شود. اما نتیجه قابل ملاحظه این است که میانگین انتگرال مربعات خطای دو برآوردگر هسته‌ای و B-اسپلاین برای داده‌هایی که از توزیع نرمال با پراکندگی زیاد پیروی می‌کنند افزایش پیدا می‌کنند، در حالیکه این مقدار برای داده‌هایی که از توزیعی غیر متقارن مانند گاما برخوردار هستند با افزایش پراکندگی داده‌ها برعکس حالت قبل کاهش پیدا می‌کنند. این تفاوت می‌تواند ناشی از این واقعیت باشد که داده‌هایی که از توزیع نامتقارن برخوردار هستند، بر خلاف توزیع‌های متقارن، به یک نسبت حول میانگین توزیع نیستند و احیاناً چولگی براست یا چپ توزیع منجر به اختلاف زیاد داده‌ها در دم توزیع می‌شوند. در اینحالت پراکندگی بیشتر داده‌ها رفتار توزیع را، بخصوص در دم‌های توزیع، بهتر مشخص می‌کنند و موجب افزایش دقت برآوردگرهای تابع چگالی می‌شوند. به هر حال دو برآوردگر هسته‌ای و B-اسپلاین برای هر دو مجموعه داده شبیه‌سازی شده از دقت مشابهی برخوردار هستند، اما سرعت بیشتر محاسبه برآوردهای هسته‌ای برتری آنها را به روش برآورد B-اسپلاین، بخصوص وقتی حجم داده‌ها زیاد باشد را بیان می‌دارد.

Reference

Ciesielski, Z. (1987) *Local Spline Approximation and Nonparametric Density Estimation*, International Conference on Constructive Theory of Function. 25-31, May Varna.

Ciesielski, Z. (1990) *Asymptotic Nonparametric Spline Density Estimation in Several Variables*, International Series of Numerical Mathematics (Birkhauser Verlag Basel). **94**, 25-53.

Cox, M.G. (1972) *The Numerical Evaluation of B-Splines*, Jour. Inst. Math. Applic. **10**, 134-149.

De Boor, C. (1972) *On Calculating with B-Splines*, Jour. Approx. Theory. **6**, 50-62.

Epanechnikov, V. (1969) *Nonparametric Estimates of a Multivariate Probability Density*, Theory of Probability and its Applications. **14**, 153-158.

Hall, P. (1982) *Limit Theorems for Stochastic Measures of the Accuracy of Density Estimators*, Stochastic Process Appl. **13**, 11-25.

Hardle, W. (1991) *Smoothing Techniques: with Implementation in S.* Springer-Verlag Newyork.

Kent, J.T., and Mohammadzadeh, M. (2000) *Global Optimization of the Generalized Cross-Validation Criterion,* Statistics and Computing. **10**, 231-236.

Krzykowski, G. (2001) *On Automatic Choice of the Window Parameter in the Nonparametric Density Estimation.* University of Gdansk. Poland.

Mohammadzadeh, M. (1998), *An Algorithm to Find the Smoothing Parameter in Smoothing Splines*, Proceeding of the 4th Iranian Statistical Conference. Shahid Beheshti University, Tehran.

Nadaraya, E.A. (1974) *On the Integral Mean Square Error of Some Nonparametric Estimates for the Density Function*, Theory Probab. Appl. **19**, 133-141.

Park, B.U., and Marron, J.S. (1989) *Comparison of Data-Driven Bandwidh Selectors*, Journal of the American Statistical Association. **84**, 66-72.

Parzen, E. (1962) *On Estimation of a Probability Density Function and Mode,* Ann. Math. Statist. **35**, 1065-1076.

Rosenblatt, M. (1956) *Remarks on Some Nonparametric Estimates of a Density Function*, Ann. Math. Statist. **27**, 642-669.

Schoenberg I.J. (1946) *Contributions to the Problem of Approximation of Equidistant Data by Analytic Function*, Quart. Appl. Math. **4**, 45-99.

Watson, G.S., and Leadbetter, M.R. (1963) *On the Estimation of the Probability Density*, Ann. Math. Stat. **34**, 480-491.