

برآورد تابع چگالی احتمال به روش هسته

رحیم چینی پرداز^۱ رباب افشاری^۲

چکیده

در بسیاری از موارد به خصوص در مسأله آنالیز داده ها، تابع چگالی احتمال معمولاً نامعلوم است، بنابراین لازم است که تابع چگالی احتمال برآورد شود. چهار روش برای برآورد تابع چگالی احتمال وجود دارد، در این مقاله سعی شده است روش هسته که کاربرد وسیعی دارد در حالت یک متغیره و چند متغیره مورد بررسی قرار گیرد. برآورد تابع چگالی بدست آمده توسط هر روشی بستگی زیادی به مقدار پارامتر هموار کننده بکار گرفته شده در برآورد دارد، بطوریکه بکارگیری مقادیر کوچک پارامتر هموارکننده باعث نا همواری بیش از حد، و مقادیر بزرگ آن باعث همواری زیاد برآورد می گردد. بنابراین برای کاهش میزان اختلاف بین تابع چگالی احتمال و برآوردش، به شرط حفظ همواری مقدار مناسبی از پارامتر هموار کننده باید انتخاب شود، در این مقاله برخی روشهای انتخاب پارامتر هموار کننده نیز مورد بررسی قرار می گیرد.

۱. مقدمه

تابع چگالی احتمال یک مفهوم مهم آماری است که شناخت جامعه آماری، بوسیله آن صورت می گیرد. مفاهیمی مانند تقارن، چولگی، نما، ... از طریق تابع چگالی احتمال قابل تشخیص هستند، اما این تابع چگالی معمولاً نامعلوم است. مجهول بودن تابع چگالی احتمال ممکن است به دو صورت باشد: یکی اینکه ساختار تابعی آن معلوم باشد و تنها پارامترهای آن نامعلوم باشند، که به این حالت، حالت پارامتری گفته می شود. دوم اینکه ساختار تابعی علاوه بر پارامترها، مجهول باشد. این حالت، حالت ناپارامتری گفته می شود. در حالت اول مسأله برآورد پارامترها مطرح است که در این مورد کتابهای کلاسیک و حتی مقدماتی آمار به آن پرداخته اند. در روشهای موجود برای برآورد پارامترها مانند روش گشتاوری، ماکزیمم درستنمایی، بیزی، ...، از ساختار تابع چگالی استفاده می شود. در حالت دوم مسأله پیچیده تر است و لازم است خود تابع چگالی برآورد شود.

تا کنون چهار روش برای برآورد تابع چگالی احتمال پیشنهاد شده است، روش هیستوگرام^۲، روش نزدیکترین همسایگی^۴، روش سری های متعامد^۵ و روش هسته^۶. خوشبختانه خواننده با روش هیستوگرام، آشنایی مقدماتی دارد. استفاده از سریهای فوریه برای برآورد توابع نیز تا حدودی شناخته شده است. اما روش نزدیکترین همسایگی و روش

هسته که دارای اهمیت بیشتری هستند نیاز به شناخت بیشتری دارند.

۲. برآورد هسته ای

فرض کنید X_1, X_2, \dots, X_n ، یک نمونه n تایی هم توزیع و مستقل از تابع چگالی مجهول f باشد. برآورد هسته ای f در نقطه x را با $\hat{f}(x)$ نمایش داده به صورت زیر تعریف می شود:

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

که در آن h ، پارامتر هموار کننده^۷ و K ، یک تابع هسته ای^۸ است. تابع K معمولاً و نه لزوماً، همیشه یک تابع چگالی احتمال متقارن با میانگین صفر در نظر گرفته می شود و در بیشتر مواقع سه شرط زیر را برای تابع K ضروری فرض می کنند:

$$\int_{-\infty}^{+\infty} K(t) dt = 1, \quad \int_{-\infty}^{+\infty} tK(t) dt = 0, \quad \int_{-\infty}^{+\infty} t^2 K(t) dt = k_2 \neq 0$$

با توجه به تعریف مشخص است که برآورد هسته ای در نقطه x ، مجموع برآمدگیها^۹ یی است که روی مشاهدات با استفاده از خود

۵- Orthogonal series

۶- Kernel

۷- Smoothing parameter or bandwidth

۸- Kernel function

۱- گروه آمار، دانشگاه شهید چمران اهواز

۲- دانشجوی کارشناسی ارشد آمار، دانشگاه شهید چمران

۳- Histogram

۴- Nearest-neighbour

همانطوریکه گفته شد، برآورد هسته‌ای، به دو عامل، پارامتر هموار کننده و تابع هسته، بستگی دارد. برای روشن شدن این رابطه، مقادیر واریانس و اریبی را محاسبه کرده و در رابطه (۲) جایگذاری می‌کنیم.

فرض کنید تابع چگالی f دارای مشتقات پیوسته از هر مرتبه‌ای باشد، برای اریبی $\hat{f}(x)$ داریم:

$$\begin{aligned} Bias_h(x) &= E(\hat{f}(x)) - f(x) \\ &= \int h^{-1}k(h^{-1}(x-y))f(y)dy - f(x) \end{aligned}$$

$$= \int k(t)f(x-h t)dt - f(x)$$

با تغییر متغیر $\frac{x-y}{h} = t$ و بسط تیلور $f(x-h t)$ حول نقطه x خواهیم داشت:

$$Bias_h(x) = \frac{1}{\nu} h^\nu f''(x)k_\nu + o(h^\nu)$$

$$\approx \frac{1}{\nu} h^\nu f''(x)k_\nu$$

در نتیجه:

$$\int (Bias_h(x))^\nu dx \approx \frac{1}{\nu} h^\nu k_\nu^\nu \int f''^\nu(x) dx \quad (3)$$

و به همین ترتیب خواهیم داشت:

$$V(\hat{f}(x)) \approx n^{-1} h^{-1} f(x) \int K^\nu(x) dx \quad (4)$$

حال روابط (۳) و (۴) را در رابطه (۲) جایگذاری کرده داریم:

$$MISE_{(x)}(\hat{f}) = \frac{1}{\nu} h^\nu k_\nu^\nu \int f''^\nu(x) dx + n^{-1} h^{-1} \int K^\nu(t) dt.$$

هدف می‌نیم کردن $MISE$ نسبت به h می‌باشد. برای اینکه این مقدار می‌نیمم شود باید هر دو مولفه سمت راست رابطه (۵) می‌نیمم شود. با کوچک گرفتن مقدار h مقدار مؤلفه اول یعنی مربع میزان اریبی کوچک می‌شود ولی در عوض مؤلفه دوم یعنی واریانس، افزایش پیدا می‌کند و برعکس.

بنابراین برای ایجاد تعادل، با مشتق گیری از (۵) نسبت به h و مساوی صفر قرار دادن آن خواهیم داشت:

$$h = h_{opt} = k_\nu^{-\frac{1}{\nu}} \left\{ \int K^\nu(t) dt \right\}^{\frac{1}{\nu}} \left\{ \int f''^\nu(x) dx \right\}^{-\frac{1}{\nu}} n^{-\frac{1}{\nu}} \quad (6)$$

که در آن $\int f''^\nu(x) dx$ را میزان نوسانات تابع چگالی تعریف می‌کنند.

مشاهده، همراه با مشاهدات همسایه ساخته می‌شود. لازم به ذکر است که در برآورد هسته ای تابع K ، شکل برآمدگیها و مقدار h ، عرض برآمدگیها را نمایش می‌دهد. به عبارت دیگر h ، پارامتری است که مقدار زیاد آن باعث همواری زیاد^۱ و مقدار کم آن باعث ناهمواری زیاد^۲ می‌شود.

از تعریف برآورد چگالی به روش هسته ای مشخص است که $\hat{f}(x)$ به داده‌ها بستگی دارد. پس خود می‌تواند به عنوان یک متغیر تصادفی در نظر گرفته شود.

۳. مقیاسهای تغییر پذیری

اگر چه در حالت پارامتری برآوردهای ناریب برای پارامترهای توزیع هایی مانند نرمال، پواسن، نمایی، هندسی، ... وجود دارند، اما روزن بلات^۳ [۱۱] نشان داد که در حالت ناپارامتری هیچ برآورد کننده ناریبی وجود ندارد. بنابراین باید در بین دنباله ای از برآورد کننده های بطور مجانبی ناریب به جستجوی بهترین برآورد پرداخته شود.

\hat{f} به طور مجانبی ناریب برای f است هرگاه برای هر مقدار حقیقی x :

$$\lim_{n \rightarrow \infty} E(\hat{f}(x)) = f(x).$$

با توجه به اریب بودن $\hat{f}(x)$ برای سنجش میزان اختلاف بین برآورد یک پارامتر با خود آن پارامتر، می‌توان از معیارهای MSE (میانگین مربع خطا) و یا $MISE$ (میانگین انتگرال مربع خطا) استفاده نمود.

طبق تعریف اگر $\hat{f}(x)$ برآورد چگالی در نقطه x باشد داریم:

$$\begin{aligned} MSE_x(\hat{f}) &= E(\hat{f}(x) - f(x))^2 \\ &= V(\hat{f}(x)) + [E(\hat{f}(x)) - f(x)]^2 \end{aligned} \quad (1)$$

همچنین

$$\begin{aligned} MISE_x(\hat{f}) &= E(\hat{f}(x) - f(x))^2 dx \\ &= \int E(\hat{f}(x) - f(x))^2 dx \\ &= \int V(\hat{f}(x)) dx + \int [E(\hat{f}(x)) - f(x)]^2 dx \end{aligned} \quad (2)$$

باید توجه کرد که $MSE_x(\hat{f})$ تابعی از x است، پس نمی‌تواند ملاک مناسب برای برآورد باشد.

اساس کار در انتخاب ذهنی، به این صورت است که شکل تابع چگالی را با استفاده از چندین پارامتر هموار کننده رسم کرده و سپس در بین این نمودارها، آن نموداری که بیشتر به ساختار داده ها نزدیک است، انتخاب می شود. البته این روش زمانی مفید واقع خواهد شد که شناخت قبلی از داده ها وجود داشته باشد که همیشه اینطور نیست. ترل^۵ [۱۷] نشان داده است که

$$h_{opt} \leq \left[\frac{243 \int K^2(x) dx}{35 k_{\gamma} n} \right]^{\frac{1}{5}} \sigma$$

که در آن σ ، انحراف استاندارد داده ها است که در صورت مجهول بودن باید برآورد آن مورد استفاده قرار گیرد. روش انتخاب ذهنی با این مقدار ماکزیمم، به جای h_{opt} شروع شده و در ادامه h را کاهش می دهند تا به برآورد مورد نظر برسند.

در روش ارجاع به یک توزیع استاندارد برای بدست آوردن h_{opt} ، مقدار $\int f^{*2}(x) dx$ را با فرض اینکه داده ها متعلق به یک خانواده توزیع استاندارد مانند توزیع نرمال استاندارد با واریانس σ^2 باشد، برآورد می کنند. بطوریکه داریم:

$$\int f^{*2}(x) dx = \sigma^{-5} \int \Phi^{*2}(x) dx = \frac{3}{8} \pi^{-\frac{1}{2}} \sigma^{-5} \approx 0.212 \sigma^{-5}$$

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

و بکاربردن $\hat{\sigma}$ به جای σ خواهیم داشت:

$$\hat{h}_{opt} \approx 1/0.6 \hat{\sigma} n^{-\frac{1}{5}} \quad (8)$$

که تابعی نزولی بر حسب n می باشد. به رابطه (۸) اصطلاحاً قانون عملی^۶ اطلاق می شود.

گروه دوم روش های مبتنی بر استفاده از داده ها در برآورد پارامتر هموار کننده هستند و اصطلاحاً به آنها روش انتخاب کاملاً خودکار و سازگار پارامتر هموار کننده^۷ اطلاق می شود. از جمله آنها می توان به روش هایی که بر پایه هم روایی^۸، هستند اشاره نمود، دو روش مهم هم روایی عبارتند از:

رابطه (۶) بیانگر ارتباط معکوس h_{opt} با n و میزان نوسانات تابع چگالی است بدین صورت که هر چه تعداد مشاهدات زیاد یا نوسانات تابع چگالی افزایش یابد، مقدار h_{opt} کاهش پیدا می کند. با جایگذاری رابطه (۶) در (۵) داریم:

$$\inf MISE(\hat{f}) = \frac{5}{4} C(K) \sqrt[5]{\int f^{*2}(x) dx} n^{-\frac{4}{5}} \quad (7)$$

که در آن

$$C(K) = k_{\gamma}^{\frac{2}{5}} \left\{ \int K^2(t) dt \right\}^{\frac{4}{5}}$$

رابطه (۷) زمانی، کمترین مقدار خود را می گیرد که $C(K)$ می نیمم شود. یا به عبارت دیگر تابع هسته K ، باید بگونه ای انتخاب شود که $C(K)$ مربوط به آن می نیمم مقدار را داشته باشد. اپانچنیکوو^۱ [۶] نشان داده است که $C(K)$ بازای تابع هسته $K_e(t)$ ، می نیمم می شود که در آن

$$K_e(t) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}t^2\right) & \text{if } -\sqrt{5} \leq t \leq \sqrt{5} \\ 0 & \text{o.w} \end{cases}$$

بنابراین قبل از بکار بردن یک تابع هسته در برآورد هسته ای، می توان میزان کارایی^۲ آنرا نسبت به هسته $K_e(t)$ ، سنجید. کارایی یک تابع هسته K بصورت زیر تعریف می شود:

$$eff(K) = \left\{ \frac{C(K_e)}{C(K)} \right\}^{\frac{5}{4}} = \frac{3}{5\sqrt{5}} \left\{ \int t^2 K(t) dt \right\}^{-\frac{1}{2}} \left\{ \int K^2(t) dt \right\}^{-1}$$

۴. انتخاب پارامتر هموار کننده

با توجه به رابطه (۶) معلوم می شود که h_{opt} به تابع چگالی مجهول f ، از طریق $\int f^{*2}(x) dx$ بستگی دارد. پس برای بدست آوردن h_{opt} باید این مقدار برآورد شود. روشهای مختلفی برای بدست آوردن h وجود دارد که این روش ها به دو گروه تقسیم می شوند. گروه اول شامل روش هایی می شود که بر پایه شناخت قبلی از داده هاست، که به این روش ها اصطلاحاً برآوردکننده های سریع و ساده می گویند. از جمله آنها می توان به روش انتخاب ذهنی^۳ و روش ارجاع به یک توزیع استاندارد^۴ اشاره نمود.

Terrel – ۵

Rule of thumb – ۶

Fully automatic and consistent bandwidth selectors – ۷

Cross-Validation – ۸

Epanechnikove – ۱

Efficiency – ۲

Subjective choice – ۳

reference to a standard distribution – ۴

$$\int \hat{f}^{\vee}(x) dx = \int \sum_{i=1}^n n^{-1} h^{-1} K\{h^{-1}(x - X_i)\} \\ \sum_{j=1}^n n^{-1} h^{-1} K\{h^{-1}(x - X_j)\} \\ = \frac{1}{n^{\vee} h} \sum_i \sum_j \int K(h^{-1} X_i - u) K(u - h^{-1} X_j) du \\ = \frac{1}{n^{\vee} h} \sum_i \sum_j K^{(\vee)}\{h^{-1}(X_i - X_j)\} \quad (11)$$

از طرف دیگر داریم:

$$\frac{1}{\vee} \sum \hat{f}_{-i}(X_i) = \frac{1}{n} \sum_i (n-1)^{-1} \sum_{j \neq i} h^{-1} K\{h^{-1}(X_i - X_j)\} \\ = \frac{1}{n(n-1)} \sum_i \sum_j h^{-1} K\{h^{-1}(X_i - X_j)\} \\ - (n-1)^{-1} h^{-1} K(\circ) \quad (12)$$

با جایگذاری دو رابطه (11) و (12) در (10) خواهیم داشت:

$$M_{\circ}(h) = \frac{\vee}{n^{\vee} h} \left(\frac{n}{\vee} K^{\vee}(\circ) \right) + \\ \frac{\vee}{n^{\vee} h} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n K^{(\vee)}(h^{-1}(X_j - X_i)) \right. \\ \left. - \frac{\vee n}{n-1} K(h^{-1}(X_j - X_i)) \right] \quad (13)$$

اگر در رابطه (13) با $\frac{1}{n}$ تقریب زده شود شکل دیگری از $M_{\circ}(h)$ بدست می آید که آنرا با $M_{\vee}(h)$ نشان داده و داریم:

$$M_{\vee}(h) = \frac{1}{n^{\vee} h} \sum_i \sum_j K^* \{h^{-1}(X_i - X_j)\} + \frac{\vee}{nh} K(\circ)$$

که در آن

$$K^*(t) = K^{(\vee)}(t) - \vee K(t).$$

استن^۷ [۱۶] نشان داد که اگر X_1, X_2, \dots, X_n یک نمونه تصادفی از یک تابع چگالی مجهول باشد و I_{ISXV} و I_{opt} به ترتیب نمایشگر انتگرال مربع خطا با استفاده از پارامتر هموارکننده بدست آمده باشد که از طریق می نیمم کردن $M_{\vee}(h)$ و $MISE$ بدست آمده است. آنگاه داریم:

الف) روش هم روایی کمترین مربعات^۱
ب) روش هم روایی درستنمایی ماکزیمم^۲.

الف) روش هم روایی کمترین مربعات

این روش که به اختصار با علامت $LSCV$ ، نشان داده می شود، اولین بار توسط رادمو^۳ [۱۲] و باومن^۴ [۳] ارائه شده است. در این روش از خود داده ها، برای برآورد پارامتر هموار کننده مناسب استفاده می شود. در روش $LSCV$ ، از مینیمم کردن انتگرال مربع خطا (ISE) ، جهت برآورد پارامتر هموار کننده، استفاده می کنند. داریم:

$$ISE = \int (\hat{f} - f)^{\vee} = \int \hat{f}^{\vee} - \vee \int \hat{f} \hat{f} + \int f^{\vee} \quad (9)$$

عبارت سوم سمت راست این رابطه یعنی $\int f^{\vee}$ ، به مقدار h بستگی ندارد. پس برای می نیمم کردن این رابطه نسبت به h باید عبارت $R(\hat{f}) = \int \hat{f}^{\vee} - \vee \int \hat{f} \hat{f}$ را نسبت به h می نیمم کرد. از طرفی چون مقدار $R(\hat{f})$ به مقدار مجهول f بستگی دارد ابتدا برآوردی برای آن بدست آورده و بعد برآورد آن را نسبت به h می نیمم می کنند.

هاردل^۵ [۸] نشان داد تابع $M_{\circ}(h)$ که به صورت زیر تعریف می شود، یک برآورد کننده نااریب، برای $E(R(\hat{f}))$ می باشد:

$$M_{\circ}(h) = \int \hat{f}^{\vee} - \frac{\vee}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i) \quad (10)$$

که در آن منظور از $\hat{f}_{-i}(X)$ برآورد تابع چگالی توسط تمامی مشاهدات بجز مشاهده i ام^۶ می باشد و به صورت زیر تعریف می شود:

$$\hat{f}_{-i}(X) = \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K(h^{-1}(X - X_j))$$

برای بدست آوردن یک فرمول روشن تر از $M_{\circ}(h)$ تابع $K^{(\vee)}$ را به عنوان پیشش تابع هسته K با خودش تعریف می کنیم. فرض کنید K یک تابع متقارن باشد با تغییر متغیر $u = h^{-1}x$ داریم:

- 1- Least squares cross-Validation
- 2- Maximum likelihood cross-Validation
- 3- Rudemo
- 4- Buwman
- 5- Hardle
- 6- Leave-one-out estimate

برای بدست آوردن h مناسب، لازم است ماکسیمم متناهی لگاریتم تابع درستنمایی بدست آورده شود. این لگاریتم تابع درستنمایی نسبت عکس با $d_{KL}(f, \hat{f})$ دارد. یعنی می نیمم کردن اطلاع کولبک-لایبلر باعث ماکزیمم شدن لگاریتم تابع درستنمایی خواهد شد.

$$E(CV_{KL}(h)) \approx -E[d_{kl}(f, \hat{f})] + \int f(x) \log(f(x)) dx$$

که در آن

$$CV_{KL}(h) = \frac{1}{n} \log \left(\prod_{i=1}^n \hat{f}_{-i}(X_i) \right)$$

برای جزئیات بیشتر به [۱] مراجعه شود. باید توجه کرد مشکل اصلی در برآورد به روش هسته ای زمانی بوجود می آید که از داده هایی که دارای توزیع دنباله دار باشند استفاده شود. بطو ریکه در برآورد حاصل از این روش، در قسمت های دنباله ای ناهمواری های بیش از حد ایجاد می شود که علت این امر به دلیل ثابت گرفتن مقدار در تمام مکانها است. برای حل چنین مشکلی، روش های دیگر برآورد مانند روش نزدیکترین همسایگی و روش هسته ای توافقی پیشنهاد شده اند که در اینجا آورده می شود.

۵. روش نزدیکترین همسایگی

در این روش، مقدار پارامتر هموار کننده در مکان های مختلف، متفاوت است. میزان همواری چگالی توسط عدد صحیح k کنترل می شود که معمولا $k = \sqrt{n}$ گرفته می شود.

برای واضح تر شدن موضوع فرض کنید فاصله $d(x, y)$ به صورت $d(x, y) = |x - y|$ تعریف شود. فاصله نقطه دلخواه t از مشاهدات را با $d_i(t) = d(t, x_i)$ ، $i = 1, \dots, n$ نشان داده و به ترتیب صعودی به صورت زیر مرتب شده اند:

$$d_1(t) \leq d_2(t) \leq \dots \leq d_n(t)$$

در آن صورت k -امین برآورد نزدیکترین همسایگی در نقطه t به صورت زیر تعریف می شود:

$$\hat{f}(t) = \frac{k}{\sum_{i=1}^k d_k(t)} \quad (14)$$

برای توضیح بیشتر موضوع، فرض کنید چگالی در نقطه t برابر $f(t)$ باشد در آن صورت برای یک نمونه به اندازه n انتظار داریم به تعداد $\sum_{i=1}^k f(t)$ مشاهده در فاصله $[t-r, t+r]$ قرار گیرد بنابراین می توان گفت رابطه (۱۴) بیانگر این واقعیت است که دقیقاً k مشاهده در فاصله $[t - d_k(t), t + d_k(t)]$ قرار بگیرد.

$$\lim_{n \rightarrow \infty} \frac{I_{LSXV}(X_1, X_2, \dots, X_n)}{I_{opt}(X_1, X_2, \dots, X_n)} = 1$$

یعنی بصورت مجانبی، با روش $LSCV$ می توان به بهترین انتخاب ممکن از پارامتر هموار کننده دست یافت.

ب) روش هم روایی درستنمایی ماکسیمم

اساس کار در این روش، بر پایه انجام آزمون زیراست:

$$\begin{cases} H_0: \hat{f}(x) = f(x) \\ H_1: \hat{f}(x) \neq f(x) \end{cases}$$

در حقیقت h مناسب مقداری خواهد بود که مطابق آن $f(x) = \hat{f}(x)$ باشد یعنی مقدار نسبت $\frac{f(x)}{\hat{f}(x)}$ برابر یک شود و یا به عبارت دیگر مقدار $E[\log(\frac{f(x)}{\hat{f}(x)})]$ تحت این h برابر صفر خواهد شد، سیلورمن^۱ [۱۶].

در این روش h طوری انتخاب می شود که باعث کمترین فاصله بین \hat{f} و f با شرط حفظ همواری شود. فاصله ای که در اینجا از آن یاد می شود فاصله متریک نیست و از آن تحت عنوان اطلاع کولبک-لایبلر^۲ یاد می شود که به صورت زیر تعریف می شود:

$$d_{KL}(f, \hat{f}) = \int \log \left(\frac{f}{\hat{f}} \right) (x) f(x) dx.$$

مقدار h ای که از می نیمم کردن $d_{KL}(f, \hat{f})$ بدست می آید مقدار مطلوب خواهد بود.

فرض کنید که یک مجموعه از مشاهدات اضافی که مستقل از بقیه هستند در اختیار ما قرار بگیرد که البته در عمل چنین مشاهداتی وجود ندارند. ولی برای در دست داشتن چنین مشاهداتی هر دفعه یکی از مشاهداتی که در اختیار داریم را حذف کرده و آنرا به عنوان مشاهده مستقل از بقیه در نظر می گیریم و تابع چگالی آن مشاهده حذف شده بصورت زیر تعریف می شود:

$$\hat{f}_{-i}(X_i) = \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K \left(\frac{X_i - X_j}{h} \right)$$

در نتیجه تابع درستنمایی برابر خواهد بود با:

$$\prod_i \hat{f}_{-i}(X_i) = \frac{1}{(n-1)^n h^n} \prod_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K \left(\frac{X_i - X_j}{h} \right).$$

مرحله سوم: برآورد هسته ای توافقی بصورت زیر تعریف می شود:

$$\hat{f}(t) = \frac{1}{\sum_{i=1}^n h^d \lambda_i^d} K\left(\frac{t - X_i}{h \lambda_i}\right)$$

بطوریکه در آن K یک تابع هسته و h پارامتر هموار کننده می باشد.

لازم به ذکر است که برای دستیابی به یک برآورد آزمایشی گفته شده در مرحله اول، باید از یکی از روش های برآورد چگالی مانند روش هسته ای یا روش نزدیکترین همسایگی استفاده شود. بریمن، مسل، پورکل^۶ [۴] در سال و ابرامسون [۲] نشان دادند که روش توافقی، به برآورد اولیه غیر حساس است و لذا می توان از هر روشی برای بدست آوردن $\tilde{f}(t)$ استفاده کرد که برای سادگی بهتر است از روش هسته ای ثابت با پارامتر هموار کننده ای که با ارجاع به توزیع استاندارد بدست آمده، استفاده نمود.

توجه می کنیم که در مرحله دوم با قرار دادن $\alpha = 0$ ، همان روش برآورد هسته ای بدست می آید. ابرامسون [۲] نشان داد که بهترین انتخاب برای α ، 0.5 می باشد.

برآورد حاصل از این روش دنباله های سنگین نخواهد داشت و بعلاوه تمام خصوصیات مشتق پذیری خود را از تابع هسته K ، می گیرد.

۷. برآورد هسته ای چند متغیره

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی n -تایی، d -متغیره از تابع چگالی مجهول f باشد. برآورد چگالی به روش هسته در نقطه دلخواه \mathbf{X} به صورت زیر تعریف می شود:

$$\hat{f}(\mathbf{X}, \mathbf{H}) = n^{-1} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X} - \mathbf{X}_i) \quad (16)$$

که در آن \mathbf{H} یک ماتریس معین مثبت و متقارن $d \times d$ است که به آن ماتریس هموار کننده گفته می شود و تابع K تابع هسته است که خود معمولاً و نه لزوماً تابع چگالی احتمال d -متغیره و به صورت شعاعی متقارن در نظر گرفته می شود و به صورت زیر تعریف می شود:

$$K_{\mathbf{H}}(\mathbf{X}) = |\mathbf{H}|^{-\frac{1}{2}} K\left(\mathbf{H}^{-\frac{1}{2}} \mathbf{X}\right).$$

روش های متفاوتی برای بدست آوردن تابع هسته d -متغیره به کمک تابع هسته یک متغیره متقارن K_0 وجود دارد که از جمله

یک حالت تعمیم یافته از این روش که به k امین نزدیکترین همسایگی تعمیم یافته معروف است به صورت زیر تعریف می شود:

$$\mathbf{X} = (X_1, \dots, X_d), \mathbf{X}_i = (X_{i_1}, \dots, X_{i_d})$$

که در آن $d_k(t)$ پارامتر همواره کننده و K تابع هسته می باشد. برآورد های حاصل از روش نزدیکترین همسایگی، برآورد های چندان رضایت بخشی نیستند، زیرا آنها بیشتر متمایل به ناهمواری های مکانی و دنباله های سنگین با انتگرال نامتناهی هستند. به عبارت دیگر اگر چه روش هسته ای باعث همواری کم در دنباله توزیعها می شود اما این روش بیش از حد این ناهمواری ها را رفع کرده و برآوردهایی را نتیجه می دهد که لزوماً در هر جایی مشتق پذیر نیستند. بدین منظور روش هسته ای توافقی مطرح می شود.

۶. روش برآورد هسته ای توافقی^۲

هدف کلی در این بخش بیان کلاسی از برآوردهاست که در حقیقت ترکیبی از روش هسته ای و روش نزدیکترین همسایگی می باشد. ایده اساسی ساختن یک برآورد هسته ای است که شامل برآمدگیهایی روی مشاهدات باشد بگونه ای که پارامتر هموار کننده از نقطه ای به نقطه دیگر تغییر کند.

فرض کنید مشاهدات در یک فضای d -بعدی قرار گرفته اند، در برآورد هسته ای توافقی، مراحل زیر باید انجام شود :

مرحله اول: پیدا کردن یک برآورد اولیه^۳ $\tilde{f}(t)$ بطوریکه برای هر f_i

$$\hat{f}(X_i) > 0$$

مرحله دوم: در نظر گرفتن عوامل هموار کننده مکانی^۴ λ_i که بصورت زیر تعریف می شوند:

$$\lambda_i = \left\{ \tilde{f}(X_i) / g \right\}^{-\alpha}$$

که در آن g میانگین هندسی $\tilde{f}(X_i)$ ها می باشد بطوریکه برای هر i :

$$\log g = \frac{1}{p} \sum \log \tilde{f}(X_i)$$

و α یک عدد در بازه $[0, 1]$ است که پارامتر حساسیت^۵ نامیده می شود.

۱- Generalized kth nearest neighbour

۲- Adaptive kernel

۳- Pilot estimate

۴- Local bandwidth factors

۵- Sensitivity parameter

برای این منظور فرض کنید تابع هسته K ، تابعی به طور شعاعی متقارن و تابع چگالی f مشتق دوم پیوسته داشته باشد، ثابتهای α و β را به صورت زیر تعریف می کنیم:

$$\alpha = \int_{R^d} K(\mathbf{t}) dt, \quad \beta = \int_{R^d} K^\vee(\mathbf{t}) dt$$

که در آن $\mathbf{t} = (t_1, t_2, \dots, t_n)$ نقطه ای است که برآورد چگالی در آن مورد نظر است.

با استفاده از قضیه تیلور برای حالت چند متغیره، مقادیر تقریبی برای اریبی و واریانس نمونه برابر خواهد شد با:

$$E(\hat{f}(\mathbf{X})) - f(\mathbf{X}) \approx \frac{1}{2} h^\vee \alpha \nabla^\vee f(\mathbf{X})$$

$$V(\hat{f}(\mathbf{X})) \approx n^{-1} h^{-d} \beta f(\mathbf{X})$$

که در آن

$$\nabla^\vee f(\mathbf{X}) = \sum_{i=1}^d \frac{\partial^\vee}{\partial^\vee X_i} f(\mathbf{X})$$

با جایگذاری روابط اخیر در رابطه مربوط به میانگین انتگرال مربع خطا خواهیم داشت:

$$MISE(\hat{f}) \approx \frac{1}{4} h^\vee \alpha^\vee \int_{R^d} \left\{ \nabla^\vee f(\mathbf{X}) \right\}^2 d\mathbf{X} + \frac{\beta}{nh^d}.$$

که با می نیم کردن این عبارت نسبت به h داریم:

$$h_{opt}^{d+\vee} = \frac{d\beta}{\alpha^\vee n} \left\{ \int_{R^d} \left(\nabla^\vee f \right)^2 \right\}^{-1} \quad (17)$$

رابطه (۱۷) بیانگر رابطه h_{opt} با تابع چگالی مجهول f از طریق $\nabla^\vee f$ می باشد. از طرفی طبق این رابطه دیده می شود که با افزایش حجم نمونه، مقدار پارامتر هموارکننده به صفر میل می کند اما سرعت این همگرایی با نرخ $\left(-\frac{1}{d+\vee}\right)$ کاهش می یابد. بنابراین به نظر می رسد که روش هسته ای در ابعاد بالاتر مناسب نباشد.

۸. کاربرد برآورد تابع چگالی به روش هسته ای

روش برآورد هسته ای کاربردهای زیادی در آنالیز داده های آماری دارد که از جمله این کاربردها می توان به برآورد تابع رگرسیون ناپارامتری، تابع تشخیص، تابع توزیع، ... اشاره کرد ([۷] و [۱۵]). از جمله روشهای برآورد تابع رگرسیون ناپارامتری مبتنی

آنها می توان به تابع هسته ضربی اشاره کرد که با K^P نشان داده و به صورت زیر تعریف می شود:

$$K^P(\mathbf{X}) = \prod_{i=1}^d K_o(X_i), \quad \mathbf{X} = (X_1, X_2, \dots, X_d)$$

یک انتخاب معروف برای K ، تابع چگالی نرمال استاندارد d - متغیره است:

$$K(\mathbf{X}) = (\sqrt{\pi})^{-d} \exp\left(-\frac{1}{2} \mathbf{X}^T \mathbf{X}\right).$$

برای ساده تر شدن محاسبات اپانچنیکوو [۶] ماتریس \mathbf{H} را یک ماتریس قطری به صورت

$$\mathbf{H} = \text{diag}(h_1^\vee, \dots, h_d^\vee)$$

در نظر گرفت. در نتیجه خواهیم داشت:

$$\hat{f}(\mathbf{X}, \mathbf{H}) = \frac{1}{n \prod_{l=1}^d h_l} \sum_{i=1}^n K\left(\frac{X_1 - X_{i1}}{h_1}, \dots, \frac{X_d - X_{id}}{h_d}\right),$$

که در آن هر $i = 1, 2, \dots, n$

$$\mathbf{X} = (X_1, \dots, X_d), \quad \mathbf{X}_i = (X_{i1}, \dots, X_{id})$$

نمایش ساده دیگر از ماتریس \mathbf{H} که کاکولوس [۵] در نظر گرفت به صورت زیر است:

$$\mathbf{H} = \text{diag}(h_1^\vee, \dots, h_d^\vee)$$

یعنی برای هر i داریم: $h_i^\vee = h^\vee$

در نتیجه خواهیم داشت:

$$\hat{f}(\mathbf{X}, h) = \frac{1}{nh^d} \sum_i K\left(\frac{\mathbf{X} - \mathbf{X}_i}{h}\right).$$

باید متذکر شد در برخی مواقع، بخصوص زمانی که پراکندگی داده ها در یکی از محورها بیشتر از بقیه باشد، بهتر است به جای استفاده از پارامتر هموار کننده یکسان در همه محورها، از یک بردار پارامتر هموار کننده یا حتی یک ماتریس غیر قطری استفاده شود. به منظور پیشگیری از مواجه شدن با فرمولهای پیچیده تر در اثر عدم استفاده از پارامتر هموار کننده یکسان، بهتر است ابتدا داده ها از پیش مقیاس بندی شود.

باید توجه کرد که در حالت چند متغیره نیز در نظر گرفتن مقادیر بزرگ از مؤلفه های پارامتر هموار کننده، باعث همواری بیش از حد برآورد، خواهد شد. همانند حالت یک متغیره درصدد یافتن تابع هسته و پارامتر هموار کننده بهینه هستیم بطوریکه باعث کوچک شدن میانگین انتگرال مربع خطا گردد ([۹] و [۱۰]).

نقطه خاصی با برازش مکانی چند جمله ای درجه p - ام به داده ها برآورد می کنند. این کلاس در حالت خاص ($p = 0$) شامل برآورد ناداریا - واتسون می شود. استفاده از روش برآورد هسته، کاربردهای دیگری نیز دارد که بحث مستقلی را طلب می کنند.

بر هموار سازی هسته ای می توان به برآورد ناداریا - واتسون^۱ اشاره کرد، کلاس دیگری از برآورد کننده های تابع رگرسیون ناپارامتری مبتنی بر هموارسازی هسته ای وجود دارد که اصطلاحاً به آن « برآورد کننده هسته چند جمله ای مکانی » اطلاق می شود (واند و جونز^۲ [۱۸]). روش کار این نوع برآورد کننده های تابع رگرسیون، به این صورت است که تابع رگرسیون را با روش حداقل مربعات وزنی در یک

مراجع

- [1] افشاری، ر، (۱۳۸۳). پایان نامه کارشناسی ارشد، دانشکده علوم ریاضی و کامپیوتر، گروه آمار.
- [2] Abramson, I. S., 1982. On bandwidth variation in kernel estimates-a square root law. *Ann. Statist.*, No. 10, 1217-1223.
- [3] Bowman, A. W., 1984. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, No. 71, 353-360.
- [4] Breiman, L., Meisel, M. and Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, No. 19, 135-144.
- [5] Cacoullous, T., 1966. Estimation of a multivariate density. *Ann Inst. Statist. Math.*, No. 18, 179-189.
- [6] Epanechnikov, V. A., 1969. Nonparametric estimation of a multidimensional probability density. *Theor. Probab. Appl.*, No. 14, 153-158.
- [7] Fix, E. and Hodges, J. L., 1951. *Discriminatory analysis nonparametric discrimination: Consistency properties*. Report No. 4, project no 21-29-004, USAF School of Aviation Medicine, Randolph Field, Texas.
- [8] Hardle, W., 1990. *Smoothing Techniques with Implementation in S*. NewYork: Springer-Verlag.
- [9] Kreider, D. L., Kuller, R. G., Ostberg, D. R. and Perkins, F. W., 1966. *An Introduction to Linear Analysis*. Reading, Mass: Addison-Wesley.
- [10] Loftsgaarden, D. O. and Quesenberry, C. P., 1965. A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, No. 36, 1049-1051.
- [11] Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, No. 27, 832-837.
- [12] Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, No. 9, 65-78.
- [13] Scott, D. W., 1992. *Multivariate Density Estimation: Theory, Practice and Visualization*. NewYork: Wiley.
- [14] Silverman, B. W., 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- [15] Simonoff, J. S. 1996. *Smoothing Methods in Statistics*. NewYork: Springer-Verlag.
- [16] Stone, C. J., 1984. An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, No. 12, 1285-1297.
- [17] Terrell, G. R., 1990. The maximal smoothing principle in density estimation. *J. Am. Statist. Ass.*, No. 85, 470-477.
- [18] Wand, M. P. and Jones, M. C., 1995. *Kernel Smoothing*. London: Chapman & Hall.