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[۳]، 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 $X_1, X_2, \ldots, X_n$، 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:

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

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:

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

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

۱- گروه آمار، دانشگاه شهید چمران اهواز

۲- دانشجوی کارشناسی ارشد آمار، دانشگاه شهید چمران

۳- Histogram

۴- Nearest- neighbour

۵- Orthogonal sereis

۶- Kernel

۷- Smoothing parametet or bandwith

۸- Kernel function

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

مشاهده، همراه با مشاهدات همسایه ساخته می شود. لازم به ذکر است که در برآورد هسته ای تابع K، شکل برآمدگیها و مقدار h، عرض برآمدگیها را نمایش می دهد. به عبارت دیگر h، پارامتری است که مقدار زیاد آن باعث هموواری زیاد[1] و مقدار کم آن باعث ناهمواری زیاد[2] می شود.

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

۳. مقیاسهای تغییرپذیری

اگر چه در حالت پارامتری برآوردهای نااریب برای پارامترهای توزیع هایی مانند نرمال، پواسن، نمایی، هندسی، ... وجود دارند، اما روزن بلات[3] [۱۱] نشان داد که در حالت ناپارامتری هیچ برآورد کننده نااریبی وجود ندارد. بنابراین باید در بین دنباله ای از برآورد کننده های بطور مجانبی نااریب به جستجوی بهترین برآورد پرداخته شود.

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

$$\lim_{n\to\infty} E\left(\hat{f}(x)\right)=f(x).$$

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:

$$\begin{aligned} MSE_x\left(\hat{f}\right) &= E\left(\hat{f}(x)-f(x)\right)^2 \\ &= V\left(\hat{f}(x)\right)+\left[E\left(\hat{f}(x)\right)-f(x)\right]^2 \end{aligned} \qquad (۱)$$

همچنین

$$MISE_x\left(\hat{f}\right)=E\left(\hat{f}(x)-f(x)\right)^2 dx$$

و چون تابع زیر انتگرال، یک تابع مثبت است داریم:

$$\begin{aligned} &= \int E\left(\hat{f}(x)-f(x)\right)^2 dx \\ &= \int V\left(\hat{f}(x)\right)dx+\int\left[E\left(\hat{f}(x)\right)-f(x)\right]^2 dx \end{aligned} \qquad (۲)$$

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

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 می کنیم.

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

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

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

$$\begin{aligned} Bias_h(x) &= \frac{1}{2}h^2 f''(x)k_2+o\left(h^3\right) \\ &\approx \frac{1}{2}h^2 f''(x)k_2 \end{aligned}$$

در نتیجه:

$$\int\left(Bias_h(x)\right)^2 dx \approx \frac{1}{4}h^4 k_2^2 \int f''^2(x)dx \qquad (۳)$$

و به همین ترتیب خواهیم داشت:

$$V\left(\hat{f}(x)\right)\approx n^{-1}h^{-1}f(x)\int K^2(x)dx \qquad (۴)$$

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

$$MISE_{(x)}\left(\hat{f}\right)=\frac{1}{4}h^4 k_2^2(x)dx+n^{-1}h^{-1}\int K^2(t)dt. \qquad (۵)$$

هدف می نیمم کردن $MISE$ نسبت به h می باشد. برای اینکه این مقدار می نیمم 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 مولفه 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:

$$h=h_{opt}=k_2^{-\frac{2}{5}}\left\{\int K^2(t)dt\right\}^{\frac{1}{5}}\left\{\int f''^2(x)dx\right\}^{-\frac{1}{5}}n^{-\frac{1}{5}} \qquad (۶)$$

که در آن $\int f''^2(x)dx$ را میزان نوسانات تابع چگالی تعریف می کنند.

۱- Over smoothing
۲- Under smoothing
۳- Rosenblatt

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

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 می‌کند. با جایگذاری رابطه (۶) در (۵) داریم:

$$\inf MISE\left(\hat{f}\right)=\frac{5}{4}C(K)\ \sqrt[5]{\int f''^{2}(x)\,dx}\ \ n^{-\frac{4}{5}} \qquad (7)$$

که در آن

$$C(K)=k_2^{\frac{2}{5}}\left\{\int K^{2}(t)dt\right\}^{\frac{4}{5}}$$

رابطه (۷) زمانی، کمترین مقدار خود را می‌گیرد که $C(K)$ می‌نیمم شود. یا به عبارت دیگر تابع هسته K، باید بگونه‌ای انتخاب شود که $C(K)$ مربوط به آن می‌نیمم مقدار را داشته باشد. اپانچنیکوو[1] [۶] نشان داده است که $C(K)$ بازای تابع هسته $K_e(t)$، می‌نیمم می‌شود که در آن

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

بنابراین قبل از بکار بردن یک تابع هسته در برآورد هسته‌ای، می‌توان میزان کارائی[2] آنرا نسبت به هسته $K_e(t)$، سنجید. کارائی یک تابع هسته K بصورت زیر تعریف می‌شود:

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

٤. 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[3] و روش ارجاع به یک توزیع استاندارد[4] اشاره نمود.

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[5] [۱۷] نشان داده است که

$$h_{opt}\le\left[\frac{243\int K^{2}(x)dx}{35k_2\,n}\right]^{\frac{1}{5}}\sigma$$

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:

$$\int f''^{2}(x)dx=\sigma^{-5}\int\Phi''^{2}(x)dx$$
$$=\frac{3}{8}\pi^{-\frac{1}{2}}\sigma^{-5}\approx 0/212\sigma^{-5}$$

با بکارگیری تابع هسته گوسین، $K(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^{2}\right)$

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

$$\hat{h}_{opt}\approx 1/06\,\hat{\sigma}\,n^{-\frac{1}{5}} \qquad (8)$$

که تابعی نزولی بر حسب n می باشد. به رابطه (۸) اصطلاحاً قانون عملی[6] 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[7] اطلاق می شود. از جمله آنها می توان به روش هایی که بر پایه هم روایی[8]، هستند اشاره نمود، دو روش مهم هم روایی عبارتند از:

۱- Epanechnikove

۲- Efficiency

۳- Subjective choice

۴- refrence to a standard distribution

۵- Terrel

۶- Rule of thumb

۷- Fully automatic and consistent bandwidth selectors

۸- Cross-Validation

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

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:

$$ISE=\int\left(\hat{f}-f\right)^{۲}=\int\hat{f}^{۲}-۲\int\hat{f}f+\int f^{۲} \qquad (۹)$$

عبارت سوم سمت راست این رابطه یعنی $\int f^{۲}$، به مقدار h بستگی ندارد. پس برای می نیمم کردن این رابطه نسبت به h باید عبارت $R(\hat{f})=\int\hat{f}^{۲}-۲\int\hat{f}f$ 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 نا اریب، برای $E(R(\hat{f}))$ می باشد:

$$M_{\circ}(h)=\int\hat{f}^{۲}-\frac{۲}{n}\sum_{i=۱}^{n}\hat{f}_{-i}(X_i) \qquad (۱۰)$$

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:

$$\hat{f}_{-i}(X)=\frac{۱}{(n-۱)h}\sum_{\substack{j=۱\\ j\neq i}}^{n}K\left(h^{-۱}\left(X-X_j\right)\right).$$

برای بدست آوردن یک فرمول روشن تر از $M_{\circ}(h)$ تابع $K^{(۲)}$ را به عنوان پیچش تابع هسته K با خودش تعریف می کنیم. فرض کنید K یک تابع متقارن باشد با تغییر متغیر $u=h^{-۱}x$ داریم:

$$\begin{aligned}\int\hat{f}^{۲}(x)dx&=\int\sum_{i=۱}^{n}n^{-۱}h^{-۱}K\left\{h^{-۱}\left(x-X_i\right)\right\}\\&\quad\sum_{j=۱}^{n}n^{-۱}h^{-۱}K\left\{h^{-۱}\left(x-X_j\right)\right\}\\&=\frac{۱}{n^{۲}h}\sum_i\sum_j\int K\left(h^{-۱}X_i-u\right)K\left(u-h^{-۱}X_j\right)du\\&=\frac{۱}{n^{۲}h}\sum_i\sum_j K^{(۲)}\left\{h^{-۱}\left(X_i-X_j\right)\right\}\end{aligned} \qquad (۱۱)$$

از طرف دیگر داریم:

$$\begin{aligned}\frac{۱}{۲}\sum\hat{f}_{-i}(X_i)&=\frac{۱}{n}\sum_i(n-۱)^{-۱}\sum_{j\neq i}h^{-۱}K\left\{h^{-۱}\left(X_i-X_j\right)\right\}\\&=\frac{۱}{n(n-۱)}\sum_i\sum_j h^{-۱}K\left\{h^{-۱}\left(X_i-X_j\right)\right\}\\&\quad-(n-۱)^{-۱}h^{-۱}K(\circ)\end{aligned} \qquad (۱۲)$$

با جایگذاری دو رابطه (۱۱) و (۱۲) در(۱۰) خواهیم داشت:

$$\begin{aligned}M_{\circ}(h)&=\frac{۲}{n^{۲}h}\left(\frac{n}{۲}K^{۲}(\circ)\right)+\\&\quad\frac{۲}{n^{۲}h}\left[\sum_{i=۱}^{n-۱}\sum_{j=i+۱}^{n}K^{(۲)}\left(h^{-۱}\left(X_j-X_i\right)\right)\right.\\&\quad\left.-\frac{۲n}{n-۱}K\left(h^{-۱}\left(X_j-X_i\right)\right)\right]\end{aligned} \qquad (۱۳)$$

اگر در رابطه (۱۳) $\frac{۱}{n-۱}$ با $\frac{۱}{n}$ تقریب زده شود شکل دیگری از $M_{\circ}(h)$ بدست می آید که آنرا با $M_{۱}(h)$ نشان داده و داریم:

$$M_{۱}(h)=\frac{۱}{n^{۲}h}\sum_i\sum_j K^{*}\left\{h^{-۱}\left(X_i-X_j\right)\right\}+\frac{۲}{nh}K(\circ)$$

که در آن

$$K^{*}(t)=K^{(۲)}(t)-۲K(t).$$

استن[۷] [۱۶] نشان داد که اگر $X_{۱},X_{۲},\ldots,X_n$ 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 $M_{۱}(h)$ و $MISE$ بدست آمده است. آنگاه داریم:

۱- Least squars cross-Validation

۲- Maximum likelihood cross-Validation

۳- Rudemo

۴- Buwman

۵- Hardle

۶- Leave-one-out estimate

۷- Stone

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

$$\lim_{n\to\infty}\frac{I_{ISXV}(X_1,X_2,\ldots,X_n)}{I_{opt}(X_1,X_2,\ldots,X_n)}=1$$

یعنی بصورت مجانبی، با روش $LSCV$ می توان بـه بهتـرین انتخـاب 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 حقیقـت h مناسـب مقـداری خواهـد بـود کـه مطـابق آن $f(x)=\hat{f}(x)$ باشد یعنی مقدار نسبت $\frac{f(x)}{\hat{f}(x)}$ برابر یک شـود و یا به عبارت دیگر مقدار $E\left[\log(\frac{f(x)}{\hat{f}(x)})\right]$ تحت ایـن h برابـر صـفر خواهد شد، سیلورمن[1] [۱۶].

در این روش h طوری انتخاب می شود که باعث کمترین فاصله بـین $\hat{f}$ و f با شرط حفظ همواری شود. فاصله ای که در اینجا از آن یاد می شود فاصله متریک نیـست و از آن تحـت عنـوان اطـلاع کولبـک-لایبلر[2] یاد می شود که به صورت زیر تعریف می شود:

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

مقـدار h ای کـه از مـی نـیمم کـردن $d_{KL}(f,\hat{f})$ بدسـت می آید مقدار مطلوب خواهد بود.

فرض کنید که یک مجموعه از مشاهدات اضـافی کـه مـستقل از بقیـه هستند در اختیار ما قرار بگیرد کـه البتـه در عمـل چنـین مـشاهداتی وجود ندارند. ولی برای در دست داشتن چنین مشاهداتی هر دفعه یکی از مشاهداتی کـه در اختیـار داریـم را حـذف کـرده و آنـرا بـه عنـوان مشاهده مستقل از بقیه در نظر می گیریم و تـابع چگـالی آن مـشاهده حذف شده بصورت زیر تعریف می شود:

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

در نتیجه تابع درستنمایی برابر خواهد بود با:

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

بـرای بدسـت آوردن h مناسـب، لازم اسـت ماکـسیمم متنـاهی لگاریتم تـابع درسـتنمایی بدسـت آورده شـود. ایـن لگـاریتم تـابع درستنمایی نسبت عکس بـا $d_{KL}(f,\hat{f})$ دارد. یعنـی مـی نـیمم کردن اطلاع کولبک–لایبلـر باعـث مـاکزیمم شـدن لگـاریتم تـابع 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 بـرآورد بـه روش هـسته ای زمـانی بوجود می آید که از داده هایی کـه دارای توزیـع دنبالـه دار باشـند استفاده شود. بطو ریکه در برآورد حاصـل از ایـن روش، در قـسمت های دنباله ای ناهمواری های بیش از حد ایجاد می شود کـه علـت این امر به دلیل ثابت گرفتن مقدار در تمام مکانها است. بـرای حـل چنین مـشکلی، روش هـای دیگـر بـرآورد ماننـد روش نزدیکتـرین همسایگی و روش هسته ای توافقی پیشنهاد شده اند کـه در اینجـا آورده می شود.

۵. روش نزدیکترین همسایگی

در این روش، مقدار پارامتر هموار کننـده در مکـان هـای مختلـف، متفاوت است. میزان همواری چگالی توسط عدد صحیح k کنتـرل می شود که معمولا $k=\sqrt{n}$ گرفته می شود.

برای واضح‌تر شدن موضوع فرض کنید فاصله $d(x,y)$ به صـورت $d(x,y)=|x-y|$ تعریـف شـود. فاصـله نقطـه دلخـواه t از مشاهدات را با $d_i(t)=d(t,x_i)$، $i=1,\ldots,n$ نشان داده و بـه ترتیب صعودی به صورت زیر مرتب شده اند:

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

در آن صورت k-امین برآورد نزدیکترین همسایگی در نقطـه t بـه صورت زیر تعریف می شود:

$$\hat{f}(t)=\frac{k}{2nd_k(t)} \qquad (۱۴)$$

برای توضیح بیشتر موضوع، فـرض کنیـد چگـالی در نقطـه t برابـر $f(t)$ باشد در آن صورت برای یک نمونه به اندازه n انتظار داریم به تعداد $2rnf(t)$ مشاهده در فاصله $[t-r,t+r]$ قـرار گیـرد بنابراین می‌توان گفت رابطـه (۱۴) بیـانگر ایـن واقعیـت اسـت کـه دقیقاً k مشاهده در فاصله $[t-d_k(t),t+d_k(t)]$ قرار بگیرد.

۱- Silverman

۲- Kullback-Leibler information

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

یک حالت تعمیم یافته از این روش که به k امین نزدیکترین همسایگی تعمیم یافته[1] معروف است به صورت زیر تعریف می‌شود:

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

که در آن $d_k(t)$ پارامتر همواره کننده و K تابع هسته می‌باشد.

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[2]

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[3] $\tilde{f}(t)$ بطوریکه برای هر فی

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

مرحله دوم: در نظر گرفتن عوامل هموار کننده مکانی[4] 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 :

$$\log g=\frac{1}{2}\sum\log\tilde{f}(X_i)$$

و α یک عدد در بازه $[0,1]$ است که پارامتر حساسیت[5] نامیده می شود.

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

$$\hat{f}(t)=\frac{1}{2}\sum_{i=1}^{n}\frac{1}{h^d\ \lambda_i^d}K\left(\frac{t-X_i}{h\lambda_i}\right)$$

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[6] [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، ۰/۵ 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 $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$ 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:

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

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

۱- Generalized kth nearest neighbour

۲- Adaptive kernel

۳- Pilot estimate

۴- Local bandwidth factors

۵- Sensitivity parameter

۶- Breiman, Meisel and Purcell

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

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

$$K^P(\mathbf{X})=\prod_{i=1}^{d} K_0(X_i) \quad , \quad \mathbf{X}=(X_1,X_2,\ldots,X_d)$$

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

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

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

$$\mathbf{H}=diag\left(h_1^2,\ldots,h_d^2\right)$$

در نظر گرفت. در نتیجه خواهیم داشت:

$$\hat{f}(\mathbf{X},\mathbf{H})=\frac{1}{n\prod_{l=1}^{d}h_l}\sum_{i=1}^{n}K\left(\frac{X_1-X_{i_1}}{h_1},\ldots,\frac{X_d-X_{i_d}}{h_d}\right),$$

که در آن برای هر $i=1,2,\ldots,n$

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

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

$$\mathbf{H}=diag\left(h_1^2,\ldots,h_d^2\right)$$

یعنی برای هر i داریم: $h_i^2=h^2$

در نتیجه خواهیم داشت:

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

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]).

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:

$$\alpha=\int_{R^d}t_1^2K(\mathbf{t})\,d\mathbf{t} \quad , \quad \beta=\int_{R^d}K^2(\mathbf{t})\,d\mathbf{t}$$

که در آن $\mathbf{t}=(t_1,t_2,\ldots,t_n)$ 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:

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

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

که در آن

$$\nabla^2 f(\mathbf{X})=\sum_{i=1}^{d}\frac{\partial^2}{\partial^2 X_i}f(\mathbf{X})$$

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

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

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

$$h_{opt}^{d+4}=\frac{d\beta}{\alpha^2 n}\left\{\int\left(\nabla^2 f\right)^2\right\}^{-1} \qquad (۱۷)$$

رابطه (۱۷) بیانگر رابطه h_{opt} با تابع چگالی مجهول f از طریق $\nabla^2 f$ 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 $(-\frac{1}{d+4})$ 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

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]

بر هموار سازی هسته ای می توان به برآورد ناداریا – واتسون[1] 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[2] [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 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.

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] Cacoullos, 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.

۱- Wand and Jones
۲- Nadarya-Watson

[Downloaded from andisheyeamari.irstat.ir on 2024-05-05]