دانشکده علوم ریاضی

گروه آمار

پایان‌نامه کارشناسی ارشد

مطالعه روش‌های تعیین پهنای باند در برآورد کرنل برای داده‌های اختر-آماری

محبوبه میرزایی

استاد راهنما

دکتر محمد آرشی

دی ۱۳۹۴

خاضعانه و خالصانه تقدیم می گردد به محضر نورانی

بانوی دو عالم

سپاس‌گزاری

سپاس خدای را عزوجل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت هر نفسی فرو می‌رود ممد حیات است و چون برمی‌آید مفرح ذات پس در هر نفسی دو نعمت موجود است و به هر نعمتی شکری واجب.

در آغاز وظیفه خود می‌دانم از زحمات بی‌دریغ استاد راهنمای خود، جناب آقای دکتر محمد آرشی، تشکر و قدردانی نمایم.

همچنین وظیفه خود می‌دانم از تمامی اساتید محترم گروه آمار دانشگاه شاهرود آقایان دکترداود شاهسونی، دکتراحمد نزاکتی رضازاده، دکترمحمدرضا ربیعی، دکترحسین باغیشنی و خانم دکتر نگار اقبال قدر دانی به عمل بیاورم.

صمیمانه از همسرم به خاطر همراهی ایشان و مهربانان زندگیم، پدر و مادر عزیزم و فرزندان خوبم کمال تشکر را دارم.

وظیفه خود می‌دانم از دوست گرانقدرم خانم مینا نوروزی‌راد به‌واسطه همراهی و هم‌فکری‌شان تشکر نمایم.

در پایان از تمامی دانشجویان آمار و ریاضی دانشکده ریاضی و همچنین از مسئولین محترم آموزش دانشکده به پاس لطف و محبت‌هایشان قدر دانی می‌نمایم.

محبوبه میرزایی
دی ۱۳۹۴

تعهدنامه

اینجانب **محبوبه میرزایی** دانشجوی کارشناسی ارشد رشته **آمار ریاضی دانشکده علوم ریاضی** دانشگاه شاهرود، نویسنده پایان‌نامه با عنوان **مطالعه روش‌های تعیین پهنای باند در برآورد کرنل**، تحت راهنمایی **دکتر محمد آرشی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود متعلق دارد، و مقالات مستخرج با نام «دانشگاه شاهرود» یا "Shahrood University of Technology" به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

محبوبه میرزایی
دی ۱۳۹۴

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

تابع احتمال مفهومی اساسی است. که به وسیله آن می‌توان به رفتار متغیرهای تصادفی پی برد. بدون دانستن تابع احتمال آماره‌های حاصل از یک نمونه تصادفی، امکان انجام استنباط بر روی نمونه‌های تصادفی و تعمیم نتایج حاصله به جامعه وجود ندارد. یکی از روش‌های برآورد تابع احتمال، روش برآورد کرنل می‌باشد که تعمیمی از برآوردگر ساده مستطیلی است. در برآورد به روش کرنل باید دو پارامتر تابع کرنل و پهنای باند را تعیین کرد. برخلاف تابع کرنل که اهمیت کمی دارد، تعیین پهنای باند در این روش از اهمیت بسزایی برخوردار است. کوچک کردن پهنای باند، موجب می‌شود که منحنی حاصل از برآورد کرنل ناهموارتر شده و جزئیات جعلی بیشتری را از تابع احتمال واقعی به نمایش گذارد و بزرگ کردن پهنای باند، موجب می‌شود که منحنی هموار گشته و باعث محو شدن جزئیات واقعی تابع احتمال گردد. لذا انتخاب پهنای باند مناسب، مهمترین مسئله در برآورد تابع احتمال به روش کرنل است. برخی از روش‌های انتخاب پهنای باند (پارامتر هموارسازی) عبارتند از: انتخاب ذهنی، قانون مقیاس نرمال، اعتبارسنجی متقابل کمترین توان‌های دوم و غیره. در این پایان‌نامه درصدد هستیم تا با بررسی موضوع برآورد تابع چگالی احتمال به روش کرنل، برخی از روش‌های انتخاب پهنای باند را مقایسه کرده و مثال واقعی از علم اخترشناسی را برای مقایسه روش‌ها به‌کار ببریم.

کلمات کلیدی: تابع کرنل، پهنای باند، اختر آمار، اختر شناسی متقابل، اخترشناسی، معیار خطای توان دوم تجمعی

پیش گفتار

در دهه‌های گذشته شاهد تغییر باورنکردنی در نظریه‌ی آمار بوده‌ایم. پیشرفت عظیم در توان کامپیوترها، باعث به وجودآمدن شیوه‌های تابعی سازگاری شده که اکنون ابزار ضروری برای تحلیل پیشرفته‌ی داده‌ها است. حال که آماردانان از فرض‌های سخت گیرانه‌ای که جزء بی‌چون و چرای مدل‌های پارامتری کلاسیک هستند، رهایی یافته‌اند، از آنان انتظار می‌رود که نه تنها قادر به برگزیدن متغیرهای مهم در یک مدل باشند، بلکه بتوانند شکل تابعی وابستگی این متغیرها را نیز تخمین بزنند. در این پایان نامه به معرفی یک روش ناپارامتری برای برآورد تابع چگالی با عنوان کرنل می‌پردازیم که وابسته به انتخاب پارامتر هموارسازی می‌باشد. سعی می‌کنیم برخی روش‌های تعیین پارامتر هموارسازی را مطرح کرده و کارآیی آنها را در قالب‌های شبیه‌سازی و مثال واقعی مقایسه کنیم.

این مجموعه، شامل ۴ فصل و یک پیوست می‌باشد. در پیوست، برنامه‌های کامپیوتری مورد استفاده در این پایان‌نامه آمده است و ساختار فصول ارائه‌شده به‌قرار زیر می‌باشند:

در فصل اول به بیان مقدماتی لزوم تعیین تابع چگالی و تعاریف مفاهیم ریاضی مورد نیاز خواهیم پرداخت. در فصل دوم به بیان برخی روش‌های برآورد تابع چگالی از جمله به بررسی روش‌های هیستوگرام، برآورد ساده و کرنل پرداخت. فصل سوم در برگیرنده‌ی برخی روش‌های تعیین پهنای باند بوده که کارایی آنها از طریق مطالعات شبیه‌سازی با یکدیگر مقایسه شده است. در فصل چهارم با استفاده از داده‌های واقعی اختری سعی درتعیین پهنای باند و برآورد تابع چگالی آن کرده و در پایان، پس از نتیجه‌گیری آینده‌ی تحقیق را مورد بررسی قرار می‌دهیم.

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

مقدمه

فرض کنید $X = (X_۱, X_۲, \ldots, X_n)$ یک نمونه تصادفی از جامعه‌ای با تابع توزیع تجمعی $F(.)$ و تابع چگالی $f(.)$ باشد. همچنین فرض کنید علاقه‌مندیم پارامتر موردنظر جامعه را با استفاده از آماره $T(X)$ برآورد کنیم. در این صورت در انجام استنباط بر روی نمونه و تعمیم آن به جامعه، چهار حالت ممکن است پیش آید:

(الف) دانستن توزیعی که از آن نمونه تصادفی $X_۱$، $X_۲$، ...، X_n گرفته شده است و از روی آن توزیع آماره $T(X)$ معلوم باشد. در این حالت مشکلی بر انجام استنباط به کمک آماره $T(X)$ از نمونه به جامعه وجود ندارد.

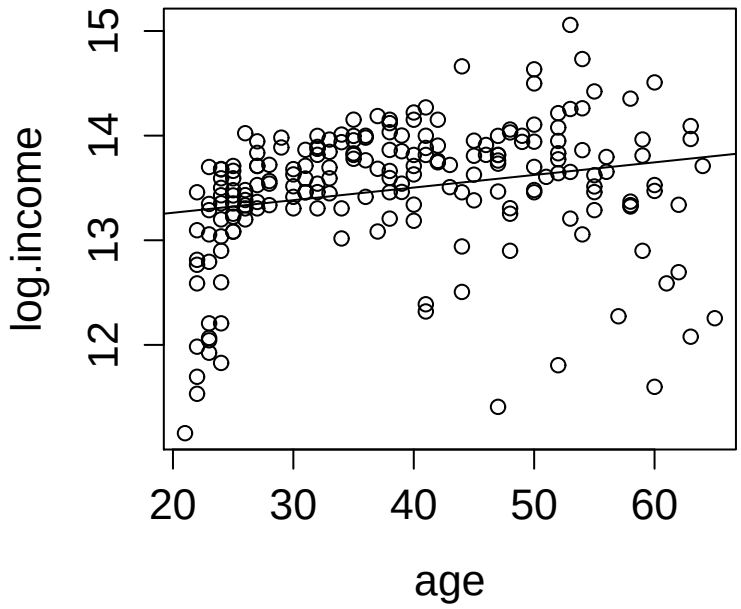
(ب) توزیع X مجهول است اما چون حجم نمونه زیاد است می‌توان با استفاده از قضیه حد مرکزی توزیع تقریبی $T(X)$ را تعیین کرد (این در حالتی است که $T(X)$ تابع خطی از $\sum_{i=۱}^{n} X_i$ باشد) در این حالت در انجام استنباط مشکلی وجود ندارد.

(ج) توزیع X معلوم است اما دست‌یابی به توزیع آماره $T(X)$ به دلیل غیرخطی بودن $T(X)$ برحسب $\sum_{i=۱}^{n} X_i$ و پیچیده بودن شکل تابعی آن امکان‌پذیر نیست. در این حالت به دلیل مجهول بودن توزیع $T(X)$ انجام استنباط با مشکل روبرو است که برای حل مشکل، الگوریتم عددی مونت‌کارلو[۱] غالبا مفید خواهد بود به این ترتیب که از توزیع معلومی که $X_۱$، $X_۲$، ...، X_n از آن نمونه‌گیری شده است به‌طور مکرر نمونه‌هایی شبیه‌سازی شده و $T(X_i)$، $i = ۱, ..., B$ که در آن X_i نمونه‌های شبیه‌سازی شده از توزیع معلوم و B تعداد تکرارهای نمونه‌گیری از توزیع معلوم می‌باشد، محاسبه می‌شود. سپس با استفاده از روش‌های برآورد چگالی، به چگالی برآورد شده‌ی $T(X)$ دست پیدا می‌کنیم.

(د) توزیع X مجهول است و امکان دستیابی به توزیع $T(X)$ وجود ندارد. در این حالت از روش‌های برآورد تابع چگالی همراه با باز نمونه‌گیری (مثلا خودگردان[۲]) استفاده می‌شود.

با توجه به حالت‌های ذکر شده فوق، دو نوع برآورد تابع چگالی داریم: (۱) پارامتری، (۲) ناپارامتری.

[۱] Monte Carlo
[۲] Bootstrap



شکل ۱.۱: نمودار پراکنش داده‌های درآمد و سن (درآمد/سن)

در حالتی‌که توزیع X معلوم است (حالت‌های الف و ج) معمولاً یک صورت پارامتری برای تابع چگالی احتمال (با فرض پیوسته بودن متغیر تصادفی) در نظر می‌گیرند و برآورد (تعیین) پارامترهای مجهول بر اساس نمونه تصادفی X صورت می‌گیرد. به عنوان مثال از میانگین نمونه $(\bar{X})$ و واریانس نمونه $(S^۲)$ برای برآورد پارامترهای مکان و مقیاس توزیع استفاده می‌شود که در آن

$$\bar{X} = \frac{۱}{n}\sum X_i, \quad S^۲ = \frac{۱}{n-۱}\sum (X_i - \bar{X})^۲$$

اما در حالتی‌که توزیع X مجهول است و هیچ‌گونه اطلاعاتی از ساختار تابعی آن نیز نداریم (حالت‌های ب و د)، از روش برآورد ناپارامتری تابع چگالی استفاده می‌شود. برخی از این روش‌ها به‌طور مختصر در فصل ۲ ذکر می‌شوند که یکی از آنها موضوع بحث این پایان نامه و موسوم به برآورد تابع چگالی به روش کرنل (هسته[۳]) بوده که یکی از روش‌های هموارسازی[۴] است.

روش هموارسازی کرنل به این دلیل انتخاب شده است که، روشی ساده را برای یافتن ساختار مجموعه داده‌ها بدون تحمیل یک مدل پارامتری است. یکی از بنیادی‌ترین روش‌های هموارسازی کرنل می‌تواند کاربرد آن در یافتن معادله‌ی رگرسیون ساده باشد که ارتباط بین دو متغیر را به‌صورت یک رابطه تابعی بیان می‌کند. معمولا یکی از متغیرها با X مشخص می‌شود که برای پیش‌بینی متغیر دیگر، Y که متغیر پاسخ است، استفاده می‌شود. در شکل ۱.۱ نمودار پراکنش داده‌های درآمد و سن نشان داده شده است که X سن و Y لگاریتم درآمد افراد را نشان

[۳] Kernel
[۴] Smoothing

می‌دهد که این داده‌ها نتیجه تحقیقی بر روی ۲۰۵ کارگر کانادایی می‌باشد. (اولاه[۵]، ۱۹۸۵). می‌خواهیم درآمد را به‌صورت تابعی از سن مدل‌بندی کنیم. ایده‌ی اولیه‌ای که به ذهن می‌رسد برازش یک خط راست به داده‌ها (برآورد رگرسیونی کمترین توان‌های دوم خطا) است که در این شکل نشان داده شده است. سوالی که مطرح می‌شود این است که چه زمانی مدل سازی Y به‌صورت یک تابع خطی از X راضی‌کننده است؟ با در اختیار داشتن نمونه‌ی تصادفی $(X_۱, Y_۱)$، $(X_۲, Y_۲)$، ... و (X_n, Y_n) یک رابطه قابل قبول بین دو متغیر درآمد (Y) و سن (X) می‌تواند با معادله خطی زیر بیان شود

$$Y_i = \beta_۰ + \beta_۱ X_i + \epsilon_i \qquad i = ۱, ..., n \tag{۱.۱}$$

که ϵ_i، متغیر تصادفی خطا با میانگین صفر است و مستلزم آن است که مشاهدات به‌طور تصادفی در اطراف خط راست پراکنده باشند. مدل خطی (۱.۱) یک نمونه از مدل رگرسیون پارامتری است. اما برای روشن کردن این نظریه، بیان مطالب بعدی لازم است. اکنون حالت کلی مسئله را در نظر بگیرید آن‌چه مسلم است باید تابع mای برای کمینه کردن خطای $E[Y - m(X)]^۲$ به‌دست آوریم که $m(X) = E(Y|X)$. این تابع بهترین انتخاب در رگرسیون خطی Y بر روی X است. در نتیجه

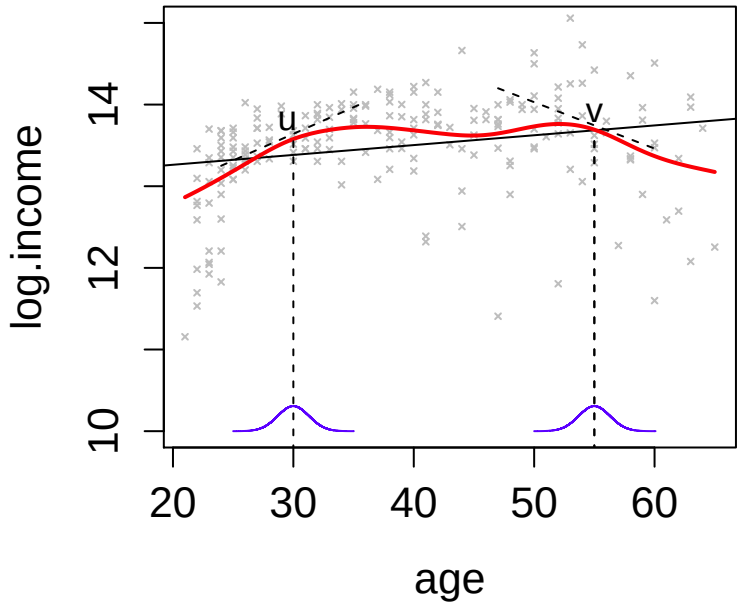
$$Y_i = m(X_i) + \epsilon_i \qquad i = ۱, ۲, ..., n$$

که در آن برای هر i، $E(\epsilon_i) = ۰$. با توجه به فرضیات در نظر گرفته شده، فقط دو پارامتر $\beta_۰$ و $\beta_۱$ مجهول می‌باشند که با یافتن برآورد مقادیر این دو پارامتر می‌توان تابع m را مشخص کرد. علاوه بر مدل رگرسیون خطی (۱.۱)، مدل‌های رگرسیونی پارامتری متفاوتی وجود دارند که از آن جمله می‌توان به موارد زیر اشاره کرد:

$$\begin{aligned} Y_i &= \beta_۱ \sin(\beta_۲ X_i) + \epsilon_i \\ Y_i &= \frac{\beta_۱}{\beta_۲ + X_i} + \epsilon_i \\ Y_i &= \beta_۰ + \beta_۱ X_i + \beta_۲ X_i^۲ + \epsilon_i \end{aligned}$$

با توجه به شکل پراکنش داده‌ها، مدل پارامتری انتخاب می‌شود. گاهی اوقات برای مدل کردن Y به‌صورت تابعی از X دلائل علمی وجود دارد در حالی‌که در بعضی موارد می‌توان از اطلاعات مطالعات پیشین استفاده کرد. انتخاب درست مدل پارامتری از اهمیت زیادی برخوردار است. تابع m می‌تواند سهمی‌وار، متناوب، یا یکنواخت باشد. اگر انتخاب یکی از پارامترها مناسب نباشد استنتاج درستی از رگرسیون نمی‌توان داشت. لذا دلیل استفاده از رگرسیون ناپارامتری را می‌توان این طور بیان نمود: در حالتی‌که از روی نمودار پراکندگی نتوان تابع ساده‌ای برای ارتباط بین متغیرها تشخیص داد، باید از خود داده‌ها برای مشخص کردن تابع استفاده کرد. رگرسیون پارامتری و ناپارامتری نباید به‌صورت دو رقیب باشند چه بسا در خیلی از حالت‌ها، برآورد یک رگرسیون ناپارامتری، یک مدل ساده پارامتری را پیشنهاد می‌کند یا در سایر حالت‌ها مشخص می‌کند که تابع رگرسیون بسیار پیچیده بوده و استفاده از یک مدل پارامتری معقول نیست. روش‌های زیادی برای به‌دست آوردن برآورد ناپارامتری m وجود دارد، روش‌هایی که از لحاظ ریاضی بسیار پیچیده هستند. روشی که بر مبنای یک ایده‌ی ساده شکل گرفته است، روش کرنل می‌باشد. شکل ۲.۱ برآورد تابع m از روی داده‌ها را نشان می‌دهد که یک برآورد کرنل خطی نامیده

[۵] Ullah



شکل ۲.۱: نمودار برآورد کرنل(درآمد/سن)

می‌شود و نشان می‌دهد نمودار تابع برآورد کرنل که معمولاً متناسب با تابع احتمال است نزدیک به تابع چگالی نرمال است

برآورد متغیر در اولین نقطه u با کمترین توان دوم وزنی به‌دست می‌آید، که در آن وزن‌ها مطابق ارتفاع تابع وزن انتخاب می‌شوند. این بدان معنی است که نقاط نزدیکتر به u نسبت به نقاطی که دورتر از u هستند، تاثیر بیشتری در خط برازش شده دارند. در شکل ۲.۱ تابع برآورد با خط شکسته نشان داده شده است. برآورد رگرسیون در نقطه u، ارتفاع خط در نقطه u است. برآورد در نقطه دیگری مانند v به همان روش به‌دست می‌آید با این تفاوت که وزن‌ها بر اساس کرنل پیرامون نقطه‌ی v تعیین می‌شود.

معمولا از برآوردگر استاندارد کرنل پارزن–روزن‌بلات[۶] استفاده می‌کنیم که در نقطه x به‌صورت زیر تعریف می‌شود:

$$\hat{f}_h(x) = \frac{۱}{nh} \sum_{i=۱}^{n} K\left(\frac{x - X_i}{h}\right)$$

که در آن K تابع کرنل و h پهنای باند[۷] است. (در فصل دوم چگونگی دستیابی به این برآوردگر شرح داده می‌شود).

[۶] Standard Parzen–Rosenblatt kernel estimator
[۷] Bandwidth

موضوع اصلی در یافتن برآوردگر مناسب بر اساس داده‌ها، تعیین h بهینه است. به منظور دست‌یابی به کارایی $\hat{f}_h$، معمولا از دو معیار خطای توان دوم تجمعی ISE[۸] و میانگین توان دوم خطای تجمعی MISE[۹] استفاده می‌کنیم.

در ادامه این فصل مروری بر تاریخچه برآورد کرنل داشته و در انتها تعاریف ریاضی که از آنها در فصول بعدی استفاده می‌شود را می‌آوریم.

۱.۱ مروری بر تاریخچه

ایده استفاده از برآورد به روش کرنل، اولین بار توسط روزن‌بلات (۱۹۵۶) و پارزن[۱۰] (۱۹۶۲) مطرح شد. البته بعدها واتسون[۱۱] (۱۹۶۹)، اپاچنیکوف[۱۲](۱۹۶۹) و نادارایا[۱۳] (۱۹۷۴) مقاله‌هایی در این زمینه ارائه کردند و تا به امروز مطالعه برای بسط و گسترش برآوردگرهای کرنلی همچنان ادامه دارد که می‌توان به واند[۱۴] و همکاران (۱۹۹۱)، یانگ و مارون[۱۵] (۱۹۹۹) در حالت پارامتری و روپرت و کلاین[۱۶] (۱۹۹۴) در حالت ناپارامتری اشاره کرد.

در فصل ۲ نشان می‌دهیم که تابع کرنل نقش چندانی در برآورد تابع چگالی نداشته و اهمیت انتخاب پهنای باند مناسب بیشتر از آن است. در راستای انتخاب پهنای باند مطالعات زیادی انجام گرفته است.

ایده روش اعتبارسنجی متقابل[۱۷] (CV) به رادمو[۱۸] (۱۹۸۲) و بومن[۱۹] (۱۹۸۴) برمی‌گردد. روش شبه‌درستنمایی اعتبارسنجی متقابل[۲۰] توسط هاباما[۲۱] و همکاران (۱۹۷۴) و داین[۲۲] (۱۹۷۶) ارائه گردید. روش‌های مختلفی مانند انتخاب‌گر پهنای باند پایا[۲۳] توسط چیو[۲۴] (۱۹۹۱a,b) و (۱۹۹۲)، اعتبارسنجی متقابل هموار توسط هال[۲۵] و همکاران (۱۹۹۲)، اعتبارسنجی متقابل اصلاح شده[۲۶] (MCV) توسط استوت[۲۷] (۱۹۹۲) و یا فلاچ و کروناکی[۲۸]

[۸] Integrated squared error
[۹] Mean integrated squared error
[۱۰] Parzen
[۱۱] Watson
[۱۲] Epanechnikov
[۱۳] Nadaraya
[۱۴] Wand
[۱۵] Yang and Marron
[۱۶] Ruppert and Cline
[۱۷] Cross-validation
[۱۸] Rudemo
[۱۹] Bowman
[۲۰] Pseudo-Likelihood CV
[۲۱] Habbeama
[۲۲] Duin
[۲۳] Stabilized Bandwidth Selector
[۲۴] Chiu
[۲۵] Hall
[۲۶] Modofied cross validation
[۲۷] Stute
[۲۸] Feluch and Koronacki

(۱۹۹۲)، اعتبارسنجی متقابل یک‌طرفه[۲۹] توسط مارتینز میراندا[۳۰] و همکاران (۲۰۰۹) هم‌چنین اعتبارسنجی غیرمستقیم[۳۱] توسط ساچاک[۳۲] و همکاران(۲۰۱۰) پیشنهاد شده است. روش اریبی اعتبارسنجی متقابل[۳۳] (BCV) توسط اسکات و ترل[۳۴] (۱۹۸۶) که کمینه کردن MISE مجانبی است شبیه روش جای‌گذاری[۳۵] عمل می‌کند با این تفاوت که از روش جک‌نایف[۳۶] بهره می‌برد. روش‌های ترکیب کردن انتخاب‌گرهای مختلف یا برآوردگرهای چگالی توسط احمد و ران[۳۷] (۲۰۰۴) پیشنهاد گردید و روش اثر کرنل[۳۸] نامیده شد و توسط مامن[۳۹] و همکاران (۲۰۱۱) به‌کار گرفته شد.

در مقایسه با CV، روش‌هایی که توابع مختلف را کمینه می‌کنند، روش‌های جای‌گذاری گفته می‌شوند یعنی از MISE به‌جای معیار ISE استفاده می‌شود. CV اجازه انتخاب پهنای باند بدون ایجاد مفروضات در مورد کلاس همواری را می‌دهد. از طرفی روش‌های جای‌گذاری سرعت همگرایی بیشتری در مقایسه با CV دارند. اما متاسفانه به‌شدت به انتخاب h وابسته هستند؛ چنانچه برآوردهای اولیه‌ای از داده‌های گذشته به‌دست آیند، انجام روش‌های جای‌گذاری خوب است. از جمله این انتخاب‌گرها روش قانون سرانگشتی[۴۰] سیلورمن[۴۱] (۱۹۸۶) است که احتمالاً محبوبترین روش است. ایرادهای این روش به مرور برطرف شده است برای مثال، شیتر و جونز[۴۲] (۱۹۹۱) یا هال و همکاران (۱۹۹۱) را ببینید. روش‌های بوت‌استرپ[۴۳] (تیلور[۴۴]، ۱۹۸۹) و همه‌ی نسخه‌های اصلاح‌شده‌ی آن (کائو[۴۵]، ۱۹۹۳ و چائوکن و همکاران[۴۶]، ۲۰۰۸) به منظور کمینه کردن MISE در روش‌های جای‌گذاری به‌کار می‌روند.

مقاله‌هایی در مورد مقایسه روش‌های مختلف خودکار انتخاب پهنای باند متغیرهای تصادفی وجود دارند. از سال ۱۹۷۰ تا سال ۱۹۸۰، مقالات متعددی در مورد برآورد چگالی توسط وگمن[۴۷] (۱۹۷۲)، تارتل و کرونمال[۴۸] (۱۹۷۶)، فریر[۴۹] (۱۹۷۷)، ورتز و اشنایدر[۵۰] (۱۹۷۹)، بین و سوکوس[۵۱] (۱۹۸۰) منتشر شده است. در ابتدا

[۲۹] One-sided cross validation
[۳۰] Martinez-Miranda
[۳۱] Indirect cross validation
[۳۲] Savchuk
[۳۳] Biased cross validation
[۳۴] Scott and Terrell
[۳۵] Plug-in
[۳۶] Jack-knife
[۳۷] Ahmad and Ran
[۳۸] Kernel contrast
[۳۹] Mammen
[۴۰] rule-of-thumb
[۴۱] Silverman
[۴۲] Sheather and Jones
[۴۳] Bootstrap
[۴۴] Taylor
[۴۵] Cao
[۴۶] Chacon et al.
[۴۷] Wegman
[۴۸] Tartel and Kronmal
[۴۹] Fryer
[۵۰] Wertz and Schneider
[۵۱] Bean and Tsokos

روش‌های انتخاب پارامتر هموار سازی توسط مارون[۵۲] (۱۹۸۸ و پارک و مارون[۵۳] (۱۹۹۰) منتشر شد. سپس مطالعات شبیه‌سازی گسترده‌ای توسط پارک و تورلاش[۵۴] (۱۹۹۲) و کائو و همکاران (۱۹۹۴) و چیو (۱۹۹۶) انتشار یافت. بررسی کوتاهی نیز توسط جونز[۵۵] و همکاران در سال ۱۹۹۶ انجام گرفت (جونز و همکاران، ۱۹۹۶a, b). پس از مدتی، لودر[۵۶] (۱۹۹۹) مقایسه‌ای از مقالات را منتشر کرد که پاسخی به مقاله‌ی جونز و همکاران بود. اخیرا حیدرنیک[۵۷] و همکاران (۲۰۱۳) برخی از روش‌های جدید انتخاب پهنای باند و روش‌های ترکیبی آن را به‌طور دقیق‌تری بررسی کرده‌اند.

۲.۱ تعاریف و قضایا

در این قسمت یک سری تعاریف مورد نیاز از آنالیز ریاضی را که عمدتاً برگرفته از کتاب آپاستل[۵۸] (۱۹۷۴) و رودین[۵۹] (۱۹۷۶) می‌باشد، می‌آوریم.

تعریف ۱.۲.۱ (گوی باز). اگر $X \in \mathbb{R}^k$ و $r > ۰$، گوی باز (یا بسته) B به مرکز X و شعاع r عبارت است از مجموعه‌ی تمام $Y \in \mathbb{R}^k$هایی که $|Y - X| < r$ (یا $|Y - X| \leq r$). در این صورت از نماد $B_{X,r}$ استفاده می‌کنیم.

تعریف ۲.۲.۱ (نقطه درونی مجموعه). نقطه $\theta \in X \subseteq \mathbb{R}^p$ را یک نقطه درونی مجموعه X می‌نامیم در صورتی‌که گوی بازی مانند $B_{\theta,\rho}$ وجود داشته باشد به طوری که زیر مجموعه X باشد. به عبارت دیگر

$$\forall\, \theta \in X \quad \exists\, B_{\theta,\rho} \subseteq X$$

تعریف ۳.۲.۱ (نقطه انباشتگی مجموعه). نقطه $\theta \in X \subseteq \mathbb{R}^p$ را یک نقطه انباشتگی X می نامیم اگر هر گوی باز به مرکز θ شامل نقطه ای از X غیر از θ باشد. به عبارت دیگر

$$\forall\, B_{\theta,\rho} \in \mathbb{R}^p, \quad (B_{\theta,\rho} - \theta) \cap X \neq \varnothing$$

تعریف ۴.۲.۱ (مجموعه باز). مجموعه $X \subseteq \mathbb{R}^p$ را باز می نامیم در صورتی که هر نقطه آن درونی باشد و Xبسته می نامیم در صورتیکه $\mathbb{R}^p - X$باز باشد.

تعریف ۵.۲.۱ (پوشش). منظور از یک پوشش باز مجموعه‌ی E در فضای متری X یعنی گردایه‌ای (مجموعه‌ای از مجموعه‌ها) از مجموعه‌های باز X مانند G_a که $E \subset \cup_a G_a$

[۵۲] Marron
[۵۳] Park and Marron
[۵۴] Park and Turlach
[۵۵] Jones
[۵۶] Loader
[۵۷] Heidenrich
[۵۸] Apostel
[۵۹] Rudin

تعریف ۶.۲.۱ (مجموعه فشرده). مجموعه‌ای که هر پوشش آن یک زیر پوشش متناهی داشته باشد مجموعه‌ای فشرده خوانده می‌شود.

در این صورت زیر مجموعه‌ای ازفضای اقلیدسی $\mathbb{R}^۲$ که بسته و کراندار باشد فشرده است مثلا در $\mathbb{R}$ فاصله بسته $[۰, ۱]$ فشرده است اما مجموعه‌ی اعداد صحیح $\mathbb{Z}$ این طور نیست زیرا کراندار نیست.

تعریف ۷.۲.۱ (نرم). تابع $f : \mathbb{R}^p \to \mathbb{R}$ را نرم گوئیم، اگر دارای شرایط زیر باشد

- $f(.)$تابعی نا منفی باشد،یعنی به ازای هر $X \in \mathbb{R}^p$ داشته باشیم، $f(X) \geq ۰$.
- $f(.)$ معین باشد،یعنی $f(X) = ۰$ اگر و تنها اگر $X = ۰$.
- $f(.)$ همگن باشد، یعنی به ازای هر $X \subseteq \mathbb{R}^p$ و $t \in \mathbb{R}$ داشته باشیم، $f(tX) = |t|f(X)$.
- $f(.)$ در نامساوی مثلثی صدق کند، یعنی به ازای هر $X, Y \in \mathbb{R}^p$ داشته باشیم

$$f(X + Y) \leq f(X) + f(Y).$$

در این صورت از نماد $f(.) = \|.\|$ برای نمایش نرم استفاده می‌کنیم.

تعریف ۸.۲.۱ (ماتریس ترانهاده[۶۰]). اگر جای سطرها و ستون های ماتریس Σ را تعویض نماییم، ماتریس حاصل ترانهاده ماتریس Σ نامیده می‌شود و با نماد Σ^T نمایش داده می‌شود.

$$\Sigma^T = (\sigma_{ij})^T = (\sigma_{ji}),$$

که در آن σ_{ij} عنصر موجود در سطر iام و ستون jام است.

تعریف ۹.۲.۱ (نرم اقلیدسی). روی بردار $X = (X_۱, \ldots, X_p)$، $\|X\|$ را نرم اقلیدسی گوئیم هر گاه

$$\|X\| = (X^T X)^{\frac{۱}{۲}} = \left(\sum_{i=۱}^{p} x_i^۲ \right)^{\frac{۱}{۲}}$$

تعریف ۱۰.۲.۱ (limsup). دنباله‌ی $\{X_n\}$ را در نظر بگیرید. در این صورت

$$\begin{aligned} \limsup_{n\to\infty} X_n &= \lim_{n\to\infty} \left(\sup_{m\geq n} X_m \right) \\ &= \lim_{n\to\infty} (\sup X_m) \\ &= \inf_{n\geq ۰} \sup_{m\geq n} X_m \\ &= \inf \{\sup \{X_m : m \geq n\} : n \geq ۰\} \end{aligned}$$

[۶۰] Transpose matrix

یا به‌طور معادل، می‌توان فرض کرد دنباله‌ی J_n به‌صورت زیر تعریف شده باشد.

$$\begin{aligned} J_n &= \sup\{X_m : m \in \{n, n+۱, n+۲, \ldots\}\} \\ &= \cup_{m=n}^{\infty} X_m = X_n \cup X_{n+۱} \cup X_{n+۲} \cup \cdots \end{aligned}$$

دنباله‌ی J_n یک دنباله‌ی ناصعودی ($J_n \supseteq J_{n+۱}$) است، زیرا $J_{n+۱}$ اجتماع مجموعه‌های کوچکتر از J_n است. بزرگترین کران پایین روی این دنباله از اجتماع X_nها برابر است با

$$\begin{aligned} \text{limsup}_{n\to\infty} X_n &= \inf\{\sup\{X_m : m \in \{n, n+۱, n+۲, \ldots\}\} : n \in \{۱, ۲, \ldots\}\} \\ &= \cap_{n=۱}^{\infty} (\cup_{m=n}^{\infty} X_m) \end{aligned}$$

تعریف ۱۱.۲.۱ (نماد Oبزرگ[۶۱]). فرض کنید X_n و Y_n دو دنباله از متغیرهای تصادفی باشند. اگر

$$\text{limsup}_{n\longrightarrow\infty} |\frac{X_n}{Y_n}| < \infty$$

به عبارت دیگر اگر $n \longrightarrow \infty$، عبارت $|\frac{X_n}{Y_n}|$ کران دار باشد می‌نویسیم $X_n = O(Y_n)$

تعریف ۱۲.۲.۱ (نماد o کوچک[۶۲]). فرض کنید X_n و Y_n دو دنباله از متغیرهای تصادفی باشند. اگر

$$\lim_{n\longrightarrow\infty} \left|\frac{X_n}{Y_n}\right| = ۰$$

می‌نویسیم $X_n = o(Y_n)$

در ریاضیات علامت O بزرگ رفتار حدی یک تابع، وقتی آرگومان های آن به یک عدد خاص میل می‌کند را بیان می‌کند علامت Oبزرگ به کاربراجازه می‌دهد که تابع را ساده کند تا بر روی نرخ رشد متمرکز شود؛ بنابراین توابع مختلف با نرخ رشد یکسان می‌توانند دارای یک علامت O مشابه باشند. تابع زیر بر حسب n را در نظر بگیرید

$$T(n) = ۴n^۲ - ۵n + ۷$$

اگر تعداد ورودی این مساله به بی‌نهایت میل کند اندازه جمله $n^۲$ بسیار نزدیک‌تر از دیگر جمله‌ها خواهد بود. در این صورت گفته می‌شود $T(n) \in O(n^۲)$ یا $T(n) = O(n^۲)$.

در ریاضیات معمولا برای نشان دادن اینکه یک سری هندسی متناهی تا چه اندازه به تابع مورد نظر نزدیک است خصوصاًدر سری تیلور ناقص از این علامت استفاده می‌شود.

تعریف ۱۳.۲.۱ (بسط تیلور). فرض کنید f یک تابع حقیقی-مقدار روی $\mathbb{R}$ باشد و همچنین $x \in \mathbb{R}$. اگر برای بعضی δها تابع f دارای مشتق‌های پیوسته تا مرتبه‌ی p در بازه‌ی $(x-\delta, x+\delta)$ باشد، آن‌گاه برای هر دنباله‌ی α_n همگرا به سمت صفر، داریم

$$\begin{aligned} f(x+\alpha_n) &= \sum_{j=۱}^{p} \left(\frac{\alpha_n^j}{j!}\right) f^{(j)}(x) + o(\alpha_n^p) \\ &= \sum_{j=۱}^{p} \left(\frac{\alpha_n^j}{j!}\right) f^{(j)}(x) + O(\alpha_n^{p+۱}) \end{aligned}$$

[۶۱] Big o notation
[۶۲] Small o notation

تعریف ۱۴.۲.۱ (تابع نشان‌گر) **.** تابع نشان‌گر مجموعه‌ی A به صورت زیر تعریف می‌شود

$$I(A) = \begin{cases} ۱ & x \in A \\ ۰ & x \notin A \end{cases}$$

تعریف ۱۵.۲.۱ (پیچش[۶۳]) **.** در ریاضیات یا به‌طور دقیق‌تر آنالیز تابعی، پیچش یک عملگر ریاضی است که بر روی دو تابع f و g عمل کرده و تابع سومی را تولید می‌کند که می‌توان به‌عنوان نسخه‌ی تصحیح‌شده‌ی یکی از دو تابع اصلی نگریسته شود.

پیچش دو تابع f و g به‌صورت $f * g$ نوشته می‌شود و به‌شکل انتگرال دو تابع که یکی از آنها برعکس شده و روی یکدیگر می‌لغزند تعریف می‌شود. با این تعریف، پیچش، یک نوع خاص از تبدیل انتگرالی است.

$$\begin{aligned}(f * g)(x) &= \int_{-\infty}^{+\infty} f(y)g(x-y)dy \\ &= \int_{-\infty}^{+\infty} f(x-y)g(y)dy\end{aligned}$$

تعریف ۱۶.۲.۱ (همگرایی تقریبا قطعی) **.** فرض کنید X و دنباله‌ی $X_۱$، $X_۲$، ... متغیرهای تصادفی باشند که بر روی یک فضای نمونه‌ی Ω تعریف شده‌اند. مجموعه‌ی C را مجموعه‌ی همه‌ی نقاط ω در Ω در نظر بگیرید به‌طوری که $X_n(\omega)$ به $X(\omega)$ همگرا باشد. یعنی

$$C = \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}$$

که آن را پیشامد همگرایی دنباله‌ی متغیرهای تصادفی $\{X_n\}$ گویند. هرگاه $C \neq \Omega$ و $P(C) = ۱$، دنباله‌ی $\{X_n\}$ دارای خاصیت همگرایی تقریبا حتمی می‌باشد به عبارت دیگر، احتمال مجموعه‌ی نقاط واگرایی برابر صفر است. این نوع همگرایی با نماد $X_n \xrightarrow{a.s.}$ نشان داده و به‌صورت زیر نوشته می‌شود.

$$P\left(\lim_{n\to\infty} X_n = X\right) = ۱ \qquad (۲.۱)$$

تعریف ۱۷.۲.۱ (همگرایی در احتمال) **.** فرض کنید $\{X_n\}$ دنباله‌ای از متغیرهای تصادفی باشد. دنباله‌ی $\{X_n\}$ در احتمال به متغیر تصادفی X همگرا است اگر برای هر $\varepsilon > ۰$،

$$\lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = ۰$$

این نوع همگرایی را با نماد $X_n \xrightarrow{P}$ نشان می‌دهند.

قضیه ۱۸.۲.۱ (قضیه فوبینی) **.** *فرض کنید $f(x.y)$ روی ناحیه‌ای مانند R پیوسته باشد. در این صورت*

(الف) *اگر تعریف R به‌صورت $a \leq x \leq b$، $f_۱(x) \leq y \leq f_۲(x)$ باشد با این شرط که $f_۱$ و $f_۲$ بر بازه‌ی $[a, b]$ پیوسته باشد، آن‌گاه*

$$\iint_R f(x,y)\mathrm{d}x\mathrm{d}y = \int_a^b \int_{f_۱(x)}^{f_۲(x)} f(x,y)\mathrm{d}y\mathrm{d}x$$

[۶۳] Convolution

(ب) *اگر تعریف* R *به‌صورت* $c \leq y \leq d$، $g_1(y) \leq x \leq g_2(y)$ *باشد با این شرط که* g_1 *و* g_2 *بر بازه‌ی* $[a,b]$ *پیوسته باشد، آن‌گاه*

$$\iint_R f(x,y)\mathrm{d}x\mathrm{d}y = \int_c^d \int_{g_1(x)}^{g_2(x)} f(x,y)\mathrm{d}x\mathrm{d}y$$

فصل ۲

مروری بر معمول‌ترین روش‌های ناپارامتری برآورد چگالی

۱.۲ مقدمه

یکی از محدودیت‌های روش پارامتری برآورد تابع چگالی، دانستن ساختار تابعی آن است. در روش ناپارامتری، معمولاً با درنظر گرفتن برخی شرایط لازم بر روی تابع چگالی، برآورد مستقیما انجام می‌شود. از آن جایی که هدف اصلی این پایان‌نامه، مطالعه روش‌های مختلف تعیین پهنای باند در روش برآورد کرنل است، لازم است که معمول‌ترین روش‌های پارامتری برآورد تابع چگالی مانند کرنل شرح داده شود. از آن‌جایی که روش برآورد کرنل، مکمل روش‌های ناپارامتری هیستوگرام و برآورد ساده است، در این فصل ابتدا مروری کوتاه بر این دو روش داشته و سپس به برآورد تابع چگالی و به‌طور خاص روش برآورد کرنل پرداخته شده است. مطالب این فصل عمدتا برگرفته از واند و جونز (۱۹۹۴) است.

۲.۲ روش‌های ناپارامتری برآورد تابع چگالی

۱.۲.۲ روش‌های برآورد چگالی

فرض کنید $X = (X_1, X_2, \ldots, X_n)$ یک نمونه تصادفی از جامعه‌ای با تابع توزیع تجمعی $F(.)$ و تابع چگالی احتمال $f(.)$، که هر دو مجهول هستند، باشد. همچنین فرض کنید توزیع جامعه پیوسته بوده به‌طوری که به ازای هر x عضو تکیه‌گاه X، $f(x) = \frac{dF(x)}{dx}$.

از روش‌های معمول برآورد $f(.)$ می‌توان به روش‌های ساده، کرنل، کرنل تطبیقی[۱]، نزدیکترین همسایگی، نزدیکترین همسایگی تعمیم یافته، سری های متعامد و ماکزیمم درستنمائی جریمه شده[۲] اشاره کرد. برای آگاهی

[۱] Adaptive kernel

[۲] Penalized likelihood maximum estimator

بیشتر درباره‌ی این روش‌ها هستی[۳] و همکاران (۲۰۰۹) را ببینید. در ادامه، تنها روش برآورد بافت نگار و برآورد ساده را برای ورود به روش کرنل توضیح می‌دهیم.

۲.۲.۲ روش بافت نگار

قدیمی‌ترین و پراستفاده‌ترین روش برآورد تابع چگالی، روش بافت نگار (هیستوگرام) است. فرض کنید X یک متغیر تصادفی با تابع چگالی $f(.)$ باشد به‌طور خلاصه، برآورد بافت نگار $f(.)$ در نقطه x با استفاده از یک نمونه‌ی nتایی به‌صورت زیر تعریف می‌شود:

$$\hat{f}(x) = \frac{۱}{nh} \times \Big(\text{تعداد } x_i\text{های موجود در باندی که } x \text{ در آن واقع است}\Big)$$

که در آن n حجم نمونه و h طول یا پهنای باند است. برای ساختن بافت نگار، نقطه شروع (موقعیت لبه‌های باند) و پهنای باند باید معلوم باشد که در آن مقدار پهنای باند، بیان‌گر میزان همواری در بافت نگاراست. انتخاب هر کدام از آن دو می‌تواند تاثیر قابل توجهی در نتایج بافت نگار داشته باشد.

۱.۲ بافت نگار وزن هنگام تولد ۵۰ کودکی که به بیماری تنفسی شدید مبتلا شدند را برای ۴ باند مختلف نشان می‌دهد (ون‌وایلت و گوپتا[۴]، ۱۹۷۳). در ردیف بالا، سمت چپ و راست به‌ترتیب مبنی بر پهنای باند ۰/۲ و ۰/۸ است و ردیف پایین شکل ۱.۲، مبنی بر پهنای باند۰/۴ است که نمودارهای سمت چپ و راست به ترتیب به‌ازای نقطه‌ی شروع ۰/۷ و ۰/۹ رسم شده‌اند. توجه کنید که هر بافت نگار، شکل مختلفی از چگالی داده‌ها را نشان می‌دهد. پهنای باند کوچک نسبتاً شکل ناهمواری را ایجاد می‌کند.

مشاهده می‌شود که پهنای باند بزرگتر، بافت نگار هموارتری را نشان می‌دهد. در شکل‌های سمت راست و چپ ردیف پایین با وجود یکسان بودن پهنای باند شکل‌های متفاوتی دیده می‌شود.

میزان پهنای باند که پارامتر هموارسازی نیز نامیده می‌شود، حساسیت بافت نگار به انتخاب مکان لبه‌ها مشکل‌ساز است؛ در صورتی‌که در سایر روش‌های برآورد مانند روش کرنل این مشکل وجود ندارد. این اشکال، یک ایراد اساسی بافت نگار است. اسکات[۵] این مشکل را با میانگین گرفتن از چند بافت نگار بیان کرده است اما این مشکل را می‌توان با استفاده از برآورد چگالی به روش کرنل نیز بر طرف کرد و بنابراین انگیزه خوبی برای مطالعه روش کرنل است. البته چندین مشکل دیگر برای برآورد بافت نگار وجود دارد که با برآورد کرنل رفع نمی‌شود از آن جمله می‌توان به پله‌ای بودن تابع و مشکلات بیشتری که در حالت چند متغیره وجود دارد به‌خصوص نمایش نموداری آن اشاره کرد.

بنابراین، دو ایراد اساسی بافت نگار، عبارتند از:

(۱) نادقیق بودن آن است خصوصاً وقتی اندازه نمونه کوچک باشد.

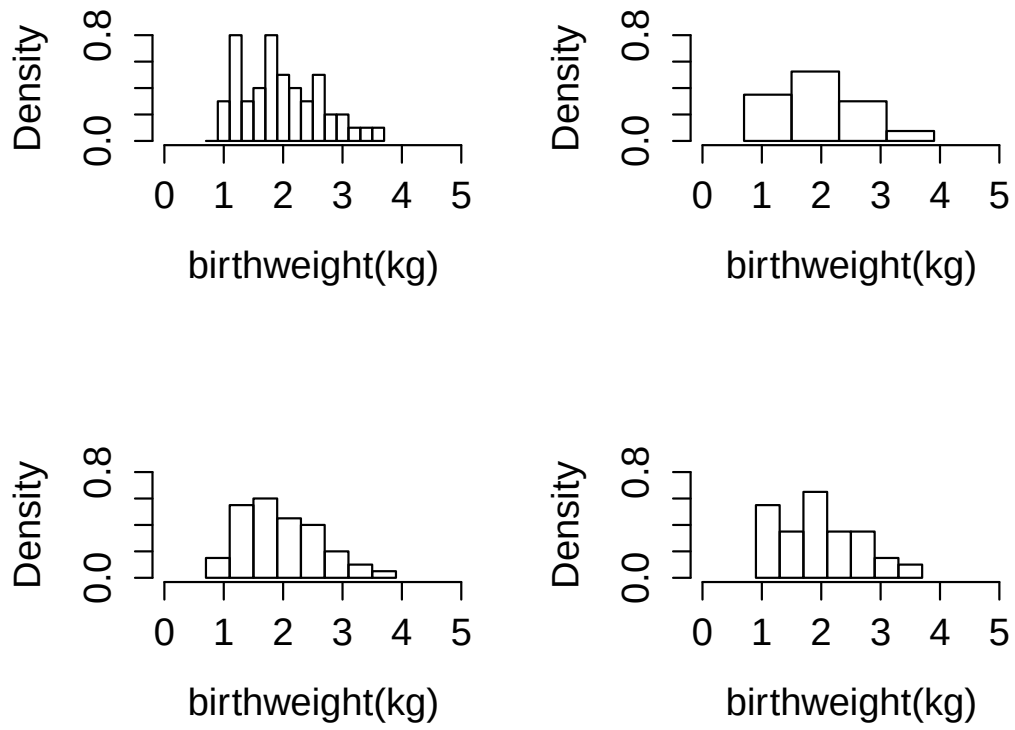
(۲) ناپیوستگی بافت نگار که در صورت نیاز به مشتق‌های آن در مراحل بعدی مطالعات (روش‌های تعیین باند و ...) باعث ایجاد مشکل خواهد شد.

---

[۳] Hastie

[۴] van Violet and Gupta

[۵] Scott 1985



شکل ۱.۲: بافت نگارهای داده‌های وزن زمان تولد. ردیف بالا–سمت چپ، پهنای باند ۰/۲ است، ردیف بالا، سمت راست، برای پهنای باند ۰/۸ رسم شده است. ردیف پایین– سمت چپ، در نقطه‌ی شروع ۰/۷، به ازای پهنای باند ۰/۴ و در ردیف پایین–سمت راست به ازای پهنای باند ۰/۴ در نقطه شروع ۰/۹ رسم شده است.

با این وجود، بافت نگار در عین سادگی روش خوبی برای نمایش داده‌ها است و به کمک آن می‌توان روش مناسب دیگری را برای برآورد چگالی انتخاب نمود.

۳.۲.۲ برآوردگر ساده

در ادامه جزئیات تکنیکی تخمین برآوردگر ساده را برای ورود به روش کرنل ذکر می‌کنیم. در این راستا، فرض کنید متغیر تصادفی و پیوسته X دارای تابع چگالی $f(.)$ باشد، در این صورت داریم

$$\begin{aligned} f(x) &= \frac{dF(x)}{dx} \\ &= \lim_{h\to ۰} \frac{F(x+h)-F(x)}{h} \\ &= \lim_{h\to ۰} \frac{۱}{h} P(x<X<x+h) \end{aligned}$$

بنابراین چنانچه بازه‌ی $(x-h, x+h)$ را برای مقادیر ممکن X در نظر بگیریم، تابع چگالی برابر است با

$$f(x) = \lim_{h \to ۰} \frac{۱}{۲h} P(x-h < X < x+h)$$

یک برآوردگر تجربی[۶] (برآوردگر طبیعی) $P(x-h < X < x+h)$ عبارت است از

$$\frac{\text{تعداد } X_۱، X_۲، \ldots، X_n \text{های واقع در بازه‌ی } (x-h, x+h)}{n}$$

بنابراین برآوردگر ساده f به صورت زیر تعریف می‌شود

$$\hat{f}(x) = \frac{۱}{۲nh} \Big(\text{تعداد } X_۱، X_۲، \ldots، X_n \text{های واقع در بازه‌ی } (x-h, x+h) \Big) \qquad (۱.۲)$$

به ازای $i = ۱, ۲, \ldots, n$، فرض کنید X_i نقطه‌ی مشاهده شده باشد. در این صورت رابطه‌ی (۱.۲) برابر خواهد بود با:

$$\hat{f}(x) = \frac{۱}{۲nh} \sum_{i=۱}^{n} I(X_i \in (x-h, x+h)) \qquad (۲.۲)$$

که در آن $I(\cdot)$ تابع نشان‌گر است. رابطه‌ی (۲.۲) را می‌توان به‌صورت زیر بازنویسی کرد

$$\begin{aligned} \hat{f}(x) &= \frac{۱}{n} \sum_{i=۱}^{n} \frac{۱}{۲h} I(X_i \in (x-h, x+h)) \\ &= \frac{۱}{n} \sum_{i=۱}^{n} \frac{۱}{h} \frac{۱}{۲} I(X_i \in (x-h, x+h)) \\ &= \frac{۱}{n} \sum_{i=۱}^{n} \frac{۱}{h} \frac{۱}{۲} I\left(\left| \frac{x - X_i}{h} \right| \leq ۱ \right) \end{aligned} \qquad (۳.۲)$$

اکنون، اگر تابع وزن $w(.)$ را به‌صورت زیر در نظر بگیریم

$$w(x) = \begin{cases} \frac{۱}{۲} & |x| < ۱ \\ ۰ & \text{سایر جاها} \end{cases}$$

آن‌گاه برآوردگر ساده در رابطه‌ی (۳.۲) را می‌توان به‌صورت تابعی زیر نمایش داد.

$$\hat{f}(x) = \frac{۱}{n} \sum_{i=۱}^{n} \frac{۱}{h} w\left(\frac{x - X_i}{h} \right) \qquad (۴.۲)$$

رابطه (۴.۲) به وسیله‌ی قراردادن یک پنجره با پهنای $۲h$ و ارتفاع $(۲nh)^{-۱}$ روی هر مشاهده و سپس جمع زدن روی کل مشاهدات به‌دست می‌آید.

برای بررسی ارتباط برآوردگر ساده با بافت نگار، فرض کنید بافت نگاری در اختیار داریم که هر باند آن پهنایی به اندازه $۲h$ دارد. همچنین هیچ مشاهده‌ای دقیقا روی لبه‌ی باندها قرار نگرفته باشد. اگر x در مرکز یکی

[۶] Emprical Estimator

از باندهای بافت نگار واقع شده باشد، نتیجه می‌شود که برآورد ساده‌ی f در x دقیقا همان ارتفاع بافت نگار در نقطه‌ی x خواهد بود.

از نظر ظاهری تشابه برآورد ساده و بافت نگار در این است که هر دو تابعی پله‌ای و در نتیجه ناپیوسته می‌باشند، یعنی در نقاط داخل یک بازه مقدار ثابت داشته و در نقاط انتهایی بازه‌ها دارای جهش و در نتیجه مشتق‌ناپذیر هستند. تفاوت این دو برآوردگر در این است که در روش بافت نگار پهنای بازه‌ها، ثابت و برابر h است اما در برآورد ساده، پهنای بازه‌ها متغیر و در هر نقطه از محور xها تابعی از تراکم مشاهدات در آن مکان و هم‌چنین تابعی از h است.

برای روشن شدن موضوع، داده‌های میزان فوران یک چشمه‌ی آب گرم قدیمی در پارک ملی یلوستون[۷] (ویسبرگ[۸] ۱۹۸۰) را در نظر بگیرید. قاب بالا شکل ۲.۲، برآورد بافت نگار به ازای پهنای باند ۰/۵ و قاب پایین برآورد ساده این داده‌ها را به ازای پهنای باند ۰/۲۵ نشان می‌دهد.

۴.۲.۲ برآوردگر کرنل

قبل از تعریف برآورد چگالی به روش کرنل، تابع کرنل را تعریف می‌کنیم. تابع کرنل، نوع خاصی از تابع چگالی احتمال است که خاصیت زوج بودن به آن اضافه شده است. بنابراین، تابع کرنل، تابعی با خصوصیات زیر است:

- نامنفی
- حقیقی-مقدار
- زوج
- انتگرال معین آن روی تکیه‌گاهش باید برابر یک باشد.

ایده اصلی برآوردگر کرنل $f(.)$ به‌صورت جایگزینی تابع وزن گسسته $w(.)$ در رابطه (۴.۲)، با یک تابع پیوسته که کرنل نامیده می‌شود، مطرح می‌گردد. همان‌طور که بیان شد در برآورد ساده تمامی نقاطی که در باند شامل x است تاثیر یکسانی در برآورد مقدار تابع احتمال در نقطه x بازی می‌کنند و به تمامی آن نقاط یک وزن تعلق می‌گیرد (به این علت یک تابع پله‌کانی است) و بنابراین می توان با تعریف تابع وزنی که مشاهدات بسته به فاصله‌ای که از متغیر X دارند تاثیر متفاوتی در تابع چگالی بازی می‌کند، برآورد مناسب‌تری به‌دست آورد به‌طوری که هر چه فاصله مشاهده از متغیر X کمتر باشد نقش بیشتری در تخمین تابع چگالی بازی می‌کند.

برآوردگر کرنل تابع چگالی $f(.)$ با تابع کرنل K به‌صورت زیر می‌باشد.

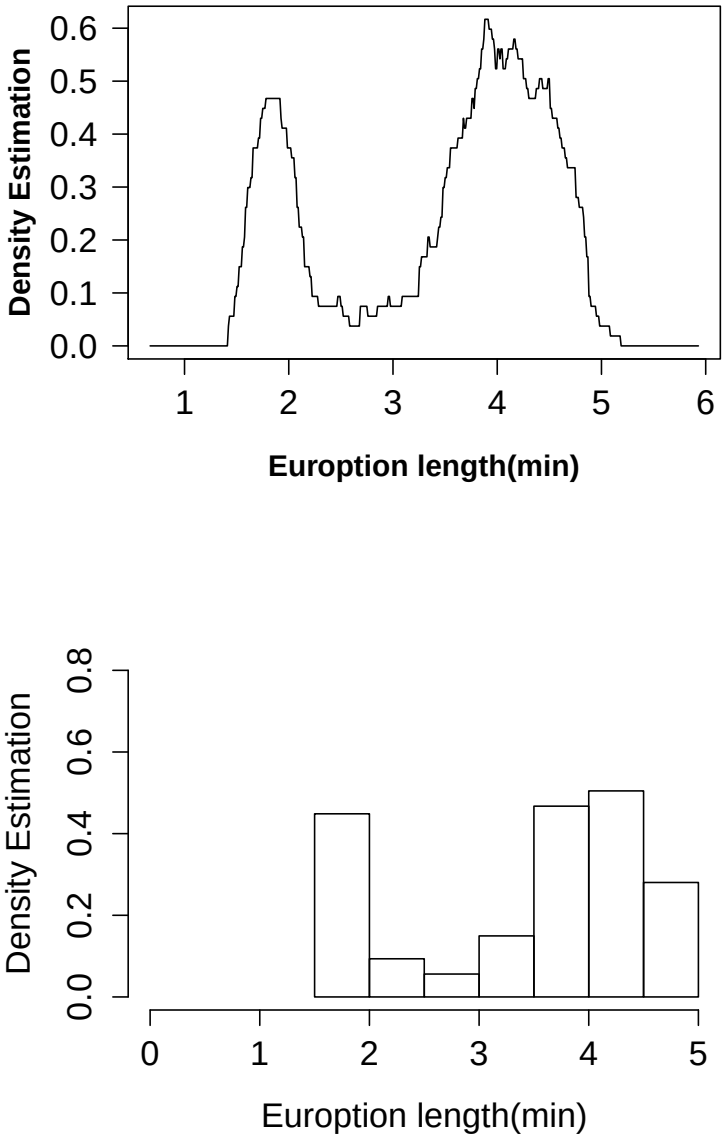
$$\hat{f}(x) = \frac{۱}{nh} \sum_{i=۱}^{n} K\left(\frac{x - X_i}{h}\right) \qquad (۵.۲)$$

[۷] Yellowstone National Park
[۸] Weisberg

که در آن h پهنای باند یا پارامتر هموارسازی نامیده می‌شود و مقدار آن میزان هموارى در برآورد کرنل را تعیین می‌کند. هم‌چنین $K(.)$ تابعی هموار و پیوسته است و اگر در شرط زیر صدق کند، آن را تابع کرنل می‌نامند

$$\int_{-\infty}^{+\infty} K(x)dx = ۱$$

معمولاً $K(.)$ یک تابع چگالی متقارن پیرامون صفر و یک مدی در نظر گرفته می‌شود که تضمین می‌کند برآوردگر $\hat{f}$ یک تابع چگالی است.



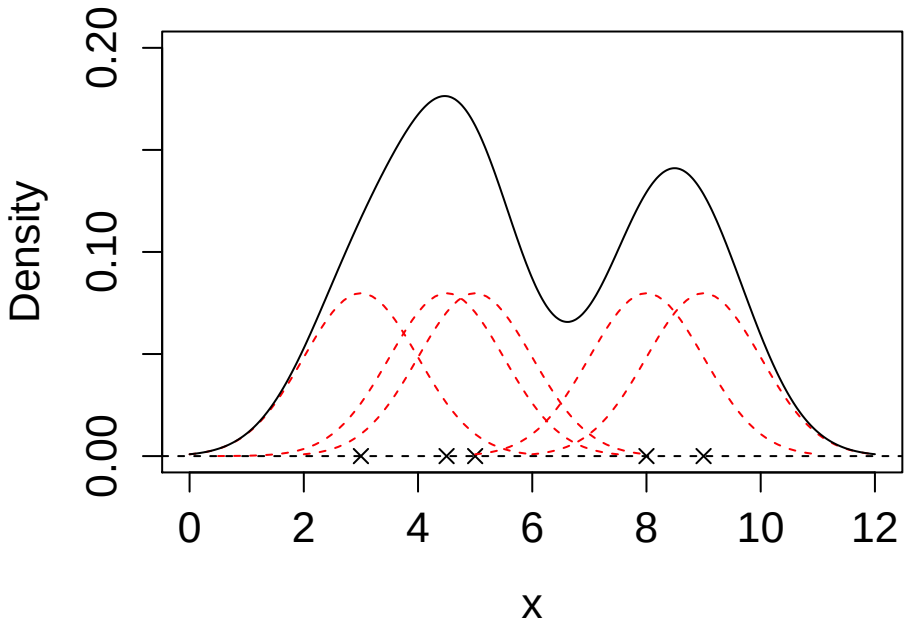
شکل ۲.۲: برآورد داده‌های فوران یک چشمه‌ی آب گرم قدیمی در پارک ملی یلوستون آمریکا به روش‌های ساده (قاب بالا) و بافت نگار (قاب پایین).

می‌توان برآوردگر کرنل را با معرفی نماد مقیاسی شده‌ی کرنل به فرم $K_h(u) = h^{-1}K(u/h)$ به صورت رابطه‌ی زیر تعریف کرد:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \qquad (۶.۲)$$

برخی از توابع کرنل معروف در جدول ۱.۲ آمده‌اند.

برآورد کرنل نیز مجموع برآمدگی‌هایی است که مشاهدات در مرکز آنها واقع شده اند. تابع کرنل K شکل برآمدگی‌ها و پهنای باند h، طول آنها را مشخص می کند. برای توضیح بیشتر، شکل ۳.۲ را در نظر بگیرید. این شکل بر اساس پنج مشاهده‌ی ۳، ۴/۵، ۵، ۷ و ۸ و برای تابع کرنل نرمال (جدول ۱.۲)، به‌دست آمده است. در این مثال، K_h یک توزیع نرمال $N(0, h^2)$ در نظر گرفته شده است و بنابراین h نقش مهمی به عنوان عامل مقیاس دارد که پراکندگی تابع کرنل را مشخص می‌کند. برآورد کرنل با مرکزی کردن تابع کرنل مقیاس شده در هر مشاهده ساخته می‌شود. مقدار برآورد کرنل در نقطه‌ی x، میانگین مقدار n کرنل در آن نقطه است. به‌عبارت دیگر، تابع کرنل به داده‌ها در همسایگی یک نقطه‌ی خاص وزن $\frac{1}{n}$ اختصاص می‌دهد. اگر مقادیر به‌دست آمده از کرنل‌های مختلف را با هم ترکیب کنیم بدین معنی است که در ناحیه‌ای که مشاهدات زیادی وجود دارد، با توجه به این‌که انتظار داریم چگالی درست مقدار نسبتا بزرگی داشته باشد، نتیجه می‌گیریم که برآوردگر کرنل باید از مقدار نسبتا بزرگی برخوردار باشد و برعکس، در نواحی که تعداد مشاهدات اندکی وجود دارد، طبیعتا انتظار تاثیر کمتری از کرنل‌ها وجود دارد.



شکل ۳.۲: برآورد تابع کرنل برای پنج مشاهده

برای روشن شدن تاثیر پارامتر هموار سازی h شکل ۴.۲ (واند و جونز ۱۹۹۵) را در نظر بگیرید. در این شکل سه برآورد چگالی به روش کرنل را بر اساس یک نمونه تصادفی ۱۰۰۰ تایی از چگالی زیر نشان می‌دهد

$$f_۱(x) = \frac{۳}{۴}\phi(x) + \frac{۱}{۴}\phi_{۱/۳}(x - \frac{۳}{۲}) \qquad (۷.۲)$$

که در آن $f_۱$، یک توزیع نرمال آمیخته است به‌طوری که داده‌ها با احتمال $\frac{۳}{۴}$ از توزیع $N(۰, ۱)$ ($\phi(x)$) و با احتمال $\frac{۱}{۴}$ از توزیع $N(\frac{۳}{۲}, (\frac{۱}{۳})^۲)$ ($\phi_{۱/۳}(x - \frac{۳}{۲})$) به‌دست آمده‌اند که در آن $\phi(x)$ و $\phi_{۱/۳}(x - \frac{۳}{۲})$ به ترتیب توابع چگالی $N(۰, ۱)$ و $N(\frac{۳}{۲}, (\frac{۱}{۳})^۲)$ هستند. تابع کرنل در هر شکل نشان داده شده است. شکل ۴.۲ با $h = ۰/۰۶$ نشان می‌دهد که پهنای کم کرنل به معنای این است که میانگین‌گیری در هر نقطه بر روی تعداد کمی از مشاهدات انجام شده است. این برآورد، توجه زیادی به داده‌ها داشته و به نمونه اجازه تغییرپذیری نمی‌دهد. چنین برآوردی، کم‌برآوردی (کم‌همواری[۹]) نامیده می‌شود. نمودار ۴.۲ (ب) برآورد کرنل دیگری را در همان داده نشان می‌دهد با این تفاوت که پهنای باند، $h = ۰/۵۴$ است. این پهنای باند، برآورد هموارتری را نتیجه می‌دهد که با توجه به شکل آن، در واقع همواری زیادی مشاهده می‌شود. در نمودار ۴.۲ (ج) با انتخاب پهنای باند $h = ۰/۱۸$ به برآورد بهتری از تابع چگالی دست خواهیم یافت. در این حالت برآورد کرنل همچنان نوسانی است اما ساختار اساسی چگالی درست داده‌ها قابل شناسایی است.

ملاحظه می‌شود برآوردگر کرنل به پارامتر هموارسازی h بستگی دارد. همان‌طوری که در شکل ۴.۲ دیده می‌شود اگر مقدار h خیلی کوچک اختیار شود برآوردگر $f(x)$ بسیار مواج خواهد بود در حالی‌که برای مقادیر بزرگ h این برآوردگر بیش از حد هموار بوده و بیان‌گر رفتار تابع چگالی نخواهد بود.

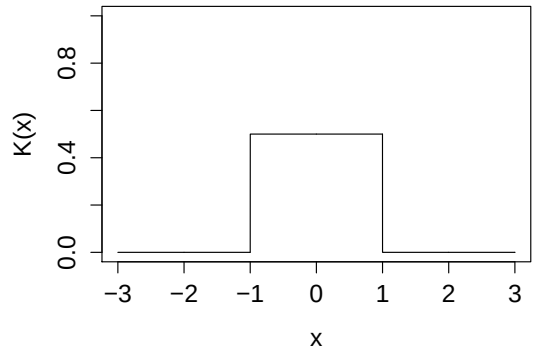
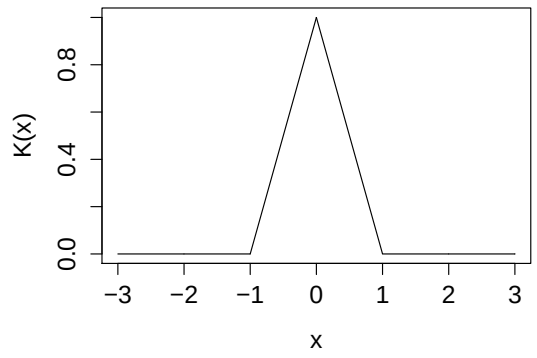
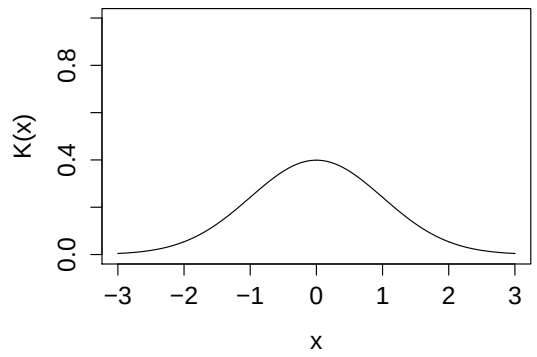
۳.۲ معیارهای ارزیابی

از آن‌جایی که در هر روش برآورد، معیاری برای ارزیابی درستی برآورد لازم است، در این بخش چهار معیار ارزیابی، توان دوم خطا (SE) و میانگین توان دوم خطا (MSE)، توان دوم خطای تجمعی (ISE) و میانگین توان دوم خطای تجمعی (MISE) را معرفی می‌کنیم. برای این منظور فرض کنید $\hat{f}(x)$ برآورد تابع چگالی $f(x)$ در نقطه x به روش کرنل باشد.

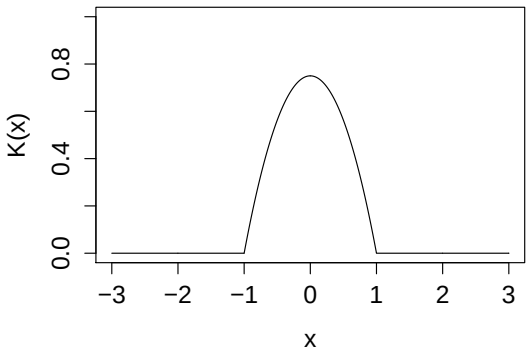
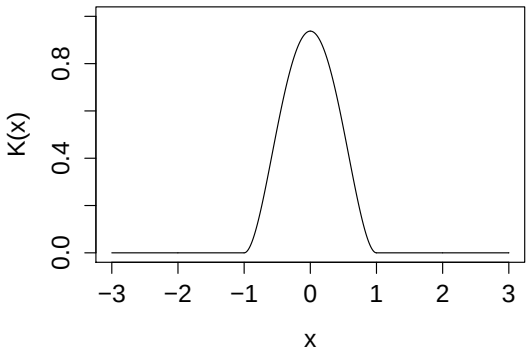
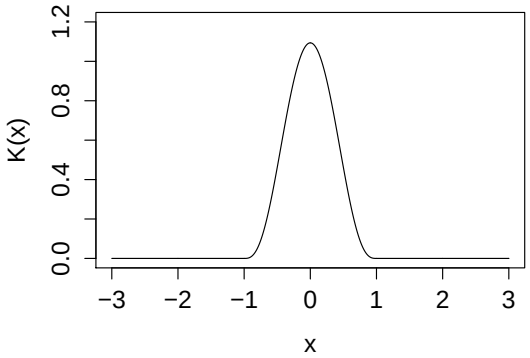
قبل از مرور نحوه انتخاب پارامتر h و هسته K به معیارهای اختلاف چگالی و برآورد آن که برای تجزیه و تحلیل کارایی برآورد چگالی لازم است اشاره می‌کنیم:

[۹] Underestimation

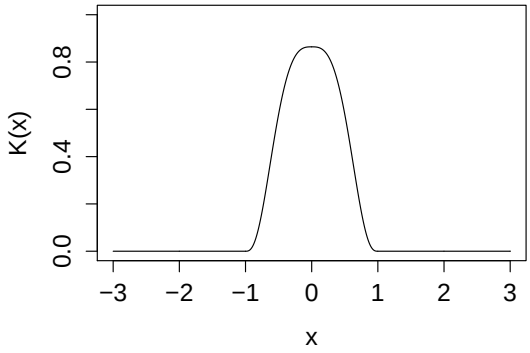
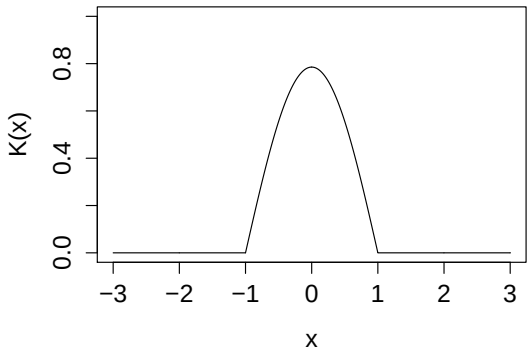
جدول ۱.۲: کرنل‌های معروف در برآورد تابع چگالی

نام کرنل	فرمول	شکل
مستطیلی (یکنواخت)	$K(x) = \frac{۱}{۲} I\left(\lvert x\rvert < ۱\right)$	
مثلثی	$K(x) = (۱ - \lvert x\rvert) I\left(\lvert x\rvert < ۱\right)$	
نرمال (گوسی)	$K(x) = \frac{۱}{\sqrt{۲\pi}} e^{-\frac{۱}{۲}x^۲}, x \in \mathbb{R}$	

جدول ۱.۲ ادامه: کرنل‌های معروف در برآورد تابع چگالی

نام کرنل	فرمول	شکل
اپاچنیکوف	$K(x) = \frac{۳}{۴}(۱ - x^۲) I\left(\lvert X\rvert < ۱\right)$	
توان چهارم (دو وزنی)	$K(x) = \frac{۱۵}{۱۶}(۱ - x^۲)^۲ I\left(\lvert X\rvert < ۱\right)$	
سه وزنی	$K(x) = \frac{۳۵}{۳۲}(۱ - x^۲)^۳ I\left(\lvert X\rvert < ۱\right)$	

جدول ۱.۲ ادامه : کرنل‌های معروف در برآورد تابع چگالی

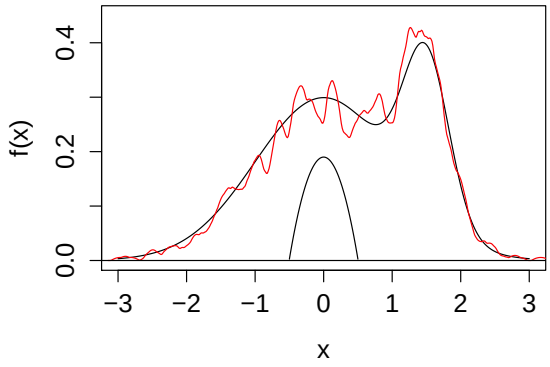
نام کرنل	فرمول	شکل
سه مکعبی	$K(x) = \frac{۷۰}{۸۱}(۱ - \|x\|^۳)^۳ I\left(\|X\| < ۱\right)$	
کسینوسی	$K(x) = \frac{\pi}{۴} cos\left(\frac{\pi}{۲}x\right) I\left(\|X\| < ۱\right)$	

توان دوم خطا

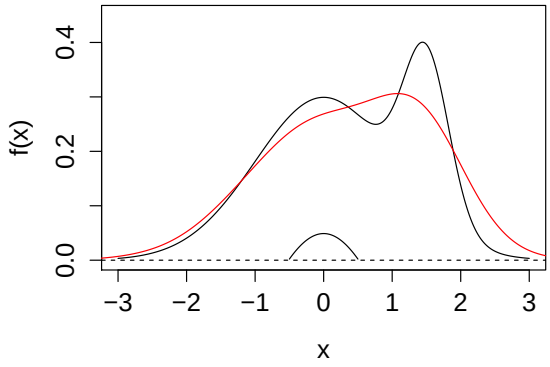
اولین معیار اندازه‌گیری خطا، توان دوم خطا است که به مراتب از قدر مطلق خطا کارایی بیشتری دارد (چه بسا کار در روابط ریاضی با قدر مطلق دشوارتر است) و برای ارزیابی دقت برآورد در یک تک نقطه از آن استفاده می‌شود. توان دوم خطا در نقطه‌ی x برای برآورد $\hat{f}(\cdot)$ به‌صورت زیر محاسبه می‌شود.

$$\mathrm{SE}(\hat{f}(x)) = \left(\hat{f}(x) - f(x)\right)^۲ \qquad (۸.۲)$$

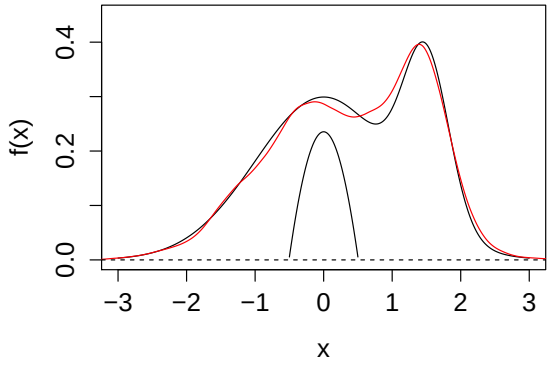
(الف)



(ب)



(ج)



شکل ۴.۲: برآورد کرنل بر اساس یک نمونه‌ی ۱۰۰۰ تایی از توزیع آمیخته نرمال f_1. نمودار مشکی‌رنگ، چگالی واقعی و نمودار قرمز رنگ، برآورد چگالی به ازای پهناهای باند مختلف است. پهنای باند (الف) $h = ۰/۰۶$، (ب) $h = ۰/۵۴$، (ج) $h = ۰/۱۸$ است. وزن کرنل برای هر نمودار با کرنل‌های کوچک، مشخص شده است.

میانگین توان دوم خطا

هنگامی که بخواهیم دقت برآورد را بر اساس یک تک نقطه (یعنی تمام اطلاعات موجود در نمونه‌ی تصادفی X_1، ... و X_n) ارزیابی کنیم و از این دقت برای سایر مجموعه داده‌ها نیز استفاده کنیم، از معیار SE میانگین می‌گیریم و آن را با MSE نمایش می‌دهیم. یکی از جذابیت‌های MSE تجزیه ساده آن به واریانس و اریبی است که ما را قادر می‌سازد که کارآیی برآورد چگالی کرنل را به خوبی نشان دهیم. میانگین توان دوم خطا در نقطه‌ی x برای برآورد $\hat{f}(\cdot)$ به‌صورت زیر محاسبه می‌شود.

$$\begin{aligned} \mathrm{MSE}_x(\hat{f}(x)) &= E\left(\hat{f}(x) - f(x)\right)^2 \\ &= \left\{\mathrm{Bias}(\hat{f}(x))\right\}^2 + \mathrm{Var}\left[\hat{f}(x)\right] \end{aligned}$$

که در آن

$$\begin{aligned} \mathrm{Bias}\hat{f}(x) &= E\left[\hat{f}(x)\right] - f(x) \\ \mathrm{Var}\hat{f}(x) &= E\left(\hat{f}(x) - E\left[\hat{f}(x)\right]\right)^2 \end{aligned}$$

توان دوم خطای تجمعی

هرگاه بخواهیم دقت برآورد را در تمامی نقاط ارزیابی کنیم از معیار ISE استفاده می‌کنیم که در آن

$$\begin{aligned} \mathrm{ISE}_x(\hat{f}(x)) &= \int \left(\hat{f}(x) - f(x)\right)^2 \mathrm{d}x \\ &= \int \hat{f}^2(x)\mathrm{d}x - 2\int \hat{f}(x)f(x)\mathrm{d}x + \int f^2(x)\mathrm{d}x \\ &= \int \hat{f}^2(x)\mathrm{d}(x) - 2E(\hat{f}(x))\mathrm{d}x + \int f^2(x)\mathrm{d}x \end{aligned} \qquad (۹.۲)$$

میانگین توان دوم خطای تجمعی

برای هدف ما که ارزیابی اختلاف $\hat{f}$ و f در کل مجموعه مقادیر x است، معیار مناسب میانگین انتگرال مربع خطا MISE می‌باشد که عبارت است از:

$$\begin{aligned} \mathrm{MISE}_x(\hat{f}(x)) &= \int E\left(\hat{f}(x) - f(x)\right)^2 \mathrm{d}x \\ &= \int \left\{\mathrm{Bias}(\hat{f}(x))\right\}^2 \mathrm{d}x + \int \mathrm{Var}\left[\hat{f}(x)\right]\mathrm{d}x \end{aligned} \qquad (۱۰.۲)$$

لازم به ذکر است که برای محاسبه جملات فرمول بالا نیاز به دانستن چگالی واقعی f است که در واقعیت مجهول است.

۱.۳.۲ خواص مجانبی

در این قسمت معیارهای ارزیابی MSE و MISE را برای برآورد کرنل $\hat{f}(.)$ به‌طور مجانبی محاسبه می‌کنیم. برای این منظور اگر $\hat{f}(.)$ برآوردگر کرنل $f(.)$ در رابطه باشد آن‌گاه:

$$\begin{aligned} E\left[\hat{f}(t)\right] &= \int \frac{1}{h} K\left(\frac{t-y}{h}\right) f(y)\mathrm{d}y \\ \mathrm{Var}\left[\hat{f}(t)\right] &= n^{-1}\int \frac{1}{h^2} K^2\left(\frac{t-y}{h}\right) f(y)\mathrm{d}y - \left\{\frac{1}{h}\int K\left(\frac{t-y}{h}\right) f(y)\mathrm{d}y\right\}^2 \end{aligned}$$

می توان این فرمول‌ها را برای رسیدن به فرمولهای دقیق MSE و MISE به‌کار برد اما جز در موارد خاص، محاسبات، پیچیده شده و فرمول‌های حاصل از آن دارای مفهوم شهودی اندکی می‌گردند لذا بهتر است از تقریب فرمول‌های بالا تحت شرایط مناسبی استفاده شود.

از آنجا که هدف ما یافتن hای است که MISE را کمینه کند، می‌توان با کمینه کردن MISE به h ایده‌آل دست یافت اما حتی برای نمونه‌های خیلی کوچک، h به‌دست آمده از این روش با مقدار hای که از فرمول مجانبی محاسبه می‌شود بسیار نزدیک به هم خواهند بود.

(الف) $\lim_{n\to\infty} h = 0$

(ب) $\lim_{n\to\infty} nh = \infty$

(ج) $\int_{-\infty}^{+\infty} xK(x)dx = 0$

(د) $\int_{-\infty}^{+\infty} x^2 K(x)dx = \sigma_K^2 < \infty$

از آن جایی که صورت انتگرالی $E[\hat{f}(t)]$ را نمی‌توان در محاسبات به‌کار برد در ادامه تقریبی برای $E[\hat{f}(t)]$ را با در نظر گرفتن برقرار بودن شرط‌های **(الف)** و **(ب)** برای سهولت در محاسبات به‌دست می‌آوریم. برای این منظور، داریم

$$\begin{aligned} E(\hat{f}(x)) &= E[K_h(x-X)] \\ &= \int K_h(x-y)f(y)\mathrm{d}y \\ &= \int K(z)f(x-hz)\mathrm{d}z \end{aligned}$$

با استفاده از بسط تیلور،

$$E(\hat{f}(x)) = \int f(x) - hzf'(x) + \frac{1}{2}h^2z^2f''(x) + o(h^2) \qquad (۱۱.۲)$$

در نتیجه

$$E(\hat{f}(x)) = f(x) + \frac{1}{2}h^2 f''(x)\int z^2K(z)\mathrm{d}z + o(h^2)$$

در این صورت تحت شرایط **(ج)** و **(د)** می‌توان نتیجه گرفت

$$\begin{aligned} \mathrm{Bias}[\hat{f}(x)] &= E[\hat{f}(x)] - f(x) \\ &= \frac{1}{2}h^2\sigma_K^2 f''(x) + o(h^2) \end{aligned}$$

از آن‌جایی که اریبی وابسته به $h^۲$ است در حالت مجانبی داریم

$$\begin{aligned}\lim_{n\to\infty}\text{Bias}[\hat{f}(x)] &= \lim_{n\to\infty}\left(\frac{۱}{۲}h^۲\sigma_K^۲ f''(x)\right)+\lim_{n\to\infty}o(h^۲)\\ &= \frac{۱}{۲}\sigma_K^۲ f''(x)\lim_{n\to\infty}h^۲ + ۰ = ۰\end{aligned}$$

و بنابراین $\hat{f}(\cdot)$ به‌طور مجانبی ناریب است.

حال، واضح است که چرا شرط (الف) را در خصوص رفتار مجانبی h به فرضیات مسئله اضافه کردیم. بدیهی است چنان‌چه این فرض اضافه نمی‌شد نمی‌توانستیم اریبی مجانبی برآوردگر کرنل را نشان دهیم.

قضیه ۱.۳.۲. *اگر $\hat{f}(x)$ برآوردگر کرنل در رابطه‌ی (۵.۲) باشد، آن‌گاه واریانس $\hat{f}(x)$ عبارتست از:*

$$Var[\hat{f}(x)] = \frac{۱}{nh}R(K)f(x)+o(\frac{۱}{nh})$$

که در آن برای هر تابع $g(\cdot)$،

$$R(g)=\int g(x)^۲\mathrm{d}x$$

برهان. بر اساس رابطه‌ی (۶.۲) می‌توان نوشت:

$$\begin{aligned}\text{Var}[\hat{f}(x)] &= \frac{۱}{n}\text{Var}[K_h(x-X)]\\ &= \frac{۱}{n}\left\{E[K_h^۲(x-X)]-(E[K_h(x-X)])^۲\right\}\\ &= \frac{۱}{nh}\int K^۲(z)f(x-hz)\mathrm{d}z-\frac{۱}{n}E^۲\left(\hat{f}_h(x)\right) \qquad (۱۲.۲)\end{aligned}$$

با استفاده از تعریف اریبی و بسط تیلور رابطه‌ی (۱۲.۲) داریم

$$\begin{aligned}\text{Var}[\hat{f}(x)] &= \frac{۱}{nh}\int K^۲(z)f(x-hz)\mathrm{d}z-\frac{۱}{n}\left[f(x)+\text{Bias}_h(x)\right]^۲\\ &= \frac{۱}{nh}\int K^۲(z)[f(x)+o(۱)]\mathrm{d}z-\frac{۱}{n}[f(x)+o(۱)]^۲\\ &= \frac{۱}{nh}(\int K^۲(z)\mathrm{d}z)f(x)+o(\frac{۱}{nh})\end{aligned}$$

و اثبات کامل می‌شود. □

بنابراین با استفاده از قضیه ۱.۳.۲، MSE برآوردگر کرنل به‌صورت زیر حاصل می‌شود

$$\begin{aligned}\text{MSE}[\hat{f}(x)] &= \frac{۱}{nh}R(K)f(x)+o\left(\frac{۱}{nh}\right)+\left[\frac{۱}{۲}h^۲\sigma_K^۲ f''(x)+o(h^۲)\right]^۲\\ &= \frac{۱}{nh}R(K)f(x)+\frac{۱}{۴}h^۴\left(\sigma_K^۲ f''(x)\right)^۲+o\left(\frac{۱}{nh}+h^۴\right) \qquad (۱۳.۲)\end{aligned}$$

و در نهایت

$$\text{MISE}[\hat{f}(x)] = \text{AMISE}+o(\frac{۱}{nh}+h^۴)$$

که در آن

$$\text{AMISE}[\hat{f}(x)] = \frac{۱}{nh}R(K) + \frac{۱}{۴}h^۴(\sigma_K^۲)^۲R(f'') \qquad (۱۴.۲)$$

که AMISE را MISE مجانبی می‌نامند زیرا تقریبی را برای MISE در نمونه‌های بزرگ فراهم می‌کند. این معیار در مقایسه با معیار MISE ساده‌تر است.

موازنه بین جملات توان دوم تجمعی، اریبی و واریانس در فرمول MISE

طبق رابطه‌ی (۱۳.۲)، MISE از دو مولفه تشکیل شده است. می‌توان اریبی و واریانس را به‌صورت زیر تقریب زد.

$$\int \text{Bias}_h(x)^۲ \mathrm{d}x \approx \frac{۱}{۴}h^۴(\sigma_K^۲)^۲R(f'')$$

و

$$\int \text{Var}\left[\widehat{f}(x)\right] \mathrm{d}x \approx \frac{۱}{nh}R(K)$$

که در نتیجه

$$\text{MISE} \approx \frac{۱}{nh}R(K) + \frac{۱}{۴}h^۴(\sigma_K^۲)^۲R(f'') \qquad (۱۵.۲)$$

حال اگر بخواهیم h را به گونه‌ای انتخاب کنیم که میانگین انتگرال توان دوم خطا تا حد امکان کوچک شود، با مقایسه دو مولفه‌ی MISE، یکی از مشکلات اساسی برآورد چگالی ظاهر می‌شود. از یک طرف اگر سعی شود که اریبی کوچک گردد، با توجه به این‌که مقدار اریبی متناسب با $h^۴$ است، مقدار خیلی کوچک h مورد نیاز خواهد بود که این انتخاب، منجر به بزرگ شدن مولفه واریانس می‌شود، از طرف دیگر با انتخاب مقدار بزرگ برای h پراکندگی تصادفی که با واریانس نشان داده می‌شود، کاهش پیدا می‌کند زیرا واریانس متناسب با $۱/nh$ است و بنابراین این کار به قیمت ایجاد خطای سیستماتیک یعنی اریبی تمام خواهد شد. در استفاده از هریک از روش‌های برآورد چگالی، انتخاب پارامتر هموارسازی موازنه‌ای را بین خطای سیستماتیک (اریبی) و خطای تصادفی (واریانس) ایجاد می‌کند. لذا هنگامی‌که n به سمت بی‌نهایت میل می‌کند، h باید به گونه‌ای تغییر کند که هر یک از مولفه‌های MISE کوچکتر شود. معمولا از آن به عنوان موازنه‌ی واریانس–اریبی[۱۰] نام می‌برند و از آن به عنوان یک معیار ریاضی برای تعیین پهنای باند استفاده می‌شود.

برای hهای خیلی کوچک، $\hat{f}(\cdot)$ خیلی نوسانی[۱۱] است و بنابراین در این حالت، به ازای نمونه‌های متفاوت تغییرپذیر است. در نمونه‌گیری از توزیع f، پله‌ها در مکان‌های متفاوتی قرار می‌گیرند و البته در این حالت، اریبی خیلی کمی وجود دارد. اگر هموارسازی بیشتری به‌کار گرفته شود، h افزایش خواهد یافت و در نتیجه، تغییرپذیری به قیمت به وجود آمدن اریبی، کاهش خواهد یافت. برای مقادیر بزرگ h، اریبی افزایش می‌یابد.

[۱۰] Variance-biased tradeoff
[۱۱] Spiky

پهنای باند ایده‌آل

برای دستیابی به h بهینه بر اساس معیار MISE، کافی است از رابطه‌ی (۱۵.۲) نسبت به h مشتق بگیریم و آن را برابر صفر قرار دهیم. یعنی

$$\begin{aligned}\frac{\partial \text{MISE}}{\partial h} &= \frac{-۱}{nh^۲}R(K) + h^۴(\sigma_K^۲)^۲R(f'') = ۰ \\ \Longrightarrow \quad & \frac{-R(K) + nh^۵(\sigma_K^۲)^۲R(f'')}{nh^۲} = ۰ \\ \Longrightarrow \quad & -R(K) + nh^۵(\sigma_K^۲)^۲R(f'') = ۰ \\ \Longrightarrow \quad & h^۵ = \frac{R(K)}{n(\sigma_K^۲)^۲R(f'')}\end{aligned}$$

در این صورت

$$h_{opt} = \left[\frac{R(K)}{n(\sigma_K^۲)^۲R(f'')}\right]^{\frac{۱}{۵}} \qquad (۱۶.۲)$$

صرف‌نظر از وابستگی h_{opt} به K و n، رابطه‌ی (۱۶.۲) نشان می‌دهد که چنان‌چه حجم نمونه افزایش یابد، پهنای باند به سمت صفر میل می‌کند و هم‌چنین h_{opt} به‌طور معکوس با $R(f'')^{\frac{۱}{۵}}$ متناسب است. از آن‌جایی‌که $|f''(x)|$ معیاری برای اندازه‌گیری خمیدگی f است، $R(f'')$ خمیدگی کل تابع را اندازه می‌گیرد. بنابراین برای یک توزیع با خمیدگی کم، $R(f'')$ کوچک خواهد شد و نیاز به پهنای باند بزرگتری وجود دارد. وقتی $R(f'')$ بزرگ باشد، بدین معنی است که هموارسازی کم، بهینه خواهد بود. متاسفانه در عمل نمی‌توان مستقیما از رابطه‌ی (۱۶.۲) برای انتخاب یک پهنای باند خوب استفاده کرد زیرا $R(f'')$ مجهول است. با این وجود، روش‌هایی برای انتخاب پهنای باند وجود دارد که مبتنی بر برآورد $R(f'')$ هستند. برای آگاهی بیشتر واند و همکاران (۱۹۹۱) را ببینید.

به‌طور کلی، رابطه (۱۶.۲) برای h بهینه دارای این اشکال است که به چگالی مجهول وابسته است، اما می‌توان نتایج مفیدی را از آن اتخاذ کرد.

بررسی اثر تغییر تابع کرنل

برای کوچک کردن h منحنی حاصل از برآورد کرنل ناهموارتر شده و جزئیات جعلی بیشتری را از چگالی واقعی به نمایش می‌گذارد و به ازای بزرگ کردن h منحنی هموار و باعث محو شدن جزئیات واقعی تابع چگالی می‌گردد. لذا، انتخاب پارامتر هموارسازی مناسب از اهمیت خاصی در برآورد چگالی برخوردار است چند روش برای این انتخاب وجود دارد: انتخاب ذهنی، مراجعه به توزیع نرمال، اعتبارسنجی متقابل کمترین توان‌های دوم (LSCV[۱۲])، اعتبارسنجی متقابل تصحیح‌شده (MCV[۱۳])، اعتبارسنجی متقابل اریب (BCV[۱۴]) و جای‌گذاری. در فصل ۳ به بررسی و مقایسه روش‌های انتخاب پارامتر هموارسازی می‌پردازیم.

در ادامه، در این بخش اثر شکل تابع کرنل K را بر برآورد تابع چگالی بررسی می‌کنیم. تقریبا در بیشتر موارد، K یک چگالی تک مدی و متقارن مانند توزیع نرمال در نظر گرفته می‌شود. کلاین[۱۵] (۱۹۹۸) نشان داد که

[۱۲] Least squares cross validation
[۱۳] Modified cross validation
[۱۴] Biased cross validation
[۱۵] Cline

برآوردگرهای تابع چگالی که بر اساس کرنل‌هایی که در دو شرط تقارن و تک مدی بودن صدق نکند، غیرمجاز هستند. مطمئنا توابع زیادی وجود دارند که متقارن و تک مدی هستند و بنابراین می‌توانند به عنوان یک تابع کرنل باشند، سوال این است که آیا استفاده از کرنلی با یک شکل خاص، بهتر از سایر کرنل‌هاست؟ برای پاسخ به این سوال، صورت مقیاسی شده‌ی کرنل را بررسی می‌کنیم که منجر به انجام مقایسه‌های بهتر در بین صورت‌های مختلف کرنل‌ها می‌شود.

اگر فرمول پهنای باند ایده‌آل، رابطه‌ی (۱۶.۲)، را در فرمول (۱۵.۲) جایگذاری کنیم، آن‌گاه

$$\text{MISE} \approx \frac{۵}{۴} C(K) R(f'')^{\frac{۱}{۵}} n^{-\frac{۴}{۵}}$$

که در آن $C(K)$ به صورت زیر محاسبه می‌شود:

$$C(K) = \left(\mu_۲^۲\right)^{\frac{۲}{۵}} R(K)^{\frac{۴}{۵}}$$

همان‌طور که مشاهده می‌شود تنها عبارت $C(K)$ به تابع کرنل وابسته است. بنابراین باید کرنل K به گونه‌ای انتخاب شود که $C(K)$ کمترین مقدار را داشته باشد.

مسئله‌ی تعیین کرنل بهینه با کمینه کردن $C(K)$ نسبت به شرط‌های (ج) و (د) برای تابع کرنل به‌دست می‌آید. هاج و لهمن[۱۶] (۱۹۵۶) نشان دادند که کرنل بهینه می‌تواند به‌صورت زیر باشد:

$$K^a(x) = \frac{۳}{۴}\left(۱ - \frac{x^۲}{۵a^۲}\right) / \left(\sqrt{۵}a\right) I_{\{|X|<\sqrt{۵}a\}}$$

که در آن a یک عدد دلخواه برای پارامتر مقیاس و $I(\cdot)$ یک تابع نشان‌گر است. به ازای $a^۲ = \frac{۱}{۵}$، ساده‌ترین شکل $K^a(x)$ به‌دست می‌آید که همان تابع کرنل اپانیچکوف است و در این‌جا آن را با K^* نمایش می‌دهیم. کارآیی کرنل دلخواه K را نسبت به K^* به‌صورت زیر تعریف می‌کنیم

$$\left\{\frac{C(K^*)}{C(K)}\right\}^{۵/۴} \qquad (۱۷.۲)$$

در جدول ۲.۲، مقادیر کارایی کرنل‌های معرفی شده در جدول ۱.۲ آمده است. به عنوان مثال کرنل دو وزنی را در نظر بگیرید. کارآیی این کرنل ۰/۹۵ محاسبه شده است و این بدین معنی است که برآورد با استفاده از کرنل اپانیچکوف، همان کمینه MISE را نتیجه می‌دهد که از ۹۵٪ داده‌ها به ازای کرنل K استفاده کنیم.

نکته قابل توجه در جدول ۲.۲ این است که کارائی های نشان داده شده همگی به یک نزدیکند. حتی برای کرنل مستطیلی که برای ساختن برآوردگر ساده استفاده می شود ، این مقدار حدوداً ۰/۹۳ است. درنتیجه این نکته استنباط می شود که تفاوت بسیار اندکی در قدرت انتخاب بین کرنل ها بر اساس معیار میانگین توان دوم خطای تجمعی وجود دارد. بنابراین برای انتخاب کرنل مناسب بهتر است ملاحظات دیگری مانند درجه مشتق‌پذیری مورد نیاز، متناهی یا نامتناهی بودن تکیه‌گاه یا میزان پیچیدگی آن در محاسبات استفاده شود.

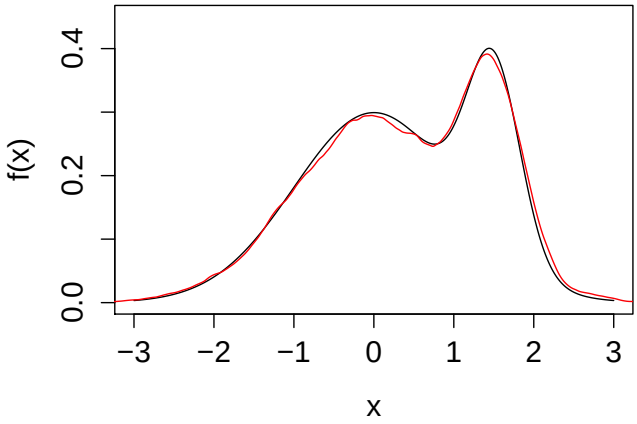
برای وضوح بیشتر در مورد عدم تاثیر انتخاب کرنل، مجددا توزیع آمیخته‌ی نرمال $f_۱$ در رابطه‌ی (۷.۲) را در نظر بگیرید. نمودار ۵.۲ (الف) و (ب)، به ترتیب برآورد کرنل ۱۰۰۰ داده‌ی تصادفی از این توزیع را به ازای کرنل‌های اپانیچکوف و نرمال نشان می‌دهد. ملاحظه می‌شود که تقریبا دو برآورد، یکسان هستند.

[۱۶] Hodge and Lehma n

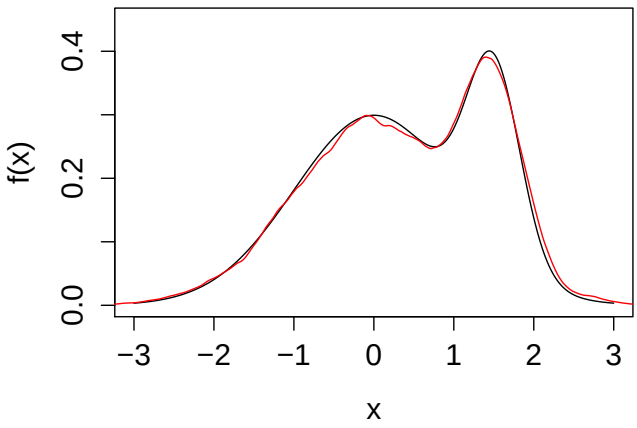
جدول ۲.۲: کارآیی چندین کرنل در مقایسه با کرنل بهینه (اپانیچکوف)

کرنل	$R(K)$	$\mu_2(K)$	کارایی
اپاچنیکوف	$\frac{3}{5}$	$\frac{1}{5}$	۱/۰۰۰
توان چهارم (دو وزنی)	$\frac{5}{7}$	$\frac{1}{7}$	۰/۹۹۴
سه وزنی	$\frac{350}{429}$	$\frac{1}{9}$	۰/۹۸۷
نرمال	$\frac{1}{\sqrt{2\pi}}$	۱	۰/۹۵۱
سه مکعبی	$\frac{175}{2447}$	$\frac{35}{243}$	۰/۹۸۶
یکنواخت	$\frac{1}{2}$	$\frac{1}{3}$	۰/۹۳۰

(الف)



(ب)



شکل ۵.۲: برآورد تابع چگالی با کرنل‌های متفاوت

فصل ۳

روش‌های تعیین پهنای باند

۱.۳ مقدمه

انتخاب پهنای باند، دشوارترین مرحله در ساخت برآوردگر چگالی به روش کرنل می‌باشد (توروست[۱]، ۲۰۱۱). انتخاب مقدار بزرگی برای h، یک انحراف استاندارد کوچک را نتیجه می‌دهد و کرنل، بیشترین احتمال را به هر نقطه نسبت می‌دهد. این انتخاب هنگامی مناسب است که اندازه‌ی نمونه بزرگ بوده و داده‌ها به‌صورت فشرده[۲] پخش شده باشند. h کوچک، منجر به یک انحراف استاندارد بزرگ می‌شود و کرنل احتمال‌های بیشتری را در مقادیر همسایگی یک نقطه گسترده می‌کند. از این روش وقتی استفاده می‌شود که اندازه‌ی نمونه کوچک باشد و داده‌ها به‌صورت تنک[۳] گسترده شده باشند.

باید این نکته در نظر گرفته شود که انتخاب پهنای باند مناسب، همیشه تحت‌تاثیر هدفی است که از برآورد چگالی استفاده می‌شود. اگر هدف برآورد چگالی، جست و جو کردن داده‌ها به منظور پیشنهاد کردن مدل‌ها و فرضیه‌های مختلف است، آن‌گاه احتمالا استفاده از روش‌های ذهنی[۴] در انتخاب پهنای باند کاملا مناسب خواهد بود. وقتی از برآورد چگالی برای نشان دادن نتایج استفاده می‌شود، می‌توان با چشم و از لحاظ بصری در خصوص همواری اظهارنظر نمود اما به راحتی نمی‌توان در مورد ناهمواری نظر داد.

با این وجود، در موارد بسیاری به یک انتخاب خودکار[۵] از پارامتر هموارسازی نیاز داریم. برای مبتدیان روش‌های کاملا خودکار بهتر است و یک انتخاب خودکار در هر مورد می‌تواند به‌عنوان یک نقطه‌ی شروع و اولیه برای یک دنباله از پیشنهادهای ذهنی باشد.

اگر برآورد چگالی به‌صورت منظم روی مجموعه داده‌های بزرگ یا به عنوان قسمتی از مجموعه داده‌های بزرگ به‌کار رود، آن‌گاه استفاده از یک روش خودکار کاملا ضروری است. استفاده از واژه‌ی «خودکار» نسبت به «ذهنی» برای روش‌هایی که نیاز به مشخص کردن مقدار صریحی برای کنترل پارامترها دارند، مطلوب‌تر است.

---

[۱] Torosset
[۲] Tightly
[۳] Sparse
[۴] Subjective methods
[۵] Automate

فرآیند انتخاب پهناهای باند با این مشکل مواجه است که ممکن است کاربر اطلاعات کافی را نسبت به فرضیه‌های پیشین نداشته باشد.

در این فصل چند روش انتخاب پارامتر هموارسازی بیان می‌شود. به‌جز روش اول که انتخاب ذهنی می‌باشد، روش‌های دیگر روش‌های خودکار برای انتخاب پارامتر هموارسازی هستند. همان‌طور که جلوتر به آن اشاره شد، انتخاب روش مناسب بستگی به هدفی دارد که برآورد چگالی به منظور آن انجام می‌شود. اگر هدف از برآورد چگالی کنکاش در داده‌ها و ارائه‌ی الگویی مناسب برای داده‌ها باشد معمولاً انتخاب پارامتر هموارسازی به صورت ذهنی مناسب و مطلوب خواهد بود ولی در بعضی موارد لازم است که یک روش خودکار در انتخاب پارامتر هموارسازی در اختیار داشته باشیم. روش های خودکار بی تردید برای کاربران بی تجربه راضی کننده تر خواهند بود و همچنین اگر از برآورد چگالی به عنوان کاری مابین کارهای دیگر استفاده شود که نتیجه‌ی آن در قسمت های بعد نیز مورد نیاز باشد، یک روش خودکار برای انتخاب پارامتر هموارسازی ضروری است. مطالب این فصل، عمدتا برگرفته از سیلورمن (۱۹۸۶) و حیدرنیک و همکاران (۲۰۱۳) می‌باشند.

۲.۳ روش‌های تعیین پهنای باند

تاکنون هنوز هیچ روش مناسبی برای تعیین پهنای باند ارائه نشده که تواما نتایج مطلوبی در تئوری و کاربرد به همراه داشته باشد و در این راستا، تحقیقات زیادی در جهت بهبود روش‌های موجود انتخاب پهنای باند انجام گرفته است.

معمولا روش‌های انتخاب پهنای باند بر اساس روش به کار برده شده به دو دسته‌ی کلی (الف) جای‌گذاری و (ب) اعتبارسنجی متقابل تقسیم می‌شوند. در ادامه هر یک از این روش‌ها و هم‌چنین ترکیب آن‌ها که اخیرا توسط حیدرنیک و همکاران (۲۰۱۳) معرفی شده است، به تفصیل مطرح می‌شود.

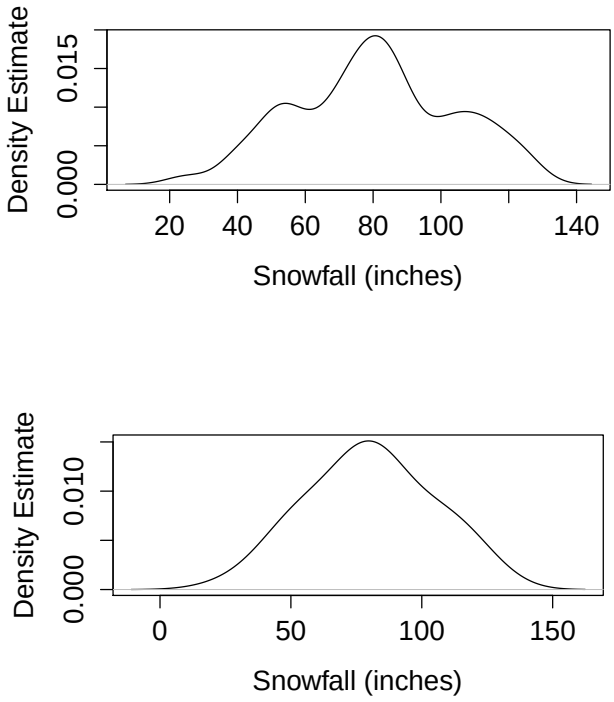
۱.۲.۳ روش جای‌گذاری

همان‌طور که از نام این روش مشخص است، اگر عبارتی شامل یک پارامتر مجهول باشد، پارامتر مجهول را با یک برآوردگر، جای‌گذاری می‌کنیم. رابطه‌ی (۱۶.۲)، برای به‌دست آوردن h_{opt} را در نظر بگیرید. همان‌طور که ملاحظه می‌شود، این رابطه، شامل کمیت مجهول $R(f'')$ است. می‌توان برای دستیابی به پهنای باند بهینه h_{opt} با استفاده از روش‌های مختلفی، f را جای‌گذاری نمود و $R(f'')$ را بر اساس آن به‌دست آورد. این روش‌ها به روش‌های «سریع و آلوده[۶]» شناخته می‌شوند که اولین بار توسط سیلورمن (۱۹۸۶) معرفی شد. این روش‌ها نیز در سه دسته‌بندی قرار می‌گیرند: (الف) انتخاب ذهنی، (ب) قاعده‌ی سرانگشتی[۷]، (ج) اصل هموارسازی بیشینه[۸].

انتخاب ذهنی

یک روش طبیعی برای انتخاب پارامتر هموارسازی، رسم چند نمودار و انتخاب برآوردی است که بیش‌ترین تطبیق

[۶] Quick and dirty methods
[۷] Rule of thumb
[۸] Maximum Smothness Principle



شکل ۱.۳: برآورد کرنل برای داده‌های سالانه‌ی بارش برف دربوفالو ایالت نیویورک، پهناهای باند: تصویر بالا، $h = ۶$ و تصویر پایین: $h = ۱۲$.

را با ایده‌های اولیه دارد. در بسیاری از موارد کاربردی، این رهیافت، کاملا رضایت‌بخش است. به‌عنوان مثال، در بسیاری از پدیده‌های اقتصادی، از توزیع پاراتو برای توصیف داده‌ها استفاده می‌شود.

رسم کردن چندین نمودار از داده‌ها که با مقادیر متفاوتی هموار شده‌اند، ممکن است منجر به بینش و شناخت بیشتری نسبت به داده‌ها در مقایسه با زمانی‌که یک منحنی با استفاده از یک پهنای باندی که به‌طور خودکار انتخاب شده، به‌دست آمده است، شود.

به‌عنوان یک مثال، برآورد در شکل ۱.۳ را در نظر بگیرید. داده‌های در نظر گرفته شده برای این برآوردها، میزان ریزش برف (بر حسب اینچ) در بوفالو، نیویورک برای ۶۳ زمستان از سال ۱۹۱۰/۱۱ تا ۱۹۷۲/۷۳ را نشان می‌دهد (برگرفته شده از سیلورمن، ۱۹۸۶). از نمودار ۱.۳ این چنین برمی‌آید که تغییر پارامتر هموارسازی اساسا منجر به دو تفسیر ممکن برای داده‌ها می‌شود. یک توزیع کاملا نرمال یا یک منحنی سه‌مدی، یک توزیع آمیخته از جامعه را نشان می‌دهد که تقریبا به‌نسبت‌های ۱، ۳ و ۱ مخلوط شده‌اند می‌تواند پیشنهادهای مناسبی باشند. در اهداف بسیاری، به‌ویژه برای تولید مدل و فرضیه‌ها، هیچ وسیله‌ی مفیدی برای آماره‌ها وجود ندارد که دانشمندان را در زمینه‌های مختلف حمایت کند. یک انتخاب بین دو مدل پیشنهاد شده با نمودار ۱.۳ یک مرحله‌ی خیلی مفید پیشرو از بین تعداد زیادی از تفسیرهای ممکنی هست که می‌تواند در نظر گرفته شود.

مراجعه به یک توزیع استاندارد

یک رهیافت ساده و طبیعی استفاده از یک خانواده‌ی استاندارد از توزیع‌هاست تا مقدار عبارت $R(f'')$ در رابطه‌ی (۱۶.۲) را تعیین کند. به عنوان مثال، فرض کنید می‌دانیم (یا به عنوان یک فرضیه) چگالی مجهول f، متعلق به خانواده‌ی توزیع‌های نرمال با میانگین صفر و واریانس $\sigma^۲$ باشد. در این صورت

$$\begin{aligned} R(f'') &= \sigma^{-۵} \int \{\phi''(x)\}^۲ \, \mathrm{d}x \\ &= \sigma^{-۵} \frac{۳}{۸} \pi^{-\frac{۱}{۲}} \approx ۰/۲۱۲ \sigma^{-۵} \end{aligned} \tag{۱.۳}$$

که در آن $\phi''(x)$ مشتق دوم تابع چگالی نرمال استاندارد نسبت به x است.

با در نظر گرفتن کرنل نرمال در رابطه‌ی (۱۶.۲)، پهنای باند بهینه با جای‌گذاری رابطه‌ی (۱.۳) به‌جای عبارت مجهول $R(f'')$ برابر است با:

$$\begin{aligned} h_{opt} &= (۴\pi)^{-\frac{۱}{۱۰}} \left(\frac{۳}{۸} \pi^{-\frac{۱}{۲}} \right)^{-\frac{۱}{۵}} \sigma n^{-\frac{۱}{۵}} \\ &= \left(\frac{۴}{۳n} \right)^{\frac{۱}{۵}} \sigma = ۱/۰۶ \sigma n^{-\frac{۱}{۵}} \end{aligned} \tag{۲.۳}$$

سوالی که ممکن است در این‌جا مطرح شود این است که فرض کردن توزیع نرمال آیا بر خلاف فلسفه‌ی روش‌های ناپارامتری برآورد چگالی نیست؟ در واقع، اگر بدانیم که X به‌طور نرمال توزیع شده است، می‌توان چگالی آن را بسیار ساده‌تر و کاراتر برآورد کرد. بدین منظور کافیست به‌سادگی μ را با میانگین نمونه و $\sigma^۲$ را با واریانس نمونه برآورد کرد و این برآوردها را در فرمول چگالی توزیع نرمال جای‌گذاری کرد.

آن‌چه در این روش با در نظر گرفتن فرضیه‌ی نرمال به‌دست می‌آید یک فرمول صریح و کاربردی برای انتخاب پهنای باند است. در کاربرد، نمی‌دانیم که آیا X به‌طور نرمال توزیع شده است یا خیر. اگر چنین باشد، آن‌گاه رابطه‌ی (۱۶.۲)، پهنای باند بهینه را نتیجه می‌دهد. اما اگر به صورت نرمال توزیع نشده باشد، در این صورت، اگر توزیع X خیلی از توزیع نرمال (توزیع مرجع[۹]) فاصله نداشته باشد، آن‌گاه h_{opt} به‌دست آمده از رابطه‌ی (۱۶.۲) پهنای باندی را نتیجه می‌دهد که از پهنای باند بهینه فاصله‌ی چندانی ندارد. به این دلیل، hای که از رابطه‌ی (۲.۳) به‌دست می‌آید، قاعده‌ی سرانگشتی سیلورمن، نتایج منطقی برای همه‌ی توزیع‌های تک‌مدی، کاملا متقارن و با دم‌های باریک ارائه می‌دهد.

آن‌چه در رابطه‌ی (۲.۳) مجهول است، انحراف معیار σ است و بنابراین انتخاب یک برآوردگر مناسب برای انحراف معیار از اهمیت بسزایی برخوردار است. یک برآوردگر معمولی برای انحراف معیار می‌تواند انحراف معیار نمونه باشد. به‌عبارت دیگر،

$$\hat{\sigma} = \sqrt{\frac{۱}{n-۱} \sum_{i=۱}^{n} (x_i - \bar{x})^۲}$$

و بنابراین h_{opt} در این حالت برابر است با

$$\hat{h}_{rot} = ۱/۰۶ \hat{\sigma} n^{-\frac{۱}{۵}} \tag{۳.۳}$$

[۹] Refrence distribution

اما نتایج بهتری برای انتخاب پهنای باند به ازای یک برآوردگر تنومند[۱۰] (که نسبت به داده‌های پرت حساس نباشد) از پراکندگی به‌دست می‌آید. در واقع، حساسیت این روش نسبت به داده‌های پرت، یک مسئله‌ی اساسی است. یک داده‌ی پرت به‌تنهایی باعث می‌شود برآورد بزرگی از σ به‌دست آید و در نتیجه به برآورد بزرگتری از پهنای باند در برآورد تابع چگالی دست پیدا می‌کنیم. برآوردگر تنومند انحراف معیار بر اساس دامنه‌ی میان‌چارکی[۱۱] به صورت زیر به‌دست می‌آید.

$$R = X_{[۰/۷۵n]} - X_{[۰/۲۵n]}$$

به عبارت دیگر، به‌سادگی دامنه‌ی میان‌چارکی نمونه بر اساس چارک سوم $(X_{[۰/۷۵n]})$ و چارک اول $(X_{[۰/۲۵n]})$ به‌دست می‌آید.

با توجه به این‌که توزیع واقعی داده‌ها نرمال در نظر گرفته شده است، می‌دانیم که اگر $X \sim N(\mu, \sigma^۲)$، آن‌گاه $Z = \frac{X-\mu}{\sigma} \sim N(۰, ۱)$ و بنابراین به‌طور مجانبی،

$$\begin{aligned} R &= X_{[۰/۷۵n]} - X_{[۰/۲۵n]} \\ &= \left(\mu + \sigma Z_{[۰/۷۵n]}\right) - \left(\mu + \sigma Z_{[۰/۲۵n]}\right) \\ &= \sigma\left(Z_{[۰/۷۵n]} - Z_{[۰/۲۵n]}\right) \\ &\approx \sigma\left(۰/۶۷ - (-۰/۶۷)\right) = ۱/۳۴\sigma \end{aligned}$$

و بنابراین،

$$\hat{\sigma} = \frac{R}{۱/۳۴} \qquad (۴.۳)$$

با جای‌گذاری رابطه‌ی (۴.۳) در رابطه‌ی (۲.۳)، داریم:

$$\hat{h}_{rot} = ۱/۰۶ \left(\frac{R}{۱/۳۴}\right) n^{-\frac{۱}{۵}} \approx ۰/۷۹ R n^{-\frac{۱}{۵}} \qquad (۵.۳)$$

می‌توان برای دستیابی به نتایج بهتری از قاعده‌ی سرانگشتی سیلورمن، روابط (۳.۳) و (۴.۳) را با هم ترکیب کرد و رابطه‌ی زیر را برای دستیابی به h_{opt} بیان کرد.

$$\hat{h}_{rot} = ۱/۰۶ \min\left\{\hat{\sigma}, \frac{R}{۱/۳۴}\right\} n^{-\frac{۱}{۵}} \qquad (۶.۳)$$

استفاده از رابطه‌ی (۶.۳) و حتی رابطه‌ی (۲.۳) کاملا نتایج خوبی دارند به‌خصوص اگر چگالی واقعی، نرمال باشد اما همان‌طور که قبلا اشاره شد، اگر چگالی واقعی از نرمال فاصله داشته باشد (به‌عنوان مثال، چندمدی باشد)، آن‌گاه ممکن است استفاده از این روش برای انتخاب پهنای باند منجر به نتایج گمراه‌کننده‌ای شود.

انتخاب پهنای باند بیش هموار

همان‌طور که می‌دانیم با بزرگ شدن h، نمودار برآورد تابع چگالی دچار همواری بیشتری می‌شود در نتیجه جزئیات واقعی بیشتری از تابع چگالی نادیده گرفته می‌شود. پس با در نظر گرفتن حد بالای پهنای باند و کوچک کردن آن تا حد امکان می‌توان پهنای باند مناسب‌تر را پیدا کرد.

[۱۰] Robust
[۱۱] Interquantile range

در بیش‌هموار یا همواری ماکزیمم معمولاً از کران بالای AMISE مطلوب یک انتخاب‌گر استفاده می‌شود. ترل (۱۹۹۰) نشان داد برای همه‌ی چگالی‌هایی با انحراف معیار σ داریم

$$h_{opt} \leq \left[\frac{۲۴۳R(K)}{۳۵n(\sigma_K^۲)^۲}\right]^{\frac{۱}{۵}} \sigma$$

که در آن $R(K) = \int K^۲(x)\,\mathrm{d}(x)$

کران بالای عبارت مذکور با چگالی سه‌وزنی یا بتای $B(۴,۴)$ به‌دست می‌آید.

بنابراین انتخاب پهنای باند بیش‌هموار به‌صورت زیر حاصل می‌شود.

$$\hat{h}_{os} = \left[\frac{۲۴۳R(K)}{۳۵n(\sigma_K^۲)^۲}\right]^{\frac{۱}{۵}} s$$

که در آن s انحراف معیار نمونه است. با استفاده از $\hat{h}_{os}$ می‌توان انتخاب ذهنی عالی برای پهنای باند داشته باشیم و یک روش گرافیکی معقول رسم نمودار با استفاده از پهنای باند $\hat{h}_{os}$ می‌باشد. در ادامه مثالی را بیان می‌کنیم که در آن با تقسیم متوالی پهنای باند شاهد تغییر در میزان همواری تابع هستیم. از مجموعه داده‌های زمان فوران ۱۰۷ آبفشان در پارک ملی یلوستون (سیلورمن، ۱۹۸۶) برای برآورد استفاده شده است. در شکل ۲.۳، (الف) برآورد تابع چگالی با استفاده از کرنل گوسی با $\hat{h}_{os} = ۰/۴۶۷$ انجام شده است نکته‌ی مهم در این شکل دو مدی بودن آن است که مبین بیش‌هموار بودن آن می‌باشد و مرکز دسته‌ها در حدود ۱/۹ دقیقه و ۴/۲ دقیقه می‌باشد. با کاهش دادن پهنای باند در شکل ۲.۳، (ب) با استفاده از $\frac{\hat{h}_{os}}{۲}$ می‌توان حفظ ساختار دو مدی را مشاهده کرد اما در شکل ۲.۳، (ج) مشاهده خواهیم کرد برای رفع حالت دو نمایی احتیاج به تقسیمات بیشتری است. در شکل ۲.۳، (ج) از $\frac{\hat{h}_{os}}{۴}$ استفاده شده است که ره‌آورد آن ۵ مدی بودن توزیع داده‌ها می‌باشد. در شکل ۲.۳، (د) با استفاده از $\frac{\hat{h}_{os}}{۵}$ برآورد دچار کاهش همواری می‌شود که مدل دور از ذهنی برای زمان فوران آبفشان می‌باشد.

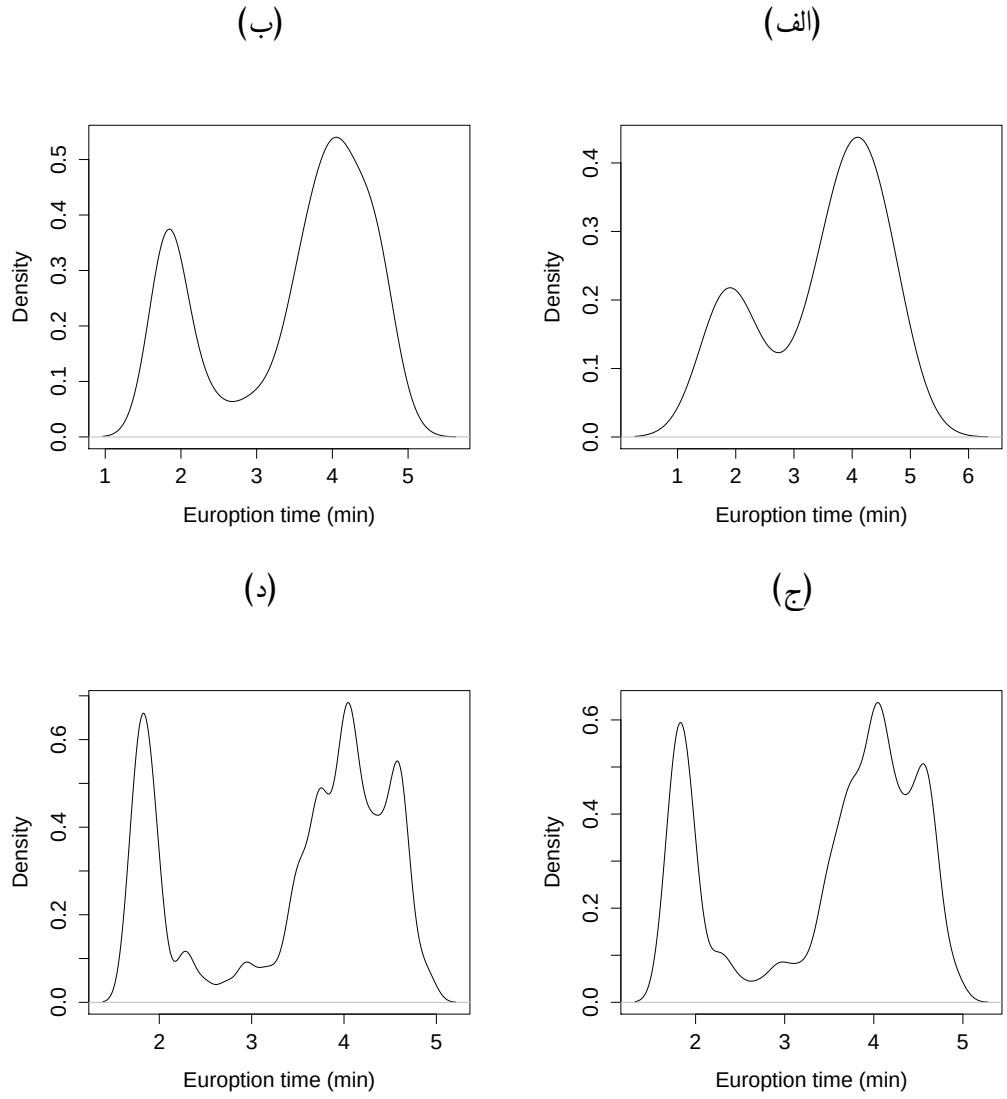
در بین مدل‌های ارائه شده در شکل ۲.۳، شکل (ب) بهترین نمایش برای داده‌ها می‌باشد و تغییرات داده‌ها را بهتر نشان می‌دهد.

انتخاب‌گرهای پهنای باند توزیع گوسی و بیش‌هموار مبتنی بر انحراف‌معیار هستند و در مفهوم به هم نزدیک می‌باشند.

در ادامه مروری خواهیم داشت بر روشهای جای‌گذاری و CV و ترکیبی از جای‌گذاری ساده و CV.

در ادامه بررسی‌ها، محدودیت‌های زیر را در نظر می‌گیریم: اولاً فقط مشاهدات مستقل را در نظر می‌گیریم، دوماً روش‌های مبتنی بر $L_۲$ را در نظر می گیریم (اگر معیار اندازه‌گیری فاصله $(X_۲ - X_۱)^۲$ باشد با $L_۲$ نمایش داده می‌شود و اگر معیار اندازه‌گیری $|X_۲ - X_۱|$ باشد با $L_۱$ نمایش داده می‌شود) سوماً مسئله و مشکل مرزی[۱۲] را بیان نمی‌کنیم چون بیان چگونگی ترکیب شدن با مشکلات انتخاب پهنای باند سخت است. ما ترجیحاً بر روی نمونه‌های کوچک و مناسب و چگالی‌های هموار متمرکز می‌شویم. درجه صحیح همواری محدوده‌ی عریضی از مشکلات را در هر ناحیه‌ی پژوهشی پنهان می‌کند اما توابع نوک‌تیز و پرنوسان مستثنی هستند. توجه کنید که مشکلات اخیر نباید به عهده کرنل‌ها باشد. این معلوم است که باید مشکلات با تغییر داده‌ها حل شوند. انتقاد اصلی

[۱۲] Boundary



شکل ۲.۳: برآورد تابع چگالی برای داده‌های فوران آبفشان پارک ملی یلوستون (سیلورمن، ۱۹۸۶) به ازای (الف) $h_{os} = ۰/۴۶۷$، (ب) $h_{os}/۲$، (ج) $h_{os}/۴$ و (د) $h_{os}/۵$

در مورد دو کلاس روش‌های انتخاب را می‌توان در موارد زیر خلاصه کرد: اعتبارسنجی متقابل (CV) منجر به هموارای کم[۱۳] می‌شود و برای مجموعه داده‌های بزرگ پایداری آن کم است (اغلب کوچکترین متغیر امکان‌پذیر از بین همه‌ی پهنای باندها انتخاب می‌شود) در حالی که جای‌گذاری وابسته به اطلاعات راهنما است و به‌طور نمونه برای نمونه‌های کوچک بد کار می‌کند. برای ایجاد تعدادی حکم در مورد نظریه مجانبی از مفروضات زیر روی کرنل‌ها و چگالی آنها استفاده می‌کنیم.

فرضیه‌ی $A_۱$: کرنل K دارای تابع چگالی با تکیه‌گاه فشرده بر روی $\mathbb{R}$ بوده و حول صفر متقارن و مشتق‌پذیر است (مشتق اول آن، K' وجود دارد)

فرضیه‌ی $A_۲$: $\mu_۲(K) < \infty$ که در آن $\mu_l(K) = \int u^l K(u)\mathrm{d}u$

[۱۳] Undersmoothing

فرضیه‌ی$A_۳$: چگالی f کران‌دار است تا مرتبه دوم مشتق‌پذیر و f' و f'' کران‌دار و انتگرال‌پذیر بوده و f'' بطور یکنواخت پیوسته است

برای تعدادی از روش‌ها ما شرایط بالا را اصلاح خواهیم کرد. در مطالعه شبیه‌سازی روش‌های انتخاب را محدود خواهیم کرد و از کرنل‌های مرتبه بالاتر[۱۴] استفاده نخواهیم کرد. در نظر داشته باشید که انگیزه اصلی برای استفاده از کرنل‌های مرتبه بالاتر برتری نظری در سرعت (نرخ) همگرایی مجانبی بالاتر است. هرچند اشکال مهم که قابلیت کاربردی آنها را از بین می‌برد وزن‌های منفی است که بنابراین برآورد چگالی منفی می‌شود (مارون، ۱۹۹۴ را ببینید). روش جای‌گذاری در مقایسه با CVهمواره دارای واریانس مجانبی کوچکتری است (هال و مارون، ۱۹۸۷a را ببینید) اما اغلب دارای اریبی بزرگتر است. با توجه به معلومات موجود بعید است که انتخاب‌گر پهنای باندی با خواص مجانبی بهتر از روش‌های جای‌گذاری وجود داشته باشد. اگرچه هال و جانستون (۱۹۹۲) توضیح دادند که چنین روش‌هایی از لحاظ فرض علمی وجود دارند، نتوانستند مثال عملی بیاورند.

قبل از ارائه روش‌های متفاوت جای‌گذاری، فرض کنید $\hat{h}_۰$ و $h_۰$ به ترتیب نشان‌دهنده‌ی پهنای باند hای باشند که به‌ترتیب $\mathrm{ISE}(h) = \mathrm{ISE}(\hat{f}_h(x))$ و $\mathrm{MISE}(h) = \mathrm{MISE}(\hat{f}_h(x))$ را کمینه می‌کنند.

روش جای‌گذاری پارک و مارون

روش دیگری که می‌توان برای جای‌گذاری عبارت مجهول $R(f'')$ در رابطه (۱۶.۲) در نظر گرفت، برآوردگر ناپارامتری زیر است.

$$\hat{f''}_g(x) = \frac{۱}{ng^۳}\sum_{i=۱}^{n} K''\left(\frac{x - X_i}{g}\right)$$

که در آن g پهنای باند پیشین است.

هال و مارون (۱۹۸۷) چندین برآورد برای $R(f'')$ پیشنهاد کرده‌اند که همه‌ی آن‌ها دوبار مجموع‌گیری روی نمونه است. در حقیقت آنها برآوردگری را پیشنهاد کردند که عامل اریبی را تا حد امکان کاهش می‌داد. برآوردگر آن ها دارای ساختار زیر است:

$$\widehat{R(f'')} = \|\hat{f''}_g\|_۲^۲ - \frac{۱}{ng^۵}\|K''\|_۲^۲ \qquad (۷.۳)$$

با جایگذاری (۷.۳) در (۱۶.۲) پهنای باند بهینه مجانبی پارک و مارون h_{PM} به‌صورت زیر حاصل می‌شود

$$\left(\frac{\|K\|_۲^۲}{\widehat{R(f'')}\mu_۲^۲(K)n}\right)^{۱/۵} \qquad (۸.۳)$$

سوالی که به ذهن می‌رسد این است که چگونه یک پهنای باند مناسب g را به‌دست آوریم. در پارک و مارون (۱۹۹۰)، g کمینه‌ی میانگین توان دوم خطای مجانبی $\widehat{R(f'')}$ است. طبق رابطه‌ی g را بر حسب $\hat{h}$ به‌صورت زیر به‌دست می‌آوریم.

$$g = C_۳(K)C_۴(f)\hat{h}^{\frac{۱۰}{۱۳}} \qquad (۹.۳)$$

[۱۴] Higher-order Kernel

که $C_۳(K)$ شامل مشتق چهارم و پیچش K و $C_۴(f)$ شامل مشتق دوم و سوم f است. (g, h_{PM}) بهینه با حل عددی دو فرمول (۸.۳) و (۹.۳) به‌دست می‌آید. نرخ همگرایی به ۰ h از مرتبه‌ی $O_P(n^{-\frac{۴}{۱۳}})$ است که بهینگی کمتری نسبت به $\sqrt{n}$ دارد (هال و مارون، ۱۹۹۱ را ببینید).

روش جای‌گذاری، شیتر و جونز[۱۵]

یکی از روش‌های پراستفاده، پهنای باند شیتر و جونز (۱۹۹۱) است. آنها از ایده‌ای شبیه پارک و مارون (۱۹۹۰) استفاده کردند با این تفاوت که برآوردگر $R(f'')$ را با نوع اریب آن جایگزین کردند و بنابراین مانع دست‌یابی به برآورد منفی $\widehat{R(f'')}$ در رابطه‌ی (۷.۳) شدند. آنها بیان کردند جمله‌ی غیرتصادفی در (۷.۳) منجر به نتیجه‌ی مثبت در مورد اریبی برآوردگر $\widehat{R(f'')}$ می‌شود. ایده‌ی آنها انتخاب برآوردگر پیشین (اولیه) g به‌گونه‌ای بود که اریبی منفی هموارسازی، تاثیر عبارت بااریبی را تعدیل کند. در نتیجه آنها $R(\hat{f}'')$ را با عبارت همیشه مثبت $R(\hat{f}''_g)$ برآورد کردند و پهنای باند پیشین g را با رابطه‌ی زیر برآورد کردند

$$g = C(K, L) \left(\frac{\|f''\|_۲^۲}{\|f'''\|_۲^۲} \right)^{\frac{۱}{۷}} h^{\frac{۵}{۷}}$$

که در آن $C(K, L)$ به L، کرنل استفاده شده در برآورد $R(f'')$ و هم‌چنین K کرنل لازم در برآورد اولیه وابسته است. سپس $R(f'')$ و $R(f''')$ با $R(f''_a)$ و $R(f''_b)$ که a و b پهنای باند قاعده‌ی سرانگشتی سیلورمن است، جای‌گذاری شوند. شیتر و جونز (۱۹۹۱) نشان دادند که پهنای باند پیشنهادی آنها نرخ همگرایی $n^{-\frac{۵}{۱۴}}$ را دارد که بهتر از روش پیشنهادی پارک و مارون (۱۹۹۰) رفتار می‌کند. در یک مثال از داده‌های واقعی، جونز و همکاران (۱۹۹۱b) متوجه شدند که برآوردگر پیشنهادی آنها رفتاری شبیه به پهنای باند پارک و مارون دارد. بنابراین، بدون اینکه خللی در کارآیی وارد شود، اندکی نرخ همگرایی بهتر خواهد شد اما از لحاظ محاسباتی هزینه‌ی بالاتری خواهد داشت. $\hat{h}_{SJ}$ پهنای باند (ساده‌تری) از پهنای باند جونز و همکاران (۱۹۹۱) است.

تکمیل روش‌جای‌گذاری

برای نمونه‌های کوچک و پهنای باند بهینه کوچک، برآوردگر $\widehat{R(f'')}$ در رابطه‌ی (۷.۳) به آسانی رد می‌شود همچنین پیدا کردن حل عددی برای (g, h_{PM}) ممکن است از لحاظ کاربردی واقعاً دشوار باشد. نتیجه نهایی به قاعده‌ی توقف بستگی دارد و ممکن است چندین ماکزیمم موضعی برای مسئله نمونه‌های متناهی دو بعدی وجود داشته باشد. برای دوری از این مشکلات و ارائه‌ی پاسخ سریع و آسان پیشنهاد می‌شود که در ابتدا از پهنای باند قاعده‌ی سرانگشتی سیلورمن برای کرنل‌های نرمال استفاده شود. کرنل توان چهارم (دو وزنی) در برآورد مشتق دوم تا حدی شبیه کرنل اپاچنیکوف است. در این راستا رابطه‌ی زیر به عنوان یک g پیشین برای رابطه‌ی (۷.۳)پیشنهاد می‌گردد

$$g = h_{rot} \frac{۲/۰۳۶۲}{۰/۷۷۶۴} n^{\frac{۱}{۵}-\frac{۱}{۹}}$$

که در آن h_{rot}، در رابطه‌ی (۶.۳) تعریف شده است.

[۱۵] Jones and Sheather

این پهنای باند منجر به برآورد منطقی مشتق دوم f شده و بنابراین $R(f'')$ محاسبه می‌شود. برتری دیگر آن، این است که g پیشین نسبتا سریع‌تر به‌دست آید.

۲.۲.۳ روش‌های اعتبارسنجی متقابل در برآورد چگالی

با توجه به فرمول (۹.۲) ظاهراً اولین جمله می‌تواند بر اساس تابع برآورد شده محاسبه شود، و دومین جمله، امید ریاضی متغیر $\hat{f}_h(X)$ است و سومین جمله، می‌تواند نادیده گرفته شود زیرا به پهنای باند بستگی ندارد. توجه کنید که برآورد $E\hat{f}_h(X)$ با $\frac{1}{n}\sum_{i=1}^{n}\hat{f}_h(X_i)$ به دلیل وابستگی ضمنی $\hat{f}_h$ ($\hat{f}_h$ به X_i وابسته است) مناسب نیست. بنابراین برآوردگرهای مختلف اعتبارسنجی متقابل (CV) به‌طور اساسی در برآورد دومین قسمت پیشنهاد شده است.

۳.۲.۳ اعتبار سنجی متقابل کمترین توان دوم (LSCV[۱۶])

روش کمترین توانهای دوم اعتبارسنجی متقابل یک روش کاملاً خودکار در انتخاب پارامتر هموار سازی است. در این روش، ایده‌ی انتخاب پهنای باند با هدف کمینه کردن ISE معادل با انتخابی از h است که کمیت $R(\hat{f})$ را که در زیر تعریف شده است را کمینه کند:

$$R\left(\hat{f}\right) = \int \hat{f}^{2}(x)\mathrm{d}x - 2\int \hat{f}(x)f(x)\mathrm{d}x$$

ایده‌ی اصلی این روش ساختن برآوردی برای $R\left(\hat{f}\right)$ با استفاده از خود داده‌ها و سپس کمینه کردن آن بر حسب h است. رهیافت کلاسیک بدین صورت است که هنگامی‌که $f(X_i)$ را برآورد می‌کنید فقط X_i را از نمونه حذف کنید و آن را برآوردگر جک‌نایف[۱۷] بنامید و آن را با نماد $\hat{f}_{-i}(X_i)$ نمایش دهید؛ به‌عبارت دیگر،

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

حال قرار می‌دهیم:

$$\mathrm{CV}(h) = \int \hat{f}^{2}(x)\mathrm{d}x - \frac{2}{n}\sum_{i}\hat{f}_{-i}(X_i)$$

مقدار CV تنها به داده‌ها بستگی دارد. روش کم‌ترین توان‌های دوم، CV را نسبت به h کمینه می‌کند. امید ریاضی CV به صورت زیر محاسبه می‌شود:

$$E(\mathrm{CV}(h)) = E\left[\int \hat{f}^{2}(x)\mathrm{d}x\right] - 2E\left[\frac{1}{n}\sum_{i}\hat{f}_{-i}(X_i)\right] \qquad (۱۰.۳)$$

باتوجه به هم‌توزیع بودن X_iها می‌توان نوشت:

$$E\left[\frac{1}{n}\sum_{i}\hat{f}_{-i}(X_i)\right] = \frac{1}{n}\sum_{i}E\left[\hat{f}_{-i}(X_i)\right]$$

[۱۶] Least squares cross-validation
[۱۷] Jack-knife estimator

$$\begin{aligned} &= \frac{1}{n}\left(nE\left[\hat{f}_{-n}(X_i)\right]\right) \\ &= E\left[\hat{f}_{-n}(X_n)\right]\left(=E\left[\hat{f}_{-1}(X_1)\right]=\cdots=E\left[\hat{f}_{-(n-1)}(X_{n-1})\right]\right) \\ &= \int \hat{f}_{-n}(x)f(x)\mathrm{d}x \end{aligned}$$

از آن‌جایی‌که بر اساس خاصیت امید ریاضی دوگانه می‌توان نوشت

$$\begin{aligned} E(X) &= EE(X|Y) \\ &= E\int xf(x|y)\mathrm{d}x \\ E\hat{f}_{-n}(X_n) &= E\left[E\hat{f}_{-n}(X_n)|\hat{f}_{-n}(x)\right] \\ &= E\int \hat{f}_{-n}(x)f(x)\mathrm{d}x \end{aligned}$$

در نتیجه داریم

$$\begin{aligned} E(CV(h)) &= E\left[\int \hat{f}^2(x)\mathrm{d}x\right] - 2E\left[\int \hat{f}(x)f(x)\mathrm{d}x\right] \\ &= E\left(R\left(\hat{f}\right)\right) \end{aligned} \qquad (۱۱.۳)$$

حال با جایگذاری در $MISE$ خواهیم داشت:

$$\begin{aligned} MISE &= E\left(R\left(\hat{f}\right)\right) + \int f^2(x)\mathrm{d}x \\ &= E\left(CV(h)\right) + \int f^2(x)\mathrm{d}x \end{aligned} \qquad (۱۲.۳)$$

لذا $CV(h)+\int f^2(x)\mathrm{d}x$ برآوردگر ناریب برای میانگین انتگرال توان دوم خطا است، چون جمله‌ی $\int f^2(x)\mathrm{d}x$ بستگی به h ندارد در نتیجه مینیمم کردن $E\left[CV_h(h)\right]$ بر حسب h معادل با مینیمم کردن میانگین انتگرال توان دوم خطا است و این نشان می‌دهد که چرا انتظار می‌رود مینیمم کردن CV_h انتخاب خوبی را برای پارامتر هموار سازی ارائه دهد. در ادامه فرض کنید $\hat{h}_{CV}$ کمینه‌کننده‌ی LSCV باشد که از رابطه‌ی زیر به‌دست می‌آید

$$\hat{h}_{CV} = \mathrm{argmin} CV(h) \qquad (۱۳.۳)$$

که در آن

$$\min_h CV(h) = \min_h \left(\int \hat{f}_h^2(x)\mathrm{d}x - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)\right)$$

استون[۱۸] (۱۹۸۴) نشان داد تحت فرضیه‌های A1 تا A3:

$$\frac{\mathrm{ISE}\hat{h}_{CV}}{\mathrm{ISE}\hat{h}_0} \xrightarrow{a.s.} 1$$

[۱۸] Stone

با این وجود، هال و مارون (۱۹۸۷) تحت فرضیه‌های A_1 تا A_2 و با مقادیر σ و c که فقط به f و K وابسته‌اند، توزیع مجانبی $\hat{h}_{CV}$ را به‌صورت زیر ارائه کردند:

$$\begin{aligned} n^{\frac{2}{10}}\left(\hat{h}_{CV}-\hat{h}_{.}\right) &\to N(0,\sigma^2 c^{-2}) \\ n\left(\mathrm{ISE}(\hat{h}_{CV})-\mathrm{ISE}(\hat{h}_{.})\right) &\to \frac{1}{2}\sigma^2 c^{-1}\chi_1^2 \end{aligned} \qquad (۱۴.۳)$$

افراد بسیاری به خاطر تعریف بدیهی و کاربرد ساده، از روش اعتبارسنجی متقابل کمترین توان‌های دوم استفاده می‌کنند. انتقادی که به این روش وارد است، فقدان ثبات و پایداری است، حتی در زمانی که اندازه‌ی نمونه زیاد باشد. سیلورمن (۱۹۸۶) توضیحاتی مبنی بر اینکه اگر فاصله $|x_i - x_j|$ برای بسیاری از مشاهدات $j \neq i$ خیلی کوچک باشد، چه اتفاقی ممکن است رخ دهد، ارائه کرده‌اند. چیو (۱۹۹۱a) رخ دادن این مسئله را با داده‌های گرد شده مورد بررسی قرار داد. در این حالت، ممکن است گره‌های زیادی به‌دست آورده شود. $(\forall i \neq j, \quad x_j \neq x_i)$.

اعتبارسنجی متقابل اریب

هدف روش اعتبارسنجی متقابل اریب[۱۹] (BCV) جایگزین کردن $R(f'')$ با برآوردگر آن، $\hat{R}(f'')$ در فرمول AMISE مجانبی، رابطه‌ی (۱۴.۲) می‌باشد که به‌صورت زیر به‌دست می‌آید

$$\begin{aligned} \hat{R}(f'') &= R(\hat{f}'') - \frac{1}{nh^5}R(K'') \\ &= \frac{1}{n^2}\sum\sum_{i\neq j}(K_h'' * K_h'')(X_i - X_j) \end{aligned} \qquad (۱۵.۳)$$

که در آن $K''(.)$ مشتق دوم K بوده و $*$ نماد پیچش است.

برای محاسبه‌ی

$$BCV(h) = \frac{R(K)}{nh} + \frac{1}{4}h^4(\sigma_K^2)^2\hat{R}(f'')$$

در رابطه‌ی (۱۶.۲)، یک برآورد f'' می‌تواند به‌صورت زیر به‌دست آید.

$$\hat{f}_g''(x) = \frac{1}{ng^3}\sum_{i=1}^{n} K''\left(\frac{x - X_i}{g}\right)$$

که در آن g پهنای باند پیشین (اولیه) است و باید به‌طور مطلوب به‌دست آید.

تثبیت انتخاب پهنای باند

بر اساس تابع مشخصه، چیو (۱۹۹۲) و (۱۹۹۱a,b) برآوردی برای h بر اساس معیار CV ارائه کرد که منشا تغییرات را نشان می‌داد. چیو نشان داد که معیار $\mathrm{CV}(h)$ تقریبا برابر است با:

$$\frac{1}{\pi}\int_0^{\infty} |\tilde{\phi}(\lambda)|^2\left\{\omega^2(h\lambda) - 2\omega(h\lambda)\right\}\mathrm{d}\lambda + \frac{2}{nh}K(0) \qquad (۱۶.۳)$$

[۱۹] Biased Cross Validation

که در آن $\tilde{\phi}(\lambda) = \frac{1}{n}\sum_{j=1}^{n} e^{i\lambda x_j}$ و $\omega(\lambda) = \int e^{i\lambda u}K(u)\mathrm{d}u$.

سهم اختلال در برآورد CV، اساسا از طریق $|\tilde{\phi}(\lambda)|$ مشخص می‌شود که شامل اطلاعات زیادی در مورد f نیست و تفاوت معیارهای CV و MISE می‌تواند برای کاهش این مشکل به‌کار گرفته شود. به عنوان یک جایگزین، او مقدار Λ را به‌صورت اولین λیی در نظر گرفت که $|\tilde{\phi}(\lambda)|^2 \leq 3/n$ و $|\tilde{\phi}(\lambda)|^2$ را با $\frac{1}{n}$ برای $\lambda > \Lambda$ تعریف کرد. در نتیجه برآوردگر زیر را ارائه کردند:

$$\begin{aligned}\hat{h}_{ST} = \arg\min_h S_n(h) &= \int_0^{\Lambda} |\tilde{\phi}(\lambda)|^2 \left\{\omega^2(h\lambda) - 2\omega(h\lambda)\right\} \mathrm{d}\lambda \\ &+ \frac{1}{n}\int_{\Lambda}^{\infty} \left\{\omega^2(h\lambda) - 2\omega(h\lambda)\right\}\mathrm{d}\lambda + 2\pi K(0)/(nh) \\ &= \frac{\pi}{nh}\|K\|_2^2 + \int_0^{\Lambda} \left\{|\tilde{\phi}(\lambda)|^2 - \frac{1}{n}\right\}\left\{\omega^2(h\lambda) - 2\omega(h\lambda)\right\}\mathrm{d}\lambda\end{aligned}$$

برای کمینه کننده‌ی $\hat{h}_{ST}$، چیو نشان داد که $\hat{h}_{ST} \xrightarrow{a.s.} \hat{h}_0$ و حتی با نرخ بهینگی $\sqrt{n}$ همگرا به h_0 است. زمانی‌که در شبیه‌سازی می‌خواهیم Λ را محاسبه کنیم با توان دوم ریشه‌های منفی مواجه می‌شویم. برای اجتناب از این مسئله، قدر مطلق عدد زیر رادیکال را محاسبه می‌کنیم.

اعتبارسنجی متقابل اصلاح‌شده

استوت[20] (۱۹۹۲) پیشنهاد کرد به جای کمینه کردن CV، اعتبارسنجی متقابل اصلاح‌شده[21](MCV) را برای تعیین پهنای مناسب h کمینه کنیم که در آن

$$\begin{aligned}\hat{h}_{MCV} &= \min_h MCV(h) \\ &= \int \hat{f}_h^2(x)\mathrm{d}x - S - \frac{\mu_2(K)}{2n(n-1)h}\sum_{i\neq j} K''\left(\frac{X_i - X_j}{h}\right). \qquad (۱۷.۳)\end{aligned}$$

که در آن

$$S = \frac{1}{n(n-1)h}\sum_{i\neq j} K\left(\frac{X_i - X_j}{h}\right)$$

تحت فرضیه‌های **A1** و فرضیه‌ی

A2′ K سه دفعه مشتق‌پذیر و پیوسته است، $\int t^4|K(t)|\mathrm{d}t < \infty$، $\int t^4|K''(t)|\mathrm{d}t < \infty$، داریم $\int t^2[K''(t)]^2\mathrm{d}t < \infty$ و $\int t^4[K'(t)]^2\mathrm{d}t < \infty$.

A3′ f چهار بار مشتق‌پذیر است، مشتق‌های آن کران‌دار و انتگرال‌پذیراست

استوت (۱۹۹۲) نشان داد در حالت مجانبی $n \longrightarrow \infty$

$$\frac{\hat{h}_0}{\hat{h}_{MCV}} \xrightarrow{P} 1 \quad \text{و} \quad \frac{ISE(\hat{h}_0)}{ISE(\hat{h}_{MCV})} \xrightarrow{P} 1,$$

[20] Stute
[21] Modified Cross Validation

پیشرفت در روش‌های اعتبارسنجی متقابل

فلاچ و کوروناکی[۲۲] (۱۹۹۲) پیشنهاد کردند به جای حذف فقط مشاهده‌ی X_i در برآورد $f(X_i)$ ترجیحا mتا X_i $(m < n)$ از نزدیکترین X_iها حذف شود. اگر $m \to \infty$، آن‌گاه $\frac{m}{n} \to 0$.

این فکر شبیه انتخاب CV برای یک دنباله از داده‌های سری زمانی (هاردل و ویو[۲۳]، ۱۹۹۲ را ببینید) است و مشابه استوت (۱۹۹۲)، آنها این روش را اعتبارسنجی متقابل اصلاح‌شده نامیدند. متاسفانه کارآیی این روش بستگی به انتخاب m دارد. از آن‌جایی که قادر نیستیم پیشنهاد مناسبی در مورد مقدار m ارائه دهیم، این روش در گروه روش‌های خودکار یا وابسته به داده‌ها قرار نمی‌گیرد.

اسکات و ترل[۲۴] (۱۹۸۷) روش اعتبارسنجی متقابل اریب (BCV) را معرفی کردند. آنها نگران نامعتبر (تغییرپذیری زیاد CV) بودن نتایج آن برای نمونه‌های کوچک بودند و مستقیما بر کمینه کردن MISE مجانبی تمرکز کردند. عبارت ناشناخته‌ی $R(f'')$ با روش‌های جک‌نایف تخمین زده شد. کارآیی کم این روش برای نمونه‌های کوچک و نیز هنگامی‌که توزیع واقعی جامعه، آمیخته‌ای از توزیع‌ها باشد، توسط نویسندگان آن پذیرفته شده است. چیو (۱۹۹۶) و جونز و همکاران (۱۹۹۶) در مطالعات شبیه‌سازی بر نامناسب بودن کارآیی این روش تاکید کرده‌اند. اعتبارسنجی متقابل هموارشده[۲۵] (SCV) توسط هال و همکاران در سال ۱۹۹۲ معرفی شد. ایده‌ی اصلی، یک نوع بیش‌هموارسازی قبل از به‌کار بردن معیار CV بود. فرآیند بیش‌هموارسازی منجر به تغییرپذیری در نمونه‌های کوچکتر می‌شود اما اریبی را افزایش می‌دهد. بنابراین اغلب پهنای باند به‌دست آمده، بیش‌هموار است و برخی از ویژگی‌های مهم چگالی را نادیده می‌گیرد. بنابراین به نظر می‌رسد که این روش تنها برای نمونه‌های بزرگ مناسب است. جونز و همکاران (۱۹۹۶b) نشان دادند که بدون کرنل مرتبه بالاتر نرخ همگرایی SCV، $n^{-۱/۱۰}$ است که برابر با نرخ همگرایی روش خودگردان تیلور[۲۶] (۱۹۸۹) و نسبتا در ارتباط با روش خودگردان کائو[۲۷] (۱۹۹۳) است این روش‌ها به روش‌های CV تعلق ندارند. علاوه بر این، با یک انتخاب خاصی از پهنای باند مقدماتی[۲۸]. نتایج SCV منجر به یک نرخ همگرایی $n^{-\frac{۵}{۱۴}}$ می‌شود.

اعتبارسنجی متقابل افرازشده[۲۹] (PCV) توسط مارون (۱۹۸۸) پیشنهاد شد. او معیار CV را با تقسیم نمونه‌ای به اندازه‌ی n به m زیرمجموعه از نمونه اصلاح کرد. PCV با کمینه کردن میانگین تابع CV برای تمام زیرنمونه‌ها محاسبه می‌شود. تعدادی از زیرنمونه‌ها با موازنه واریانس و اریبی انتخاب می‌شوند. بنابراین، انتخاب یک نمونه‌ی مقدماتی در این حالت مسئله‌ی دشواری است و همان‌طور که پارک و مارون (۱۹۹۰) اشاره کردند این روش، یک روش کاملا عینی[۳۰] (شدنی) نیست. برای به‌دست آوردن افرازهای منطقی، الزاما اندازه‌ی نمونه باید بزرگ باشد.

---

[۲۲] Feluch and Koronacki
[۲۳] Hardel and Vieu
[۲۴] Scott and Terrell
[۲۵] Smoothed Cross Validation
[۲۶] Taylor
[۲۷] Cao
[۲۸] Pilot bandwidth
[۲۹] Partitioned cross validation
[۳۰] Objective

اعتبارسنجی متقابل شبه‌درستنمایی[۳۱] (گاهی اوقات کولبک-لایبلر[۳۲] نامیده می‌شود) توسط هاباما[۳۳] و همکاران (۱۹۷۴) و نیز دیون[۳۴] (۱۹۷۶) معرفی شد. هدف این روش، پیدا کردن پهنای باندی است که معیار شبه‌درستنمایی را با خارج کردن X_i بیشینه کند. با توجه به انتقاد زیادی که در مورد نامناسب بودن این روش برای برآورد چگالی مطرح شده است، معمولا از این روش استفاده نمی‌شود.

ویکمپ[۳۵] (۱۹۹۹) روشی که وابستگی زیاد به تکنیک CV دارد را پیشنهاد کرد. این روش، بر اساس مفهوم بهینگی دوروی و لوگوسی[۳۶] (۱۹۹۶) به وجود آمدند اما در مطالعاتی که در مورد نرم $L_۲$ بودند قرار گرفتند. در راستای سایر مسائل، این فرآیند به تفکیک‌سازی نمونه احتیاج دارد که می‌تواند کاملا برای نمونه‌های با اندازه‌های کوچک و متوسط مشکل‌ساز باشد.

۴.۲.۳ روش ترکیب در روش‌های انتخاب پهنای باند

می‌دانیم که انتقادات زیادی به روش CV به دلیل همواری کم و عدم ثبات برای نمونه‌های بزرگ وجود دارد و در همان زمان روش جای‌گذاری دارای ثبات بیشتری است اما باعث بیش‌همواری می‌شود. به این دلیل پیشنهاد ترکیب پهنای باندهای متفاوت یا برآوردهای چگالی مطرح می‌شود.

می‌توان با استفاده از ترکیب پارامترهای هموارسازی برآوردگرهای چگالی مختلف، پهناهای باند ترکیبی دیگری پیدا کرد. از آن جمله می‌توان به مطالعات ریگولت و سیباکوف[۳۷] (۲۰۰۷)، ساماروف و سیباکوف[۳۸] (۲۰۰۷) و یانگ[۳۹] (۲۰۰۰) اشاره کرد. با در نظر گرفتن این حقیقت که برآوردگرهای پهناهای باند بر اساس روش‌های اعتبارسنجی متقابل و جای‌گذاری به ترتیب کم‌هموار و بیش‌هموار هستند و هم‌چنین آن پهناهای باندی که پارامترهای مقیاسی هستند باید به‌صورت لگاریتمی (ضربی) ترکیب شوند، می‌توان شکل کلی زیر را در نظر گرفت.

$$\left(\hat{h}_{CV}^{\alpha}\hat{h}_{PM}^{\beta}\right)^{\frac{۱}{\alpha+\beta}} \qquad (۱۸.۳)$$

لازم به ذکر است که در شبیه‌سازی‌ها مدل‌های زیر در نظر گرفته شده‌اند.

mix 1 $\alpha = ۱$ و $\beta = ۲$

mix 2 $\alpha = ۲$ و $\beta = ۱$

mix 3 $\alpha = ۱$ و $\beta = ۱$

[۳۱] Pseudo-likelihood cross validation
[۳۲] Kullback-Leible
[۳۳] Habama
[۳۴] Duin
[۳۵] Wegkamp
[۳۶] Devroye and Lugosi
[۳۷] Rigollet and Tsybakov
[۳۸] Samarov and Tsybakov
[۳۹] Yang

۵.۲.۳ روش‌های خودگردان

در این قسمت علاوه بر روش‌های ارائه شده در بخش ۳.۲، روش کاربردی دیگری را ارائه می‌کنیم که در شبیه‌سازی‌ها از آن استفاده می‌شود.

اساس روش‌های انتخاب بر مبنای روش خودگردان، انتخاب پهنای باند در راستای برآورد خودگردان ISE یا MISE است. برای توصیف کلی این ایده‌ی بازنمونه‌گیری، مطالعات ناپارامتری هال (۱۹۹۰) را ببینید. فرض کنید که برآورد پارزن–رزنبلات $\hat{f}_g$ به ازای پهنای باند پیشین g فرض شده در اختیار داریم. از $\hat{f}_g$ می‌توان نمونه‌های خودگردان $(X_1^*, X_2^*, \ldots, X_n^*)$ را تولید کرد. چگالی کرنل نمونه‌های خودگردان به‌صورت زیر خواهند بود.

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

ISE و MISE به‌ترتیب می‌تواند با عبارات زیر تقریب زده شوند.

$$\begin{aligned} \mathrm{ISE}^*(h) &:= \int \left(\hat{f}_{h^*}(x) - \hat{f}_{g^*}(x)\right)^2 \mathrm{d}x \\ \mathrm{MISE}^*(h) &:= E_* \left[\int \left(\hat{f}_{h^*}(x) - \hat{f}_{g^*}(x)\right)^2 \mathrm{d}x\right] \end{aligned} \qquad (۱۹.۳)$$

که در آن E_* و MISE^* فقط به نمونه اصلی بستگی دارند و به نمونه‌های خودگردان وابسته نیستند. در نتیجه، نیازی به انجام نمونه‌گیری دیگری برای به‌دست آوردن MISE^* نیست. بر اساس قضیه فوبینی[۴۰] و تجزیه‌ی $MISE^*(h) = V^*(h) + SB^*(h)$ می‌توان MISE^* را تقریب زد. در این راستا عبارت واریانس $V^*(h)$ برابر است با

$$V^*(h) = \frac{1}{nh}.\|K\|_2^2 + \frac{1}{n}.\int \left(\int K(u).\hat{f}_g(x - hu)\mathrm{d}u\right)^2 \mathrm{d}x \qquad (۲۰.۳)$$

همچنین عبارت توان دوم اریبی SB به‌صورت زیر است

$$SB^*(h) = \int \left(\int K(u).(\hat{f}_g(x - hu) - \hat{f}_g(x))\mathrm{d}u\right)^2 \mathrm{d}x \qquad (۲۱.۳)$$

می‌توان $V^*(h)$ و $SB^*(h)$ را بر حسب پیچش (نماد $*$) به‌صورت زیر بازنویسی کرد.

$$\begin{aligned} V^*(h) &= \frac{1}{nh}\|K\|_2^2 + \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(K_h * K_g) * (K_h * K_g)\right](X_i - X_j) \\ SB^*(h) &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(K_h * K_g - K_g) * (K_h * K_g - K_g)\right](X_i - X_j). \end{aligned}$$

در کاربرد به‌دست آوردن فرمول صریحی برای انتگرال، وقتی که کرنل دارای تکیه‌گاه کراندار است دشوار است. با این وجود، استفاده از کرنل گوسی به ما کمک می‌کند مستقیماً پهنای باند مطلوب را با کمینه کردن $\mathrm{MISE}^*(h)$ به‌صورت زیر حساب کنیم

$$MISE^*(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{\sqrt{2\pi}}\left[\frac{\sum_{i,j}\left(\exp\left(-\frac{1}{2}\left(\frac{X_i - X_j}{g\sqrt{2}}\right)^2\right)\right)\sqrt{2g^2}.n^2}{\sqrt{2g^2}n^2}\right.$$

[۴۰] Fubini

$$- \frac{۲/\sum_{i,j}\left(\exp\left(-\frac{۱}{۲}\left(\frac{X_i-X_j}{\sqrt{h^۲+۲g^۲}}\right)^۲\right)\right)}{\sqrt{h^۲+۲g^۲}.n^۲} + \frac{(n+۱)\sum_{i,j}\left(\exp\left(-\frac{۱}{۲}\left(\frac{X_i-X_j}{\sqrt{۲(h^۲+g^۲)}}\right)^۲\right)\right)}{\sqrt{۲(h^۲+g^۲)}.n^۳}\Bigg] \quad (۲۲.۳)$$

پهنای باند معادل برای هر کرنل دیگر را می‌توان همان‌طور که در مارون و نلان[۴۱] (۱۹۸۸) یا هاردل[۴۲] و همکاران (۲۰۰۴) توضیح داده شده است، به‌دست آورد.

روش خودگردان در برآورد چگالی کرنل اولین بار توسط تیلور[۴۳] (۱۹۸۹) ارائه شد، اگر چه نسخه‌های اصلاح شده زیادی بعداً منتشر شد (فاراوی و یوهان،[۴۴] ۱۹۹۰، هال، ۱۹۹۰ و کائو، ۱۹۹۳ را ببینید). تفاوت اصلی انواع مختلف این برآوردگر در چگونگی انتخاب پهنای باند پیشین g و تولید نمونه‌های خودگردان است.

تیلور (۱۹۸۹) پیشنهاد کرد که $g = h$ را در نظر گرفته و از کرنل گوسی استفاده کنیم. مارون (۱۹۹۲) نشان داد که این انتخاب نامناسب است و اریبی مثبتی را به وجود می‌آورد. فاراوی و یوهان (۱۹۹۰) برای پیدا کردن g، برآورد LSCV را پیشنهاد کردند. هال (۱۹۹۰) پیشنهاد کرد استفاده از توزیع تجربی برای تولید نمونه‌های خودگردان با اندازه‌ی m $(m < n)$ مناسب است. وی پیشنهاد کرد $m \simeq n^{۱/۲}$ و $h = g(m/n)^{۱/۵}$ و MISE را نسبت به g کمینه کرد. کائو و همکاران (۱۹۹۴) نشان دادند که روش خودگردان پیشنهادی هال، کاملا ناپایدار است. همچنین آنها متوجه شدند روش‌های فاراوی و یوهان (۱۹۹۰) و هال (۱۹۹۰) با روش کائو (۱۹۹۳) بهبود می‌یابد.

اریبی برآورد خودگردان تصحیح شده توسط گروند و پولزهل[۴۵] (۱۹۹۷) توسعه یافت. آنها یک برآورد پهنای باند با نرخ همگرایی $\sqrt{n}$ به‌دست آوردند که نتایج خوبی برای حجم نمونه‌های بزرگتر ارائه می‌دهد اما در مورد نمونه‌های کوچک و متوسط نتایج خوبی نداشت. علاوه بر این، به منظور دستیابی به تئوری مجانبی آنها از مفروضات زیاد و قویتری در مقایسه با سایر روش‌ها باید استفاده می‌کردند.

در روش خودگردان هموارشده کائو (۱۹۹۳) پهنای باند پیشین g کمینه‌کننده‌ی عبارت مجانبی میانگین توان دوم خطا است. برای جزئیات بیشتر به کائو (۱۹۹۳) مراجعه کنید. وی دریافت که در رابطه‌ی (۲۱.۳) به‌ازای $i = j$ اریبی زیاد می‌شود و بنابراین، پیشنهاد کرد از انتگرال توان دوم اریبی خودگردان اصلاح شده استفاده شود

$$MB^*(h) = \frac{۱}{n^۲}\sum_{i\neq j}\big[(K_h * K_g - K_g) * (K_h * K_g - K_g)\big](X_i - X_j).$$

[۴۱] Marron and Nolan
[۴۲] Hardle
[۴۳] Taylor
[۴۴] Faraway and Juun
[۴۵] Grund and Polzehl

در ارتباط با نرخ همگرایی، او نشان داد کمینه‌کننده $MB^*(h)$ که با h_o^* نشان داده می‌شود، نتایج زیر برقرار است.

$$\begin{aligned}\frac{\mathrm{MISE}(h_o^*)-\mathrm{MISE}(h_o)}{\mathrm{MISE}(h_o)} &= O_p(n^{-\frac{۵}{۷}})\\ \frac{\mathrm{MISE}(h_{oM}^*)-\mathrm{MISE}(h_o)}{\mathrm{MISE}(h_o)} &= O_p(n^{-\frac{۳}{۱۳}}).\end{aligned} \qquad (۲۳.۳)$$

توجه کنید که نرخ همگرایی برای نسخه‌ی خودگردان اصلی اندکی سریع‌تر است.

اخیراً چاکن و همکاران[۴۶] (۲۰۰۸) برآوردگر خودگردانی کاملاً شبیه کائو (۱۹۹۳) منتشر کردند.

در مجموع، در مطالعات شبیه‌سازی، کلاس‌های برآوردهای خودگردان کائو (۱۹۹۳) را در نظر می‌گیریم. او ثابت کرد که پهنای باند پیشین g، رابطه‌ی (۱۹.۳) را کمینه می‌کند و با کمینه کردن قسمت میانگین توان دوم خطا تعیین می‌شود. لذا او نشان داد g با رابطه‌ی زیر به‌دست می‌آید.

$$g = \left(\frac{\|K\|_۲^۲}{R(\hat{f}''')\mu_۲^۲(K)n}\right)^{۱/۷}. \qquad (۲۴.۳)$$

این فرمول برای پهنای باند پیشین g در (۲۲.۳) به‌کار می‌رود.

۳.۳ شبیه‌سازی

در این بخش، کارآیی روش‌های مختلف انتخاب پهناهای باند را با هم مقایسه می‌کنیم. در اینجا روش‌های سازگار با داده‌ها را بدون تصحیح مرزها در نظر می‌گیریم. شش مثال مختلف را برای شبیه‌سازی داده‌ها در نظر می‌گیریم.

مثال ۱: توزیع لاپلاس با تابع چگالی احتمال $f(x)=۴\exp\left(-|۸(x-۰/۵)|\right)$.

مثال ۲: توزیع گامای ساده: $\Gamma(۱,b)$ با پارامتر مقیاس $b=۱/۵^۲$ و شکل $a=b^۲$ که بر روی $x\in\mathbb{R}$، $۵x$ تعریف شده است. و تابع چگالی احتمال آن به‌صورت $f(x)=۵\frac{b^a}{\Gamma(a)}(۵x)^{a-۱}\exp\{-۵xb\}$ است.

مثال ۳: آمیخته‌ای از سه توزیع گاما، $\Gamma(a_j,b_j)$ که در آن $a_j=b_j^۲$ و $b_۱=۱/۵$، $b_۲=۳$ و $b_۳=۶$ که بر روی $۸x$ به‌کار برده شده است که دو برآمدگی ایجاد می‌کند.

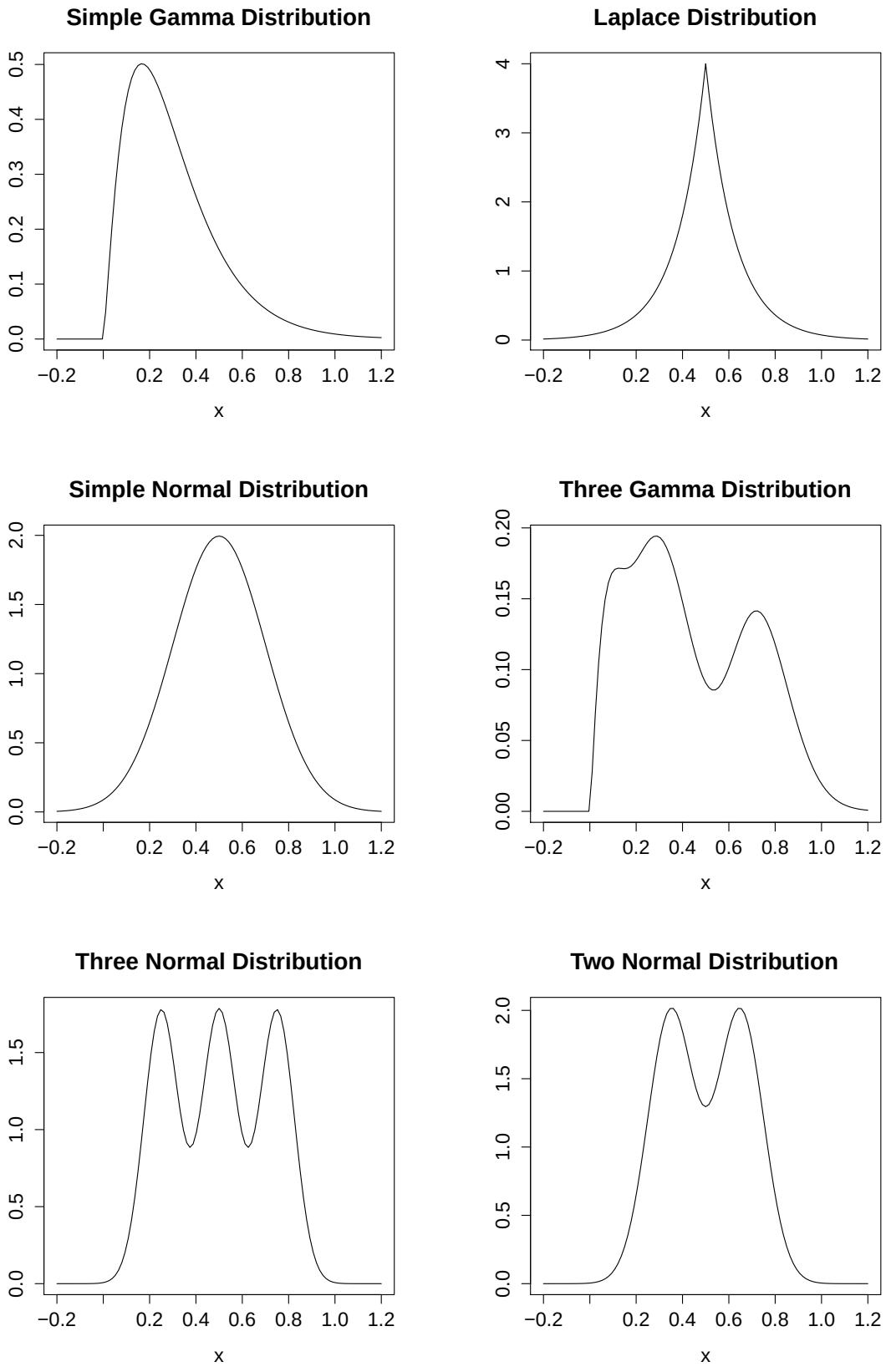
مثال ۴: توزیع نرمال ساده، $N(۰/۵,۰/۲^۲)$ که فقط یک مد دارد.

مثال ۵: آمیخته‌ای از توزیع‌های $N(۰/۳۵,۰/۱^۲)$ و $N(۰/۶۵,۰/۱^۲)$ که دو مد دارد.

مثال ۶: یک توزیع سه‌مدی، آمیخته‌ای از سه توزیع $N(۰/۲۵,۰/۰۷۵^۲)$، $N(۰/۵,۰/۰۷۵^۲)$ و $N(۰/۷۵,۰/۰۷۵^۲)$.

همان‌طور که از نمودار ۳.۳ مشخص است، همه‌ی چگالی‌ها، جرم اصلی خود را در بازه‌ی $[۰,۱]$ می‌پذیرند و با نرخ نمایی به سمت دم‌ها کاهش می‌یابند. بنابراین، می‌توانیم اثر مرزی ممکن را نادیده بگیریم.

[۴۶] Chacon et al.



شکل ۳.۳: نمودار چگالی‌هایی که برای تولید داده از آنها استفاده شده است. (مثال‌های ۱ تا ۶)

کارآیی روش‌های مختلف بر اساس پنج معیار متفاوت از ISE تابع برآورد کرنل به‌ازای hهای گوناگون به شرح زیر محاسبه شده است:

c_1 میانگین $(\hat{h} - \hat{h}_\cdot)$ ، که نشان‌دهنده‌ی اریبی پهنای باند است.

c_2 میانگین $\mathrm{ISE}(\hat{h})$.

c_3 انحراف‌معیار $\mathrm{ISE}(\hat{h})$ که پراکندگی ISE را نشان می‌دهد.

c_4 میانگین $\left(\left[\mathrm{ISE}(\hat{h}) - \mathrm{ISE}(h_{opt})\right]^2\right)$، که معیاری بر اساس نرم L_2 است.

c_5 میانگین $\left(\left|\mathrm{ISE}(\hat{h}) - \mathrm{ISE}(h_{opt})\right|\right)$، که معیاری بر اساس نرم L_1 است.

که در آن‌ها $\hat{h}_\cdot$، همان پهنای باندی است که از کمینه کردن MISE در هر مرحله‌ی نمونه‌گیری به‌دست آمده است.

می‌توان به جای بررسی معیارهای (c_2, c_3)، توزیع ISE را برای مثال با کمک نمودار جعبه‌ای مطالعه کرد.

در ادامه، روش‌های مختلف پهناهای باند که در شبیه‌سازی استفاده شده است، معرفی شده است.

h1 پهنای باندی که از کمینه کردن MISE به‌دست آمده است. ($\hat{h}_\cdot$، رابطه‌ی ۱۶.۲)

h2 پهنای باندی که از کمینه کردن رابطه‌ی (۶.۳) به‌دست می‌آید که همان قاعده‌ی پهنای تصحیح‌شده‌ی سیلورمن است.

h3 پهنای باندی که از کمینه کردن رابطه‌ی (۳.۳) به‌دست می‌آید که همان قاعده‌ی پهنای سیلورمن است.

h4 پهنای باندی که با روش کمترین توان‌های دوم اعتبارسنجی متقابل به‌دست آمده است.

h5 پهنای باندی که با روش اعتبارسنجی متقابل اریب به‌دست آمده است.

h6 پهنای باندی که از روش جای‌گذاری شیتر و جونز (۱۹۹۱) به‌دست آمده است.

h7 پهنای باندی که از روش جای‌گذاری پارک و مارون (۱۹۹۰) به‌دست آمده است.

h8 پهنای باندی که به روش خودگردان و از کمینه کردن رابطه‌ی (۲۲.۳) به‌دست آمده است.

h9 پهنای باند ترکیبی که به‌ازای $\alpha = 1$ و $\beta = 2$ در پهنای باند پیشنهادی $\left(\hat{h}_{CV}^{\alpha}\hat{h}_{PM}^{\beta}\right)^{\frac{1}{\alpha+\beta}}$ به‌دست آمده است.

h10 پهنای باند ترکیبی که به‌ازای $\alpha = 2$ و $\beta = 1$ در پهنای باند پیشنهادی $\left(\hat{h}_{CV}^{\alpha}\hat{h}_{PM}^{\beta}\right)^{\frac{1}{\alpha+\beta}}$ به‌دست آمده است.

h11 پهنای باند ترکیبی که به‌ازای $\alpha = 1$ و $\beta = 1$ در پهنای باند پیشنهادی $\left(\hat{h}_{CV}^{\alpha}\hat{h}_{PM}^{\beta}\right)^{\frac{1}{\alpha+\beta}}$ به‌دست آمده است.

در همه‌ی روش‌ها، جستجوی پهناهای باند روی دنباله‌ای از ۵۱۲ نقطه انجام گرفته است. در این‌جا نتایج را به‌ازای اندازه‌های نمونه ۲۵، ۵۰، ۱۰۰ و ۲۰۰ ارائه می‌کنیم.

۱.۳.۳ نتایج شبیه‌سازی

برای خلاصه و مقایسه کردن پهناهای باند مختلف، ابتدا پهناهای باند را انتخاب می کنیم و سپس، اریبی‌ها را برای هر یک از این روش‌ها محاسبه می‌کنیم. در نهایت، می‌توان روش‌های مختلف پهناهای باند را با معیارهای متفاوت ISE مقایسه کرد. هم‌چنین همه‌ی نتایج بر اساس ۲۵۰ تکرار به‌دست آمده است.

مقایسه‌ی اریبی برای پهناهای باند متفاوت

در نمودار ۴.۳، اریبی (c_1) برای روش‌های متعدد انتخاب پهنای باند در حالت‌های مختلف اندازه‌ی نمونه و شش مثال بیان شده نشان داده شده است.

از آن‌جایی که به جز روش خودگردان در مثال ۱، مقادیر اریبی منفی است، می‌توان از نمودار ۴.۳ چنین استنباط کرد که هر چه اندازه‌ی حجم نمونه بیشتر شود، از مقدار اریبی کاسته می‌شود. در مثال ۲، این مقدار کاسته شدن محسوس نیست اما در سایر روش‌ها، آن به خوبی مشاهده می‌شود. روش‌های انتخاب ذهنی به‌خصوص پهنای باند h2 که از رابطه‌ی (۶.۳) به‌دست آمده است و هم‌چنین روش جای‌گذاری شیتر و جونز (۱۹۹۱) که با h6 نشان داده شده است، به همواری کم گرایش دارند.

در توزیع‌های آمیخته و هم‌چنین توزیع نرمال (مثال‌ها ۳ تا ۶) رفتار برآوردگرهای پهناهای باند به روش جای‌گذاری و هم‌چنین خودگردان تقریبا مشابه است. اگر چه که در اندازه نمونه‌ی کوچک، اریبی پهنای باند به‌دست آمده از روش خودگردان کمتر از پهنای باند نتیجه شده از روش‌های جای‌گذاری است که هر چه حجم نمونه افزایش یابد، مقدار اریبی آن‌ها برابر می‌شود. هم‌چنین رفتار مشابهی از اریبی برآوردگرهای ترکیبی در مثال‌های ۱ تا ۶ مشاهده می‌شود.

مقایسه‌ی مقادیر ISE برای پهناهای باند متفاوت

در این بخش، می‌توان مقادیر ISE از برآوردهای چگالی بر اساس انتخاب پهناهای باند متفاوت به‌دست آمده‌اند. نتایج برحسب نمودار جعبه‌ای، توزیع ISEها را برای ۲۵۰ تکرار در فرآیند شبیه‌سازی نشان می‌دهد.

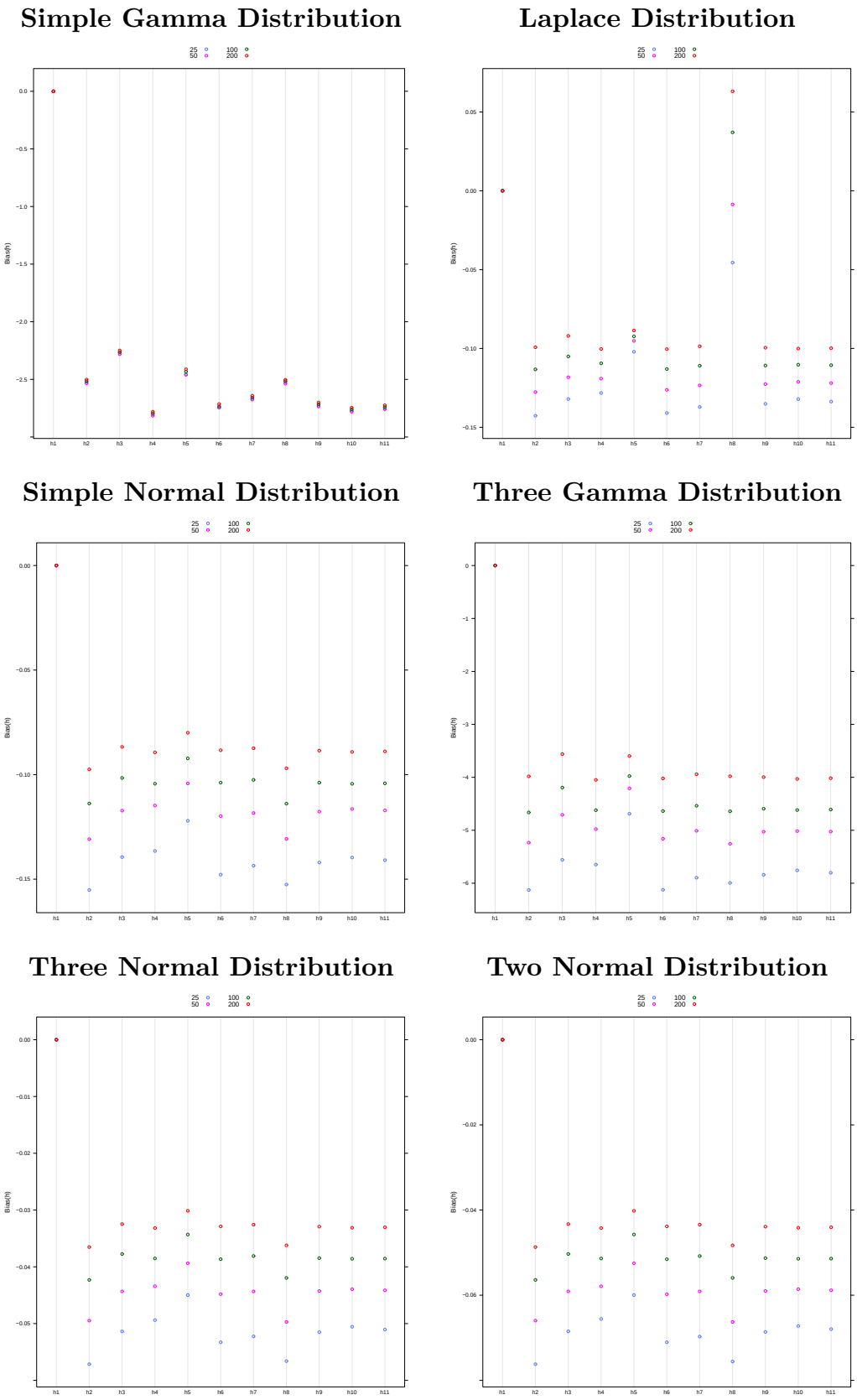
در نمودارهای ۵.۳ تا ۱۰.۳، نمودارهای ISEهای توزیع‌های مختلف در نظر گرفته شده رسم شده است.

همان‌طور که انتظار می‌رفت با افزایش حجم نمونه، مقدار ISE کاسته می‌شود که تمامی نمودارهای ۵.۳ تا ۱۰.۳ این موضوع را تایید می‌کنند. طبیعتا، با افزایش پیچیدگی روش برآورد، مقدار ISE افزایش می‌یابد.

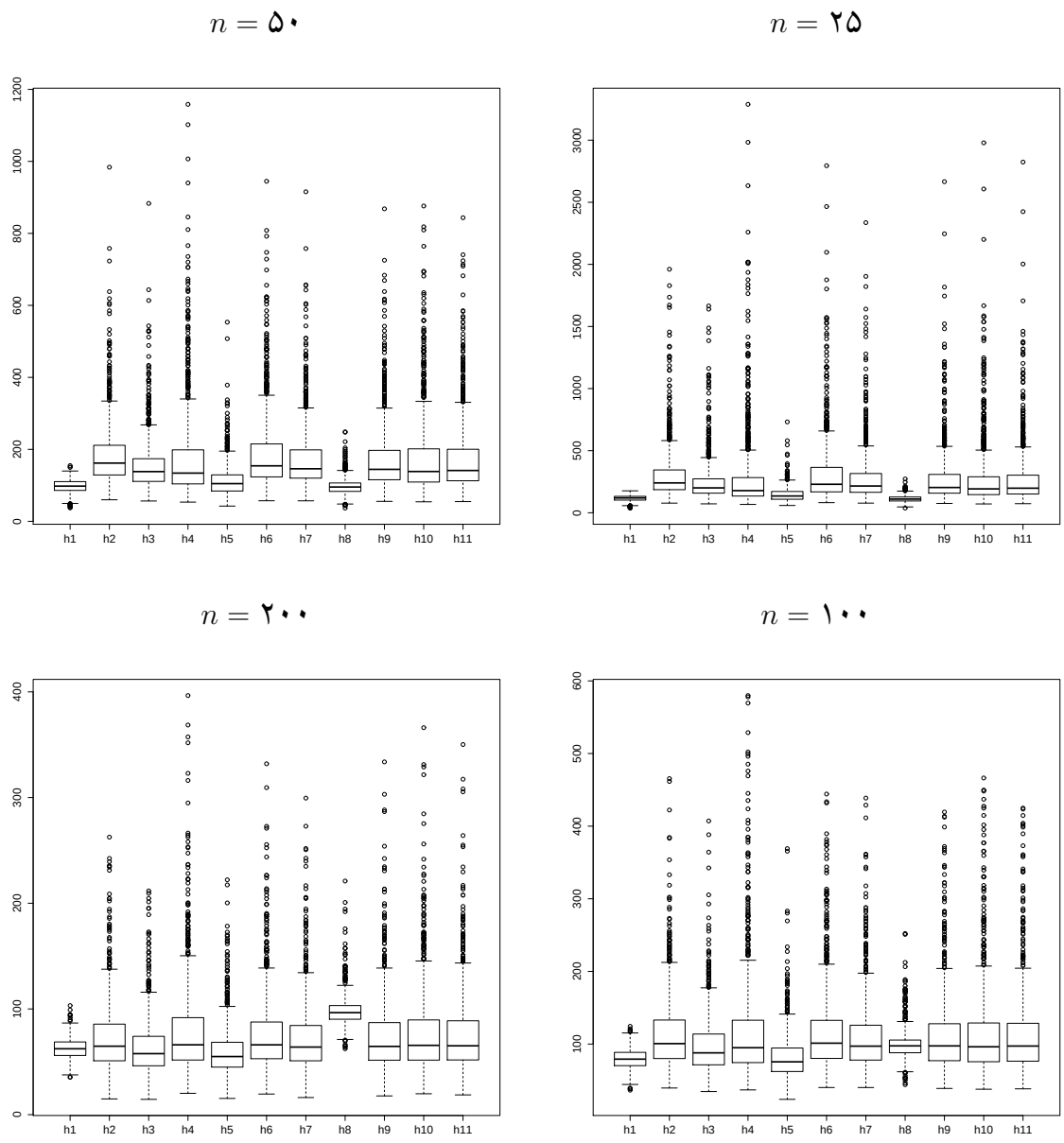
روش‌های اعتبارسنجی متقابل (شامل h4 و h5) بی‌ثباتی زیادی را در همه‌ی مثال‌ها نشان می‌دهد. روش‌های ترکیبی، در همه‌ی مثال‌های شبیه‌سازی رفتار مشابهی را دارند.

مقایسه‌ی فاصله‌ی L_1 و L_2 از ISE برای پهناهای باند متفاوت

برای داشتن ایده‌ی بهتری از فاصله‌ی بین مقادیر ISEای که با روش‌های مختلف پهناهای باند متفاوت و پهنای باند بهینه به‌دست آمده‌اند. نگاه جزئی‌تری به معیارهای c_4 و c_5 داریم. معمولا این دو معیار، از رایج‌ترین معیارهای مورد استفاده هستند.



شکل ۴.۳: مقایسه‌ی اریبی برای اندازه نمونه‌های متفاوت در مثال‌های ۱ تا ۶

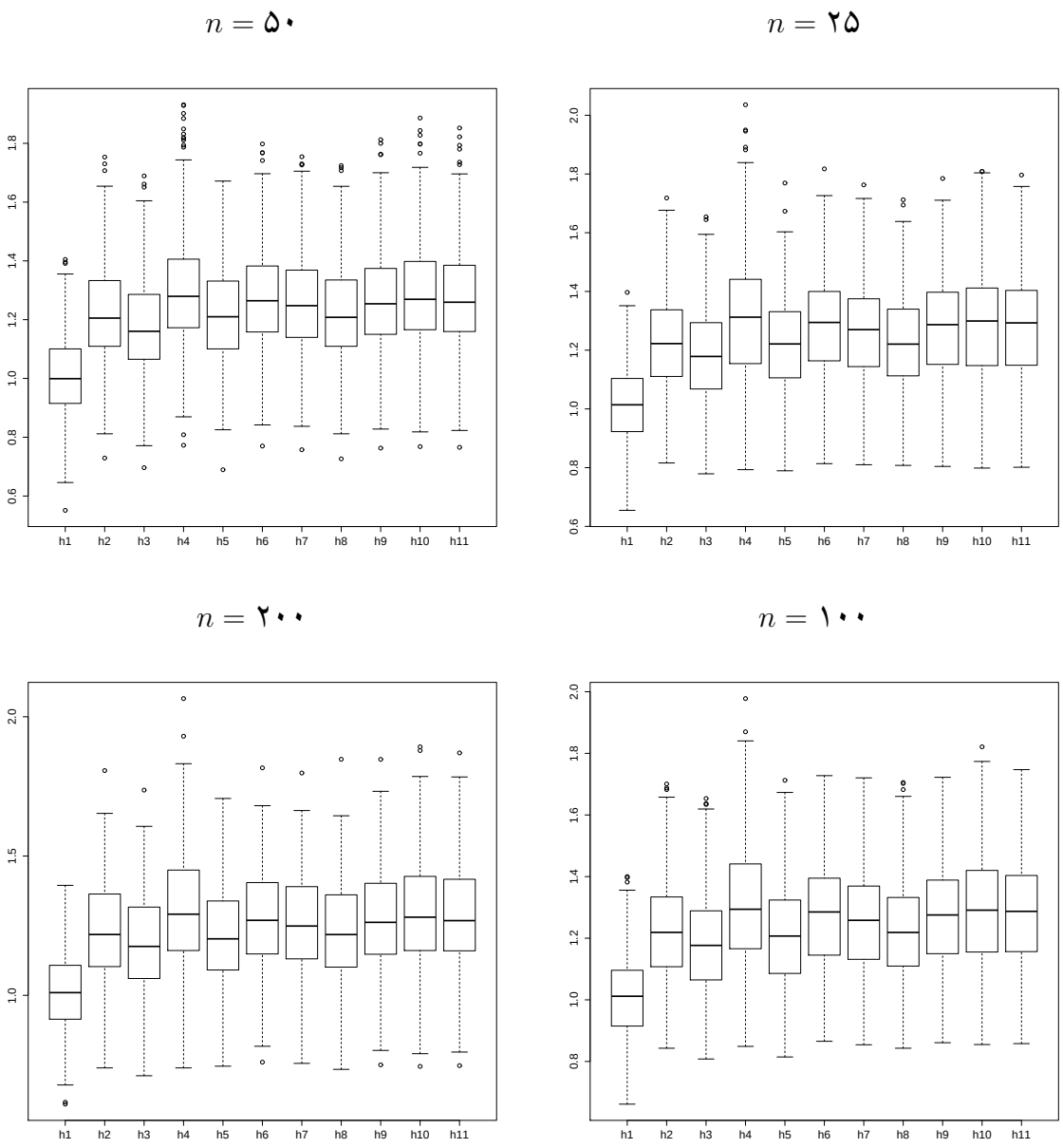


شکل ۵.۳: نمودار جعبه‌ای از مقادیر ISE برای توزیع لاپلاس (مثال ۱)

شکل‌های ۱۱.۳ و ۱۲.۳ فاصله‌ی $L_۱$ و $L_۲$ را برای مثال‌ها و اندازه‌ی نمونه‌های متفاوت نشان می‌دهد.

برای پهناهای باند به روش اعتبارسنجی متقابل، در مثال ۲ مقدار $c_۵$ بزرگ است. از آن‌جایی که چگالی مثال ۲ خیلی ناهموار نیست، بزرگی این مقدار حاکی از آن است که پهناهای باند تغییرپذیری زیادی دارند. اما با این وجود، اگر چگالی یک تابع را ندانیم، بر اساس فاصله‌ی $L_۱$، پیشنهاد می‌شود که از روش اعتبارسنجی متقابل اریب برای انتخاب پهنای باند استفاده کنیم. اگر بنا باشد از بین دو روش قاعده‌ی سرانگشتی سیلورمن، بر اساس این فاصله یکی انتخاب شود باید از روش $\mathrm{h3}$ یعنی روشی که از رابطه‌ی (۳.۳) به‌دست می‌آید، استفاده شود.

برای نمونه‌های بزرگ، فاصله‌ی $L_۱$ کوچکتر می‌شود. اما باید بزرگی مقادیر این فاصله را نیز به‌خصوص برای توزیع‌های نرمال و آمیخته‌ای از آن در نظر گرفت.

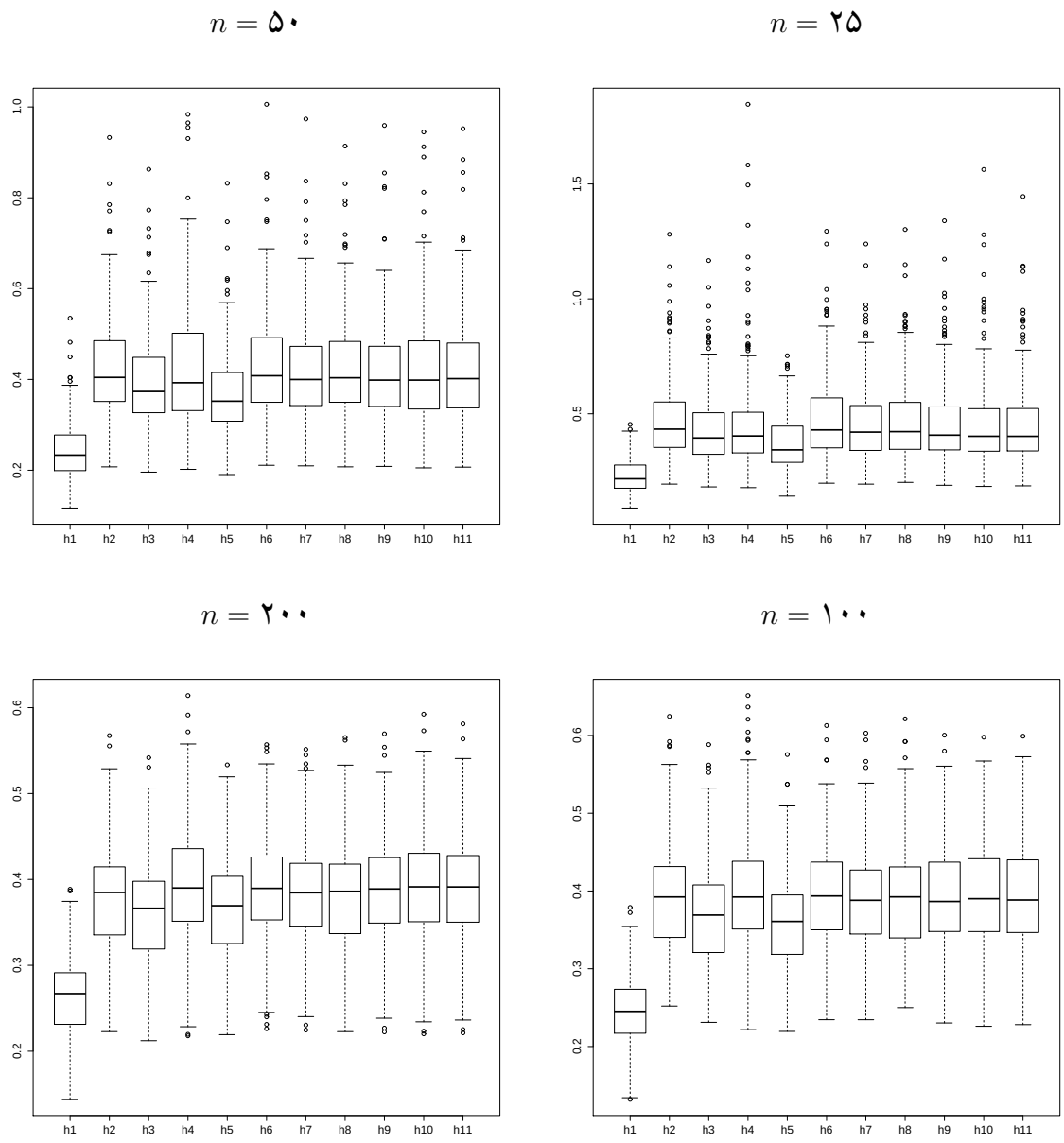


شکل ۶.۳: نمودار جعبه‌ای از مقادیر ISE برای توزیع گاما (مثال ۲)

برتری روش‌های ترکیب در برآورد پهناهای باند بر اساس فاصله‌ی $L_۱$، پایداری این روش‌ها نسبت به سایر روش‌ها می‌باشد.

از نمودار ۱۲.۳ نتایجی مشابه نمودار ۱۱.۳ مشخص می‌شود. در این‌جا مقادیر بزرگی برای نمونه‌های کوچک نیز به‌دست می‌آید. اگر چه تغییر در اندازه‌ی نمونه، تغییر محسوسی در نتایج به‌دست آمده برای توزیع گاما ندارد.

روش‌های اعتبارسنجی متقابل اریب و نیز قاعده‌ی سرانگشتی سیلورمن که از رابطه‌ی (۳.۳) به‌دست می‌آید، روش‌های پیشنهادی بر اساس فاصله‌ی $L_۲$ هستند اگر تابع چگالی واقعی را ندانیم.



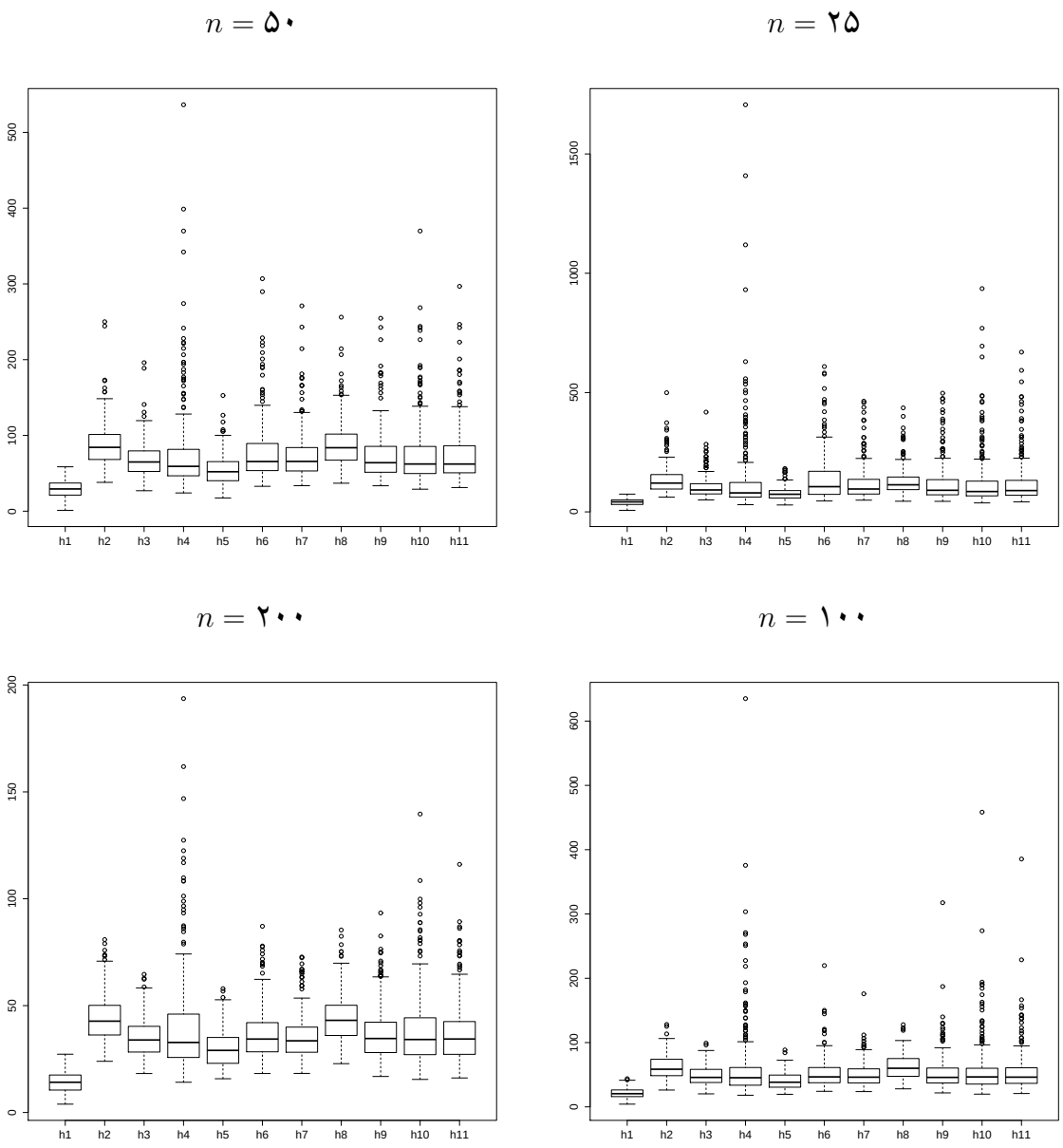
شکل ۷.۳: نمودار جعبه‌ای از مقادیر ISE برای آمیخته‌ی از سه توزیع گاما (مثال ۳)

۴.۳ نتیجه‌گیری

اولین نکته‌ای که باید در انتخاب یک برآوردگر برای پهنای باند در نظر گرفت توجه به خصوصیات داده‌ها و اینکه در چه وضعیت و موقعیتی، یک برآوردگر برتر است ، می‌باشد. تنها توجه به ملاک‌های عددی، رضایتبخش نیست.

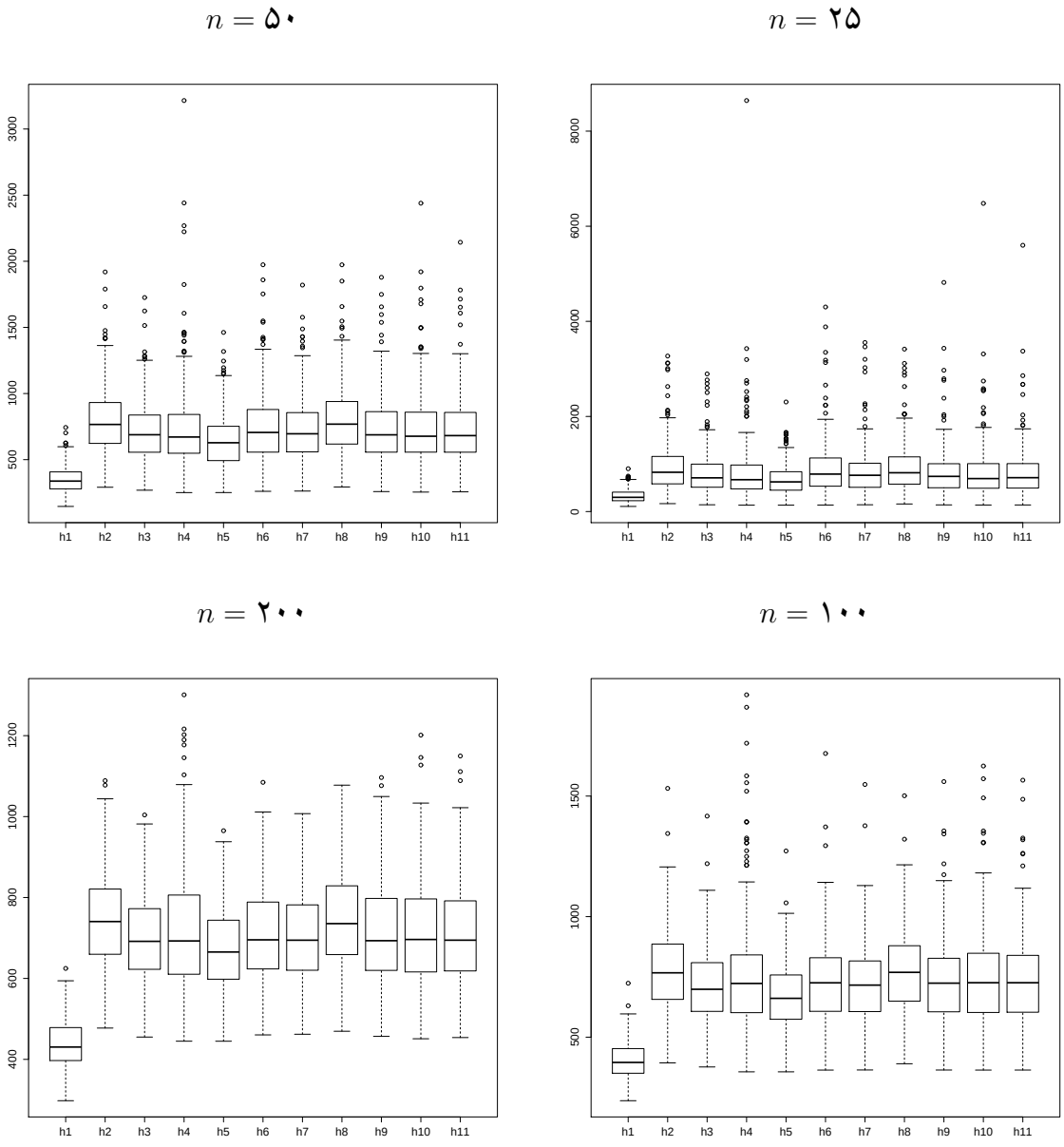
روش‌های اعتبارسنجی متقابل به برآوردهایی با اریبی کم ولی واریانس زیاد گرایش دارند و معمولا از آنها برای چگالی‌های مواج (ناهموار) در حالتی که حجم نمونه مناسب باشد، مورد استفاده قرار می‌گیرد. اما این روش، دارای تغییرپذیری زیادی است.

در روش‌های جای‌گذاری تغییرپذیری کمتر است اما بیش‌همواری را نتیجه می‌دهند. همان‌طوری که در

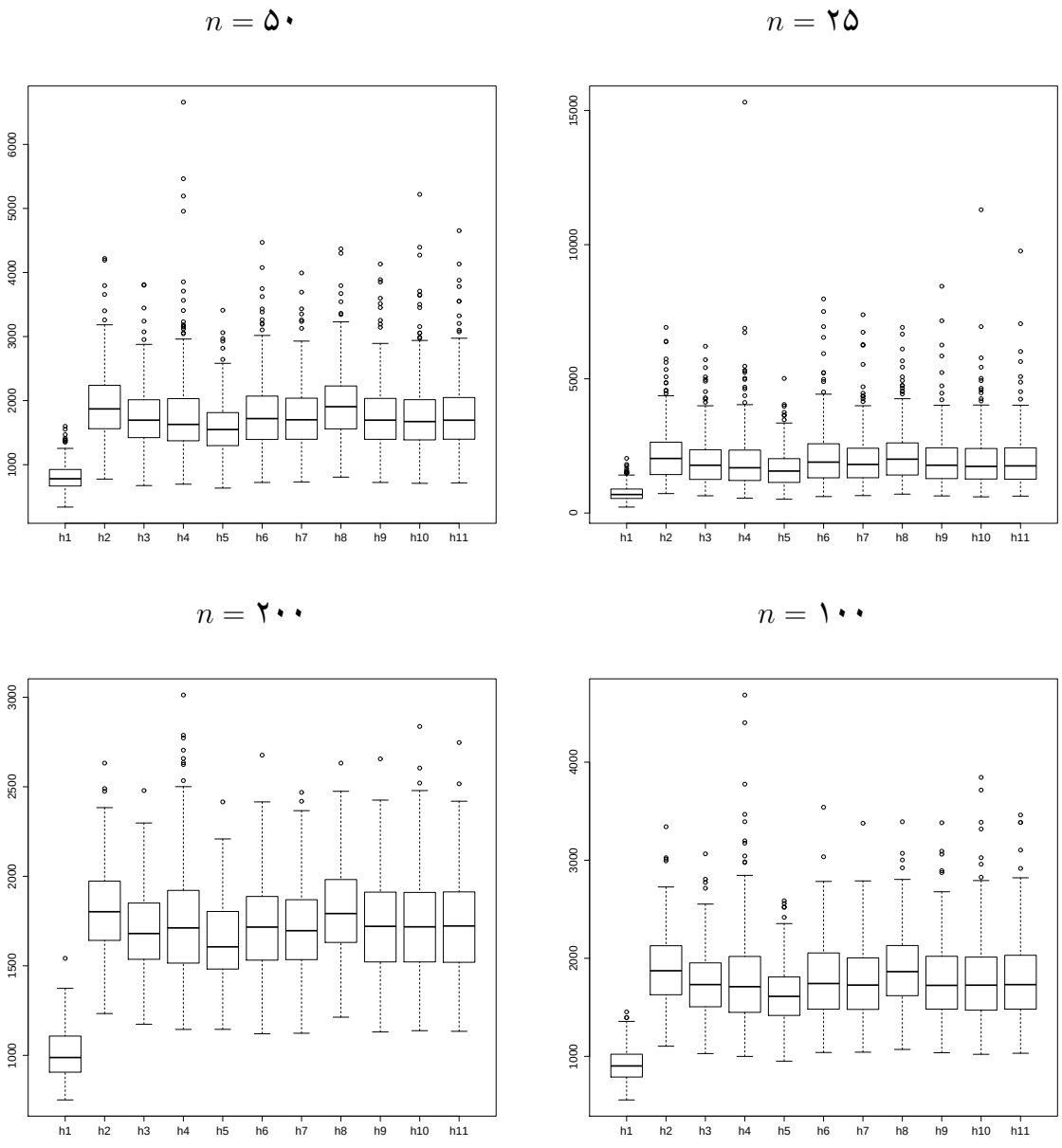


شکل ۸.۳: نمودار جعبه‌ای از مقادیر ISE برای توزیع نرمال (مثال ۴)

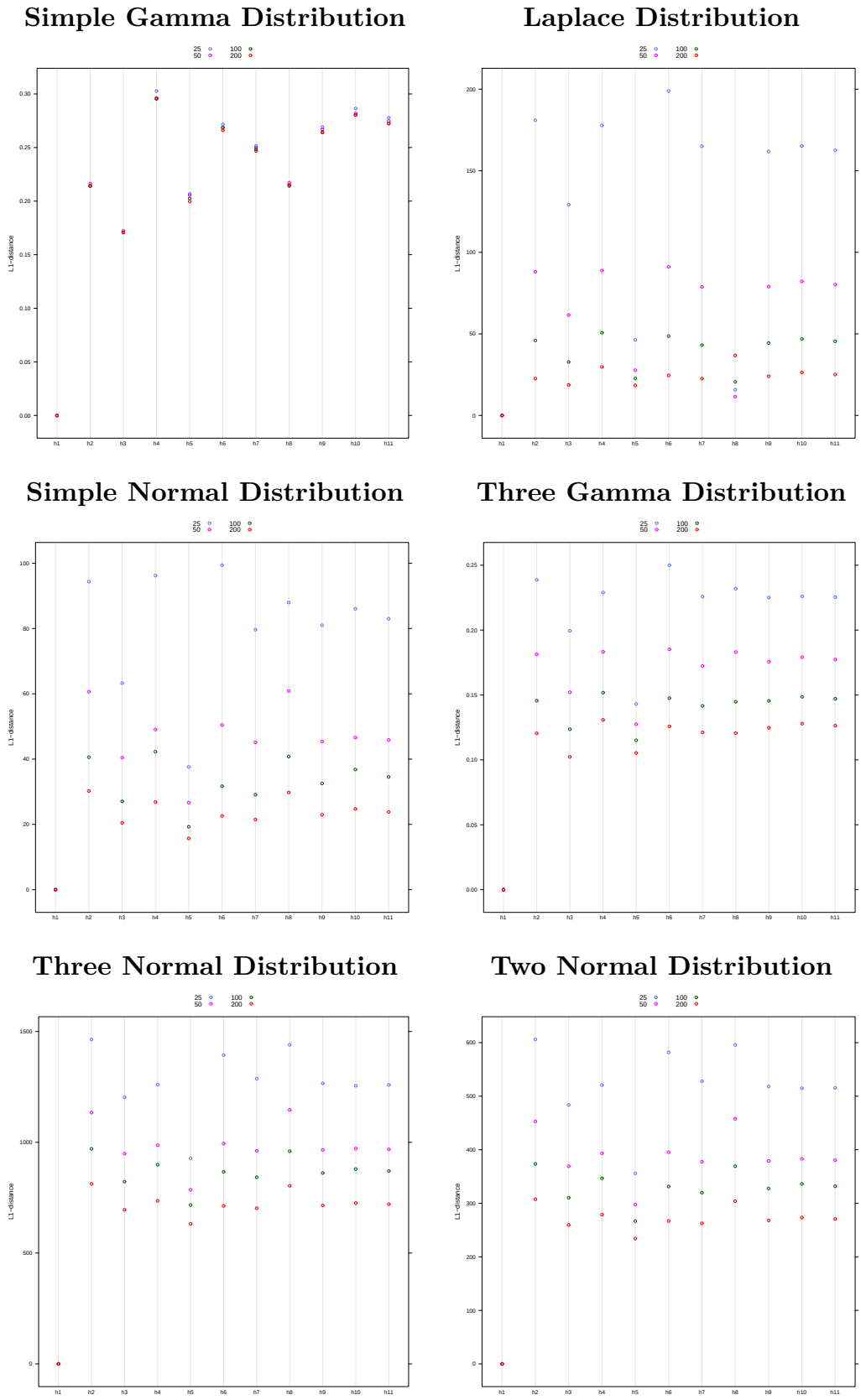
مثال‌های شبیه‌سازی نشان داده شد، با پیشنهاد روش‌های ترکیبی، تلاش‌هایی در راستای حل این مشکل صورت گرفته است.



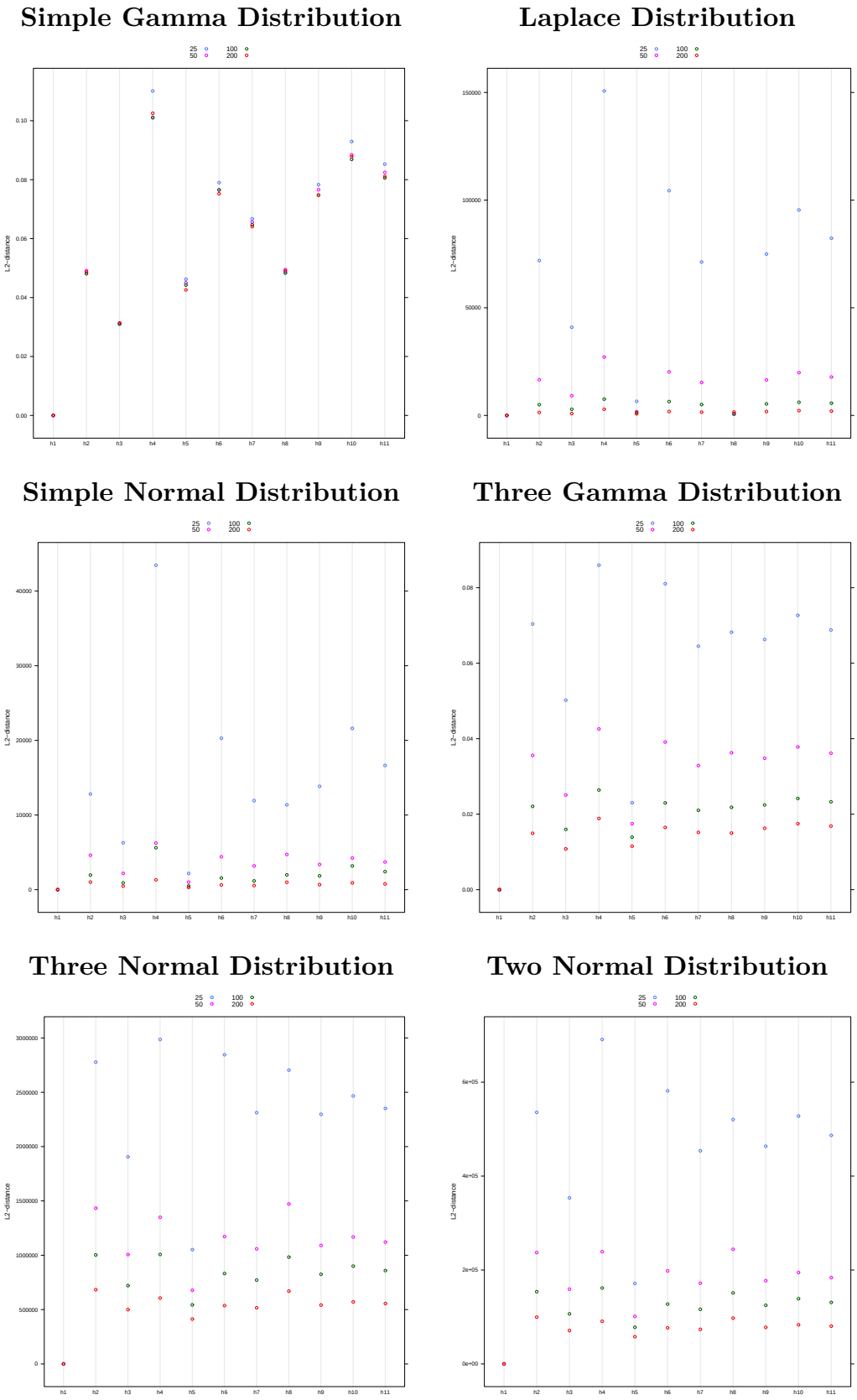
شکل ۹.۳: نمودار جعبه‌ای از مقادیر ISE برای آمیخته‌ی از دو توزیع نرمال (مثال ۵)



شکل ۱۰.۳: نمودار جعبه‌ای از مقادیر ISE برای آمیخته‌ی از سه توزیع نرمال (مثال ۶)



شکل ۱۱.۳: مقایسه‌ی فاصله‌ی L_1 از $\mathrm{ISE}(h_{opt})$ برای اندازه نمونه‌های متفاوت در مثال‌های ۱ تا ۶



شکل ۱۲.۳: مقایسه‌ی فاصله‌ی L_2 از $\mathrm{ISE}(h_{opt})$ برای اندازه نمونه‌های متفاوت در مثال‌های ۱ تا ۶

فصل ۴

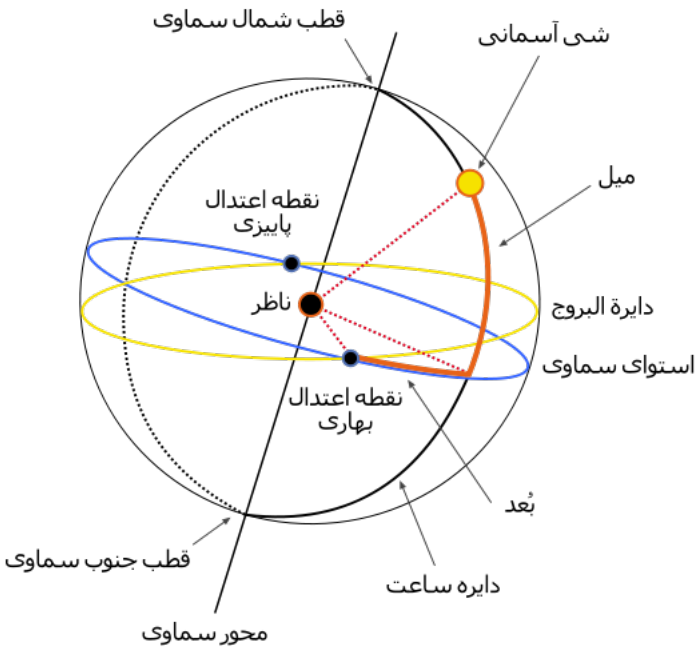
کاربرد روش‌های تعیین پهنای باند در یک مثال داده - اختری

۱.۴ مقدمه

اطلاعات و دانش انسان درباره سیارات، ستارگان، میانگین خلاء بین ستاره‌ای، کهکشان‌ها یا پدیده‌های رشد و نمو و بزرگ شدن، معمولاً محدود به مشاهدات متعددی است که اطلاعات محدودی در مورد شرایط اصولی یا اساسی به‌دست می‌دهد. دانش نجومی ما معمولاً مربوط به نمونه‌های ابتدایی فرآیندهای پیچیده روی پراکندگی‌های اتم‌ها است. با توجه به این دشواری‌ها برای ستاره‌شناسی، اغلب اصل مختصری برای فرض کردن پراکندگی‌های آماری خاص از آن، و روابط بین متغیرهای مشاهده شده، وجود دارد. برای مثال، ستاره‌شناسان به دفعات ثبت، متغیر مشاهده‌شده را اندازه می‌گیرند تا دامنه‌های گسترده را کاهش دهند و واحدهای فیزیکی را بردارند و بنابراین با توجیه مختصر فرض می‌کنند که باقیمانده متغیرها به‌طور نرمال توزیع شده‌اند. در صورتی‌که در عمل و به منظور افزایش کارآیی روش‌های به‌کار رفته در داده‌های اختری، لازم است این فرض آزمون شود. اما در شرایطی که حتی نمی‌توان نرمال بودن باقیمانده متغیرها را به‌عنوان توزیع مشاهدات در نظر گرفت استفاده از برآورد تابع چگالی روش کرنل کارساز می‌باشد.

مسئله برآورد تابع چگالی در تخمین‌های نجومی گوناگونی مختلف به‌کار می‌رود. از جمله، این کاربردها می‌توان به یافتن توزیع‌های جلیدیّه یا کهکشانی برای ردیابی توزیع ماده تیره شالوده، استفاده کرد. توزیع فوتون‌ها در یک تصویر اشعه ایکس یا اشعه گاما را می‌توانیم برای مجسم ساختن اشعه ایکس یا اشعه گاما آسمان، به‌کار ببریم. منحنی‌های (توزیع) نور حاصل از متغیر مشاهده شده ستاره‌ها و کوازارها (اخترنماها) به‌صورت گذرا، را می‌توانیم برای درک ماهیت تغییر پذیری‌شان، استفاده کنیم. توزیع جریانات ستاره‌ای در هاله کهکشان را می‌توانیم برای ردیابی دینامکی (مسیر حرکتی) کهکشان‌های کوتوله قطعه‌برداری شده مورد استفاده قرار دهیم.

نمودارهای فوتومتری (نورسنجی) در علم اخترشناسی کاربردهای فراوانی دارند که گاهی اوقات نمی‌توان توزیع پارامتری را به آن‌ها برازش داد. در این وضعیت‌ها، اغلب مطلوب است که چگالی را برای مقایسه با تئوری (نظریه) فیزیک نجومی تخمین بزنیم، تا روابط ما بین متغیرها، را در ذهن مجسم کنیم.



شکل ۱.۴: زاویه‌ی حرکت اجرام آسمانی

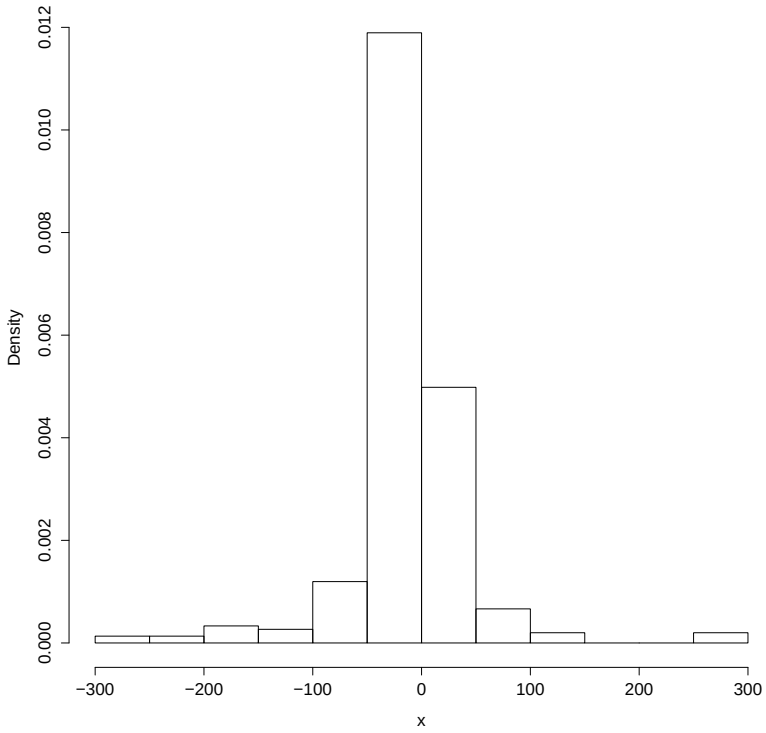
یک تخمین زننده ساده، که عموماً خیلی در اخترشناسی (نجوم) مورد استفاده قرار می گیرد بافت نگار است. اما برآوردکننده‌های پیوسته کرنل اغلب می تواند جمعیت های طبیعی واقعی‌تر را نشان دهد. در این راستا، گاهی اوقات پهنای باند را به صورت عینی انتخاب نموده یا به دلایل فیزیک نجومی، مستقل از مجموعه داده ای که در دست داریم انتخاب می کنیم. در وضعیت های دیگر، این پهنای باند را می توانیم از داده ها بر مبنای یک معیار آماری نیز تخمین بزنیم.

۲.۴ داده‌های اختری مورد مطالعه

در ادامه داده‌هایی را از هایپرکوس[۱] که کاتالوگی شامل حدوداً صد هزار داده مربوط به حرکت صدها هزار ستاره مختلف که توسط تلسکوپی به همین نام جمع آوری شده است می‌باشد، در اختیار گرفتیم تا روش‌های متفاوت تعیین پهنای باند در برآورد چگالی کرنل را مورد بررسی قرار دهیم. لازم به ذکر است که حرکت ستاره را در دو بعد مطالعه می‌شود. حال اگر سرعت حرکت ستارگان را برحسب زاویه بر زمان در نظر بگیریم، واحد اندازه‌گیری آن $\frac{mas}{yr}$ [۲] می‌باشد. شکل ۱.۴، زاویه‌ی حرکت اجرام آسمانی را نشان می‌دهد. در این مطالعه، یک مجموعه ۳۰۰ تایی از مجموعه داده‌های ذکر شده را در نظر گرفته‌ایم.

در شکل ۲.۴، بافت نگار این داده‌ها رسم شده است.

[۱] Hipparcos
[۲] Mili arc second/ year

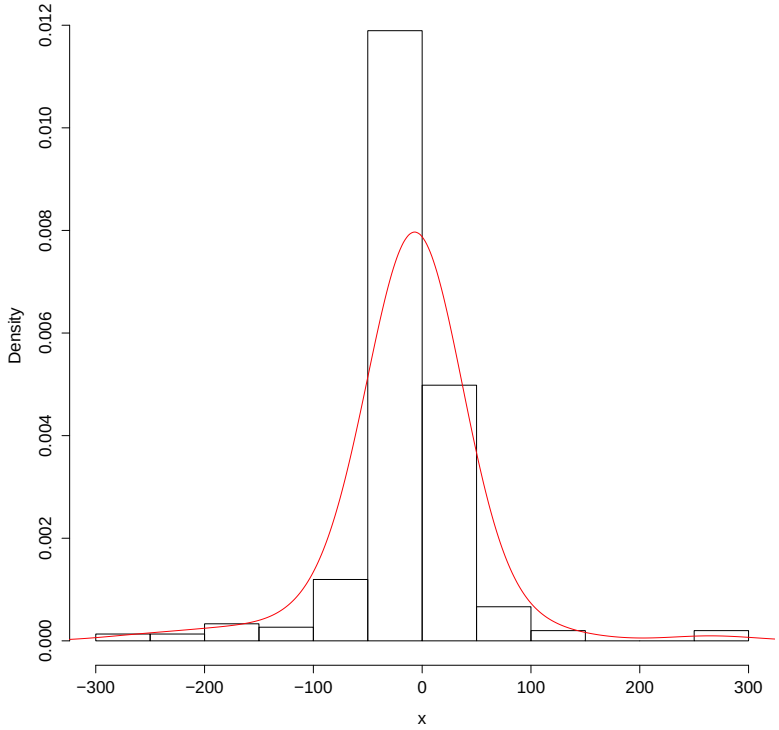


شکل ۲.۴: بافت نگار داده‌های ستاره‌شناسی

در ادامه، هر یک از پهناهای باند متفاوت به‌همراه نمودار برازش آنها به داده‌ها آورده شده است و در پایان مقایسه‌ای بین پهناهای باند مختلف از نظرهای اریبی و ISE انجام گرفته است.

همان‌طور که از جدول ۱.۴ می‌توان برداشت کرد، اریبی برآوردگری که به روش خودگردان به‌دست آمده است از سایر برآوردگرها کمتر است. همچنین برآوردگری که به روش قاعده‌ی سرانگشتی سیلورمن که از رابطه‌ی (۶.۳) به‌دست آمده است، در جایگاه بعدی قرار می‌گیرد.

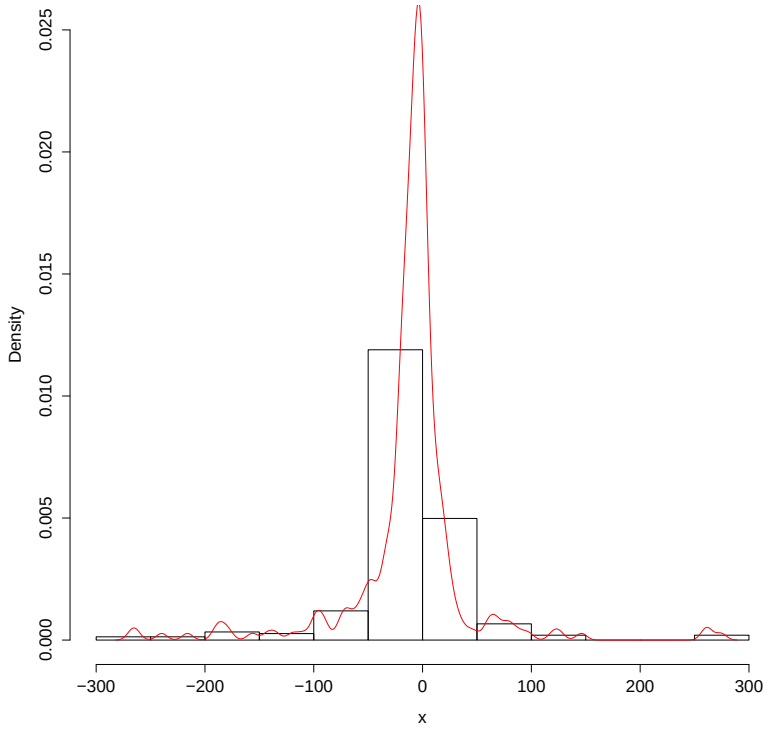
مقادیر ISE در جدول ۲.۴ نشان می‌دهد که بهترین کارآیی از آن برآوردگر خودگردان در انتخاب پهنای باند است. روش قاعده‌ی سرانگشتی سیلورمن که از رابطه‌ی (۳.۳) به‌دست می‌آید و همچنین روش اعتبارسنجی متقابل اریب، مقادیر ISE کمتری در مقایسه با سایر روش‌ها دارند.



شکل ۳.۴: نمودار بافت نگار و نمودار برازش داده‌ها به‌ازای پهنای باند بهینه

جدول ۱.۴: مقادیر اریبی پهناهای باند مختلف

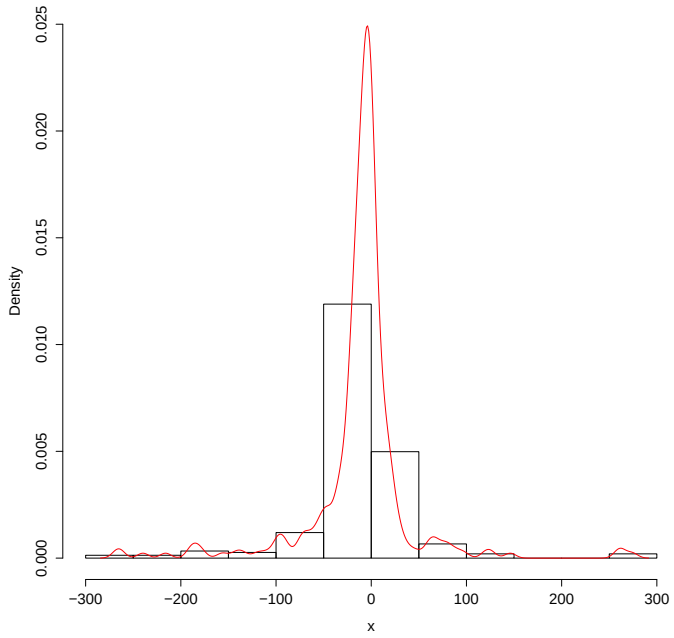
پهناهای باند	اریبی
h1	۰
h2	−۳۵/۷۳۲۶۱
h3	−۳۴/۸۷۳۰۹
h4	−۳۷/۶۷۱
h5	−۳۵/۲۴۵۵۸
h6	−۳۷/۱۱۹۳۳
h7	−۳۶/۴۶۷۵۸
h8	−۱/۷۷۹۵۳۳
h9	−۳۶/۹۱۵۹۷
h10	−۳۷/۳۱۵۳۲
h11	−۳۷/۱۲۱۴۳



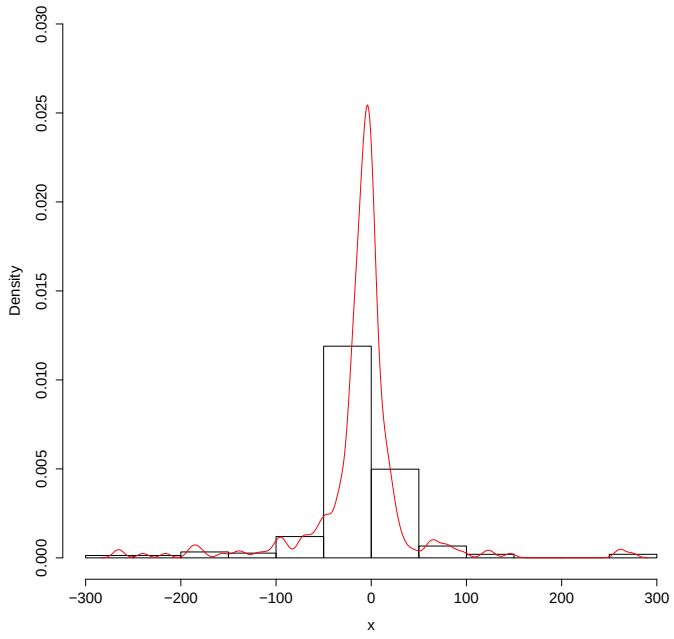
شکل ۴.۴: نمودار بافت نگار و نیز برازش داده‌ها به ازای پهنای باند قاعده‌ی سرانگشتی سیلورمن که از رابطه‌ی (۶.۳) به‌دست آمده است

جدول ۲.۴: مقادیر ISEهای پهناهای باند مختلف

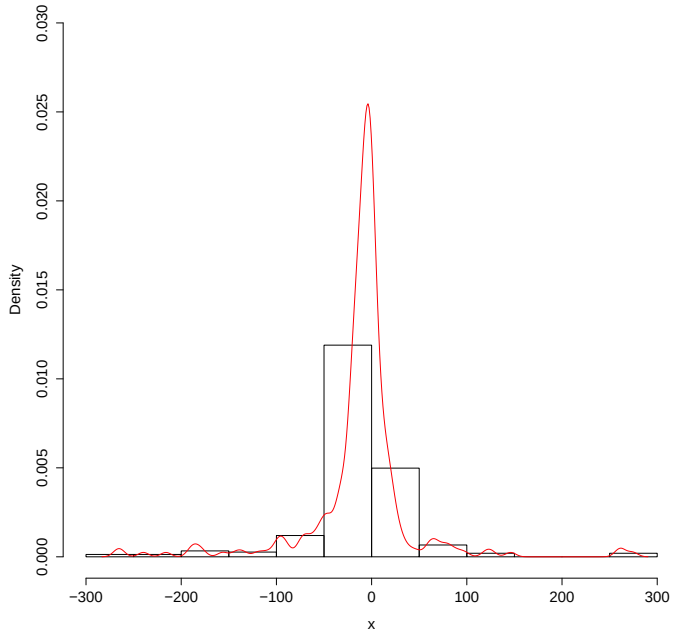
پهناهای باند	ISE
h1	۰
h2	۰/۰۵۹۹۰۹۸۹
h3	۰/۰۵۶۰۴۱۳۷
h4	۰/۰۶۹۳۴۷۳۷
h5	۰/۰۵۷۶۸۹۵۶
h6	۰/۰۶۶۶۰۵۲۶
h7	۰/۰۶۳۴۱۱۶۲
h8	۰/۰۰۰۰۸۰۷
h9	۰/۰۶۵۵۴۴۳۱
h10	۰/۰۶۷۵۹۶۵
h11	۰/۰۶۶۶۱۶۲۹



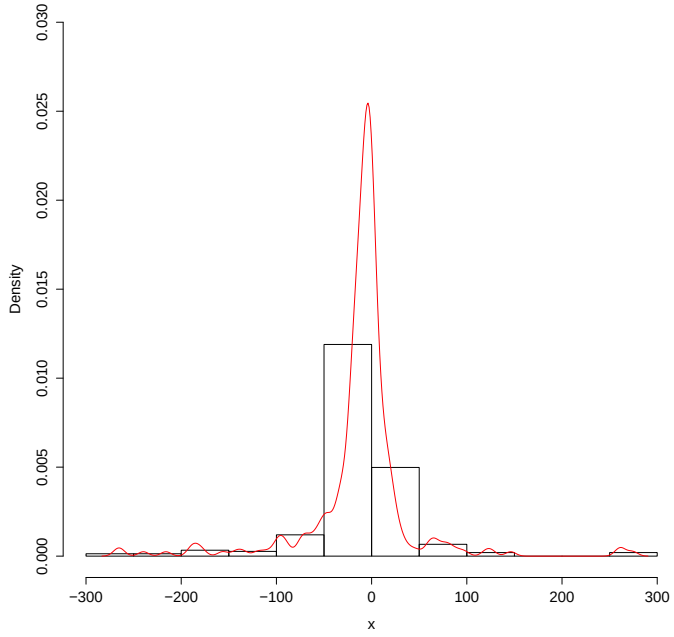
شکل ۵.۴: نمودار بافت نگار و نیز برازش داده‌ها به ازای پهنای باند قاعده‌ی سرانگشتی سیلورمن که از رابطه‌ی (۳.۳) به‌دست آمده است.



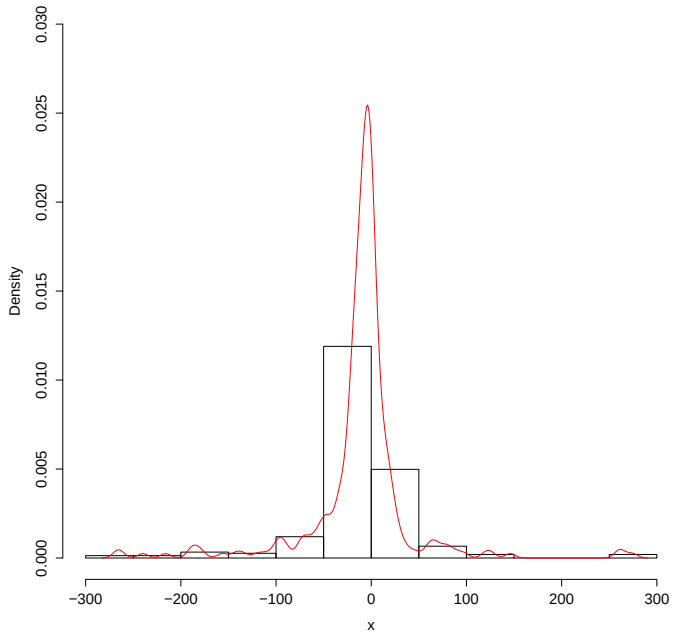
شکل ۶.۴: نمودار بافت نگار و نیز برازش داده‌ها به‌ازای پهنای باندی که بر اساس روش اعتبارسنجی متقابل کمترین توان‌های دوم به‌دست آمده است.



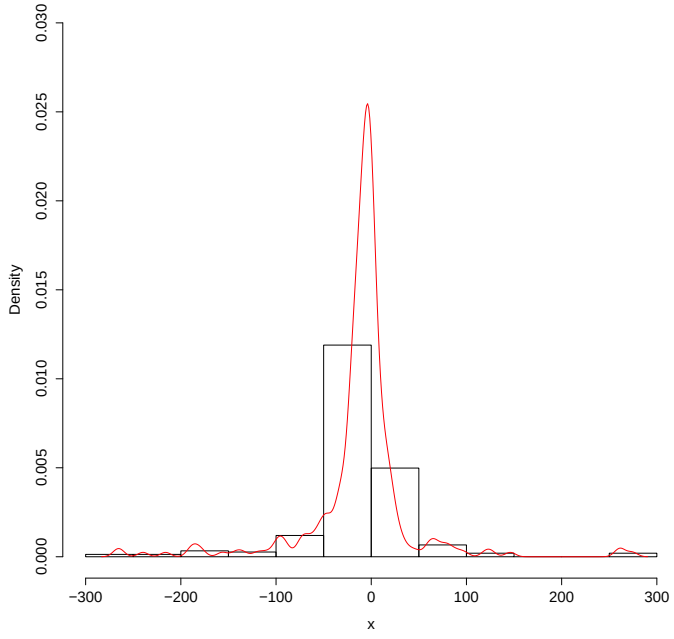
شکل ۷.۴: نمودار بافت نگار و نیز برازش داده‌ها به‌ازای پهنای باندی که بر اساس روش اعتبارسنجی متقابل اریب به‌دست آمده است.



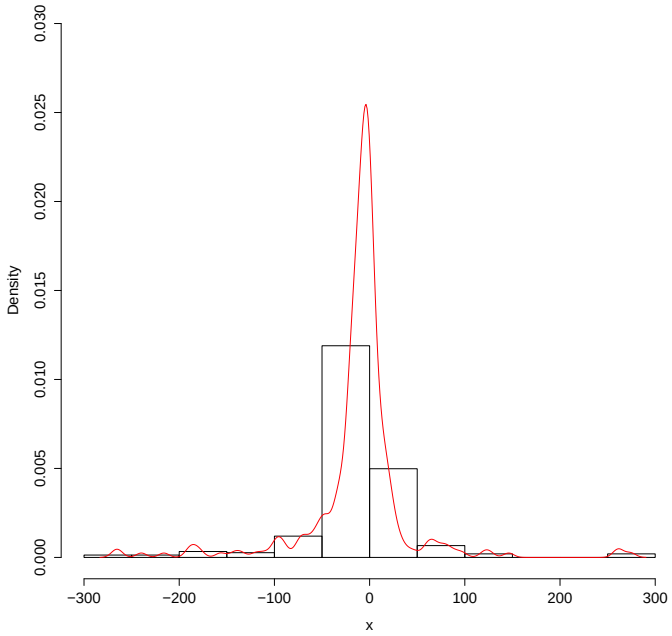
شکل ۸.۴: نمودار بافت نگار و نیز برازش داده‌ها به‌ازای پهنای باندی که از روش جای‌گذاری پارک و مارون به‌دست آمده است.



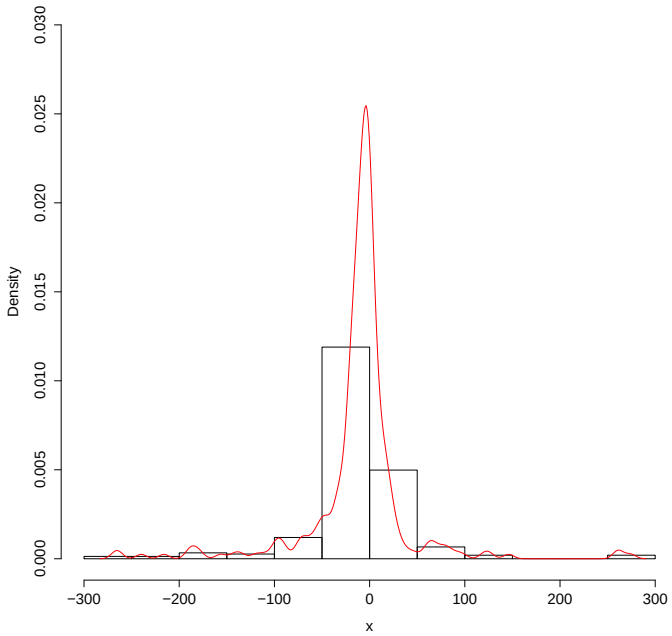
شکل ۹.۴: نمودار بافت نگار و نیز برازش داده‌ها به‌ازای پهنای باندی که از روش جای‌گذاری شیتر و جونز به‌دست آمده است.



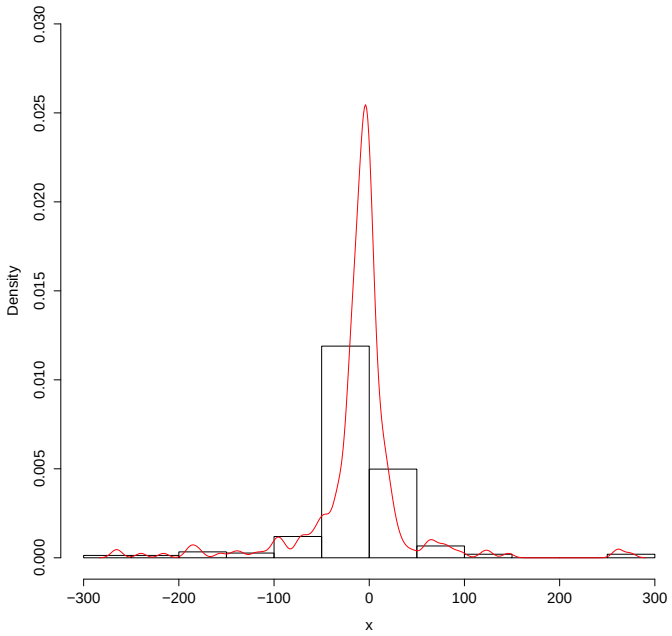
شکل ۱۰.۴: نمودار بافت نگار و نیز برازش داده‌ها به ازای پهنای باندی که از روش خودگردان به‌دست آمده است.



شکل ۱۱.۴: نمودار بافت نگار و نیز برازش داده‌ها به ازای پهنای باندی که با روش mix1 به‌دست آمده است.



شکل ۱۲.۴: نمودار بافت نگار و نیز برازش داده‌ها به‌ازای پهنای باندی که با روش mix2 به‌دست آمده است.



شکل ۱۳.۴: نمودار بافت نگار و نیز برازش داده‌ها به‌ازای پهنای باندی که با روش mix3 به‌دست آمده است.

۳.۴ نتیجه‌گیری و آینده تحقیق

در این پایان‌نامه، روش‌های مختلف برآورد تابع چگالی مورد بررسی قرار گرفته است. ابتدا مرور مختصری بر روش‌های رایج بافت نگار و ساده داشتیم و سپس به بررسی جامع‌تری از روش کرنل پرداخته‌ایم. از آن‌جایی پارامتر پهنای باند یک نقش کلیدی در این روش‌ها و به‌خصوص روش کرنل دارد، با تعریف و بیان معیارهایی برای اندازه‌گیری خطا، به مقایسه‌ی روش‌های مختلفی که در تعیین پهنای باند وجود داشت پرداخته‌ایم و در این راستا به معرفی دو دسته‌ی کلی روش‌های انتخاب جایگزین و اعتبارسنجی متقابل و هم‌چنین ترکیبی از این دو روش پرداخته‌ایم. در نهایت، با انجام شبیه‌سازی برای شش توزیع مختلف (لاپلاس، گاما، نرمال، آمیخته‌ای از سه توزیع گاما، آمیخته‌ای از دو توزیع نرمال و هم‌چنین آمیخته‌ای از توزیع نرمال) معیارهای اریبی، میانگین و انحراف‌معیار ISE، و همچنین میانگین فاصله‌های L_1 و L_2 کمیت ISE را در روش‌های مختلف به‌کار گرفته شده برای برآورد پهنای باند محاسبه شدند.

در راستای مطالعات انجام شده در این پایان‌نامه می‌توان موارد زیر را به عنوان آینده‌ی تحقیق مورد ارزیابی و مطالعه‌ی بیشتری قرار داد:

۱. بررسی کاربرد برآوردگرهای پهنای باند متفاوت در مدل‌های رگرسیونی ناپارامتری.

۲. دستیابی به برآوردگرهای بهینه پهنای باند بر اساس معیار AMISE

۳. ارائه‌ی برآوردگرهای پهنای باند بر اساس مشتق کرنل یا کرنل‌های مرتبه‌های بالاتر

پیوست آ

برنامه‌های نوشته‌شده با نرم‌افزار R

کد مربوط به شکل ۱.۱:

```
library(SemiPar)
data(age.income)
attach(age.income)
plot(age,log.income,ylim=c(10,15),pch=4,cex=0.4,col="gray")
abline(lm(log.income~age))
lines(ksmooth(age,log.income, "normal", bandwidth = 8),lwd=2,
 col = "red")
```

کد مربوط به شکل ۲.۱:

```
library(SemiPar)
require(ks)
require(kerdiest)
data(age.income)
attach(age.income)
x<-seq(25,35,0.001)
y.1<-function(x) 10+dnorm(x,mean=30,sd=1.3)
lines(x,y.1(x),col=4)
x<-seq(50,60,0.001)
y.2<-function(x) 10+dnorm(x,mean=55,sd=1.3)
lines(x,y.2(x),col=4)
segments(30, 0, 30, 13.58, col=1,lty=2)
```

```
segments(55, 0, 55, 13.6, col=1,lty=2)
segments(55, 0, 55, 13.6, col=1,lty=2)
segments(24, 13.25, 35.5, 14, col=1,lty=2,lwd=1)
segments(47, 14.2, 60.2, 13.45, col=1,lty=2,lwd=1)
text(29.5,13.8,"u",cex=0.8)
text(55,13.9,"v",cex=0.8)
```

کد مربوط به شکل ۲.۱:

```
library(graphics)
bwt<-c(1050,1175,1130,1230,1310,1500,1575,1600,1680,
       1720,1750,1760,1770,1930,2015,2090,2275,2500,
       2600,2700,2950,3160,3400,3640,2830,1030,1100,
       1185,1225,1262,1295,1300,1410,1550,1715,1720,
       1820,1890,1940,2041,2200,2200,2270,2400,2440,2550,
       2570,2560,2730,3005)
x <- bwt/1000
op<-par(mfrow=c(2,2))
h <- 0.2
bins <- seq(0.7, max(x)+h, by=h)
hist(x, breaks=bins,freq=F,main=" ",xlim=c(0,5),ylim=c(0,.8),
xlab="birthweight(kg)")

h <- 0.8
bins <- seq(0.7, max(x)+h, by=h)
hist(x, breaks=bins,freq=F,main=" ",xlim=c(0,5),ylim=c(0,.8),
xlab="birthweight(kg)")

h <- 0.4
bins <- seq(0.7, max(x)+h, by=h)
hist(x, breaks=bins,freq=F,main=" ",xlim=c(0,5),ylim=c(0,.8),
xlab="birthweight(kg)")

h <- 0.4
bins <- seq(0.9, max(x)+h, by=h)
hist(x, breaks=bins,freq=F,main=" ",xlim=c(0,5),ylim=c(0,.8),
```

```
xlab="birthweight(kg)")
par(op)
```

کد مربوط به شکل ۲.۲:

```
library(graphics)
EL<-c(4.37,4.70,1.68,1.75,4.35,1.77,4.25,4.10,4.05,
      1.90,4.00,4.42,1.83,1.83,3.95,4.83,3.87,1.73,
      3.92,3.20,2.33,4.57,3.58,3.70,4.25,3.58,3.67,
      1.90,4.13,4.53,4.10,4.12,4.00,4.93,3.68,1.85,
      3.83,1.85,3.80,3.80,3.33,3.73,1.67,4.63,1.83,
      2.03,2.72,4.03,1.73,3.10,4.62,1.88,3.52,3.77,
      3.43,2.00,3.73,4.60,2.93,4.65,4.18,4.58,3.50,
      4.62,4.03,1.97,4.60,4.00,3.75,4.00,4.33,1.82,
      1.67,3.50,4.20,4.43,1.90,4.08,3.43,1.77,4.50,
      1.80,3.70,2.50,2.27,2.93,4.63,4.00,1.97,3.93,
      4.07,4.50,2.25,4.25,4.08,3.92,4.73,3.72,4.50,
      4.40,4.58,3.50,1.80,4.28,4.33,4.13,1.95)

plot(dkde(x = EL, deriv.order = 0, kernel = "uniform",
          h=0.25),type="l",xlab="Euroption length(min)",
            ylab="Density Estimation",sub=" ",main=" ")

h <- 0.5
bins <- seq(1.5, max(EL)+h, by=h)
hist(EL, breaks=bins,freq=F,xlim=c(0,5),ylim=c(0,.8),
     xlab="Euroption length(min)",ylab="Density Estimation",sub=" ",
     main=" ")
```

کد مربوط به شکل ۳.۲:

```
x=c(3, 4.5, 5.0, 8, 9)
plot(x,rep(0,length(x)),type="p",pch=4,ylim=c(0,0.2),xlim=c(0,12),
ylab="Density",)
lines(seq(0,6,0.1),0.2*dnorm(seq(0,6,0.1),mean=3,sd=1),
lty=2,col=2)
```

```
lines(seq(0.5,7.5,0.1),0.2*dnorm(seq(0.5,7.5,0.1),mean=4.5,
sd=1),
lty=2,col=2)
lines(seq(2,8,0.1),0.2*dnorm(seq(2,8,0.1),mean=5,sd=1),
lty=2,col=2)
lines(seq(5,11,0.1),0.2*dnorm(seq(5,11,0.1),mean=8,sd=1),
lty=2,col=2)
lines(seq(6,12,0.1),0.2*dnorm(seq(6,12,0.1),mean=9,sd=1),
lty=2,col=2)
lines(density(x,kernel = "gaussian",bw=1))
abline(h=0,lty=2)
```

کد مربوط به جدول ۱.۲:

```
RecT=function(x) 0.5*(abs(x)<1)
x=seq(-3,3,0.001)
plot(x,RecT(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
#Triangular Function
TriA=function(x) (1-abs(x))*(abs(x)<1)
x=seq(-3,3,0.001)
plot(x,TriA(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
#Gaussian
Gaus=function(x) (1/sqrt(2*pi))*exp(-0.5*(x^2))
x=seq(-3,3,0.001)
plot(x,Gaus(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
#Epanechnikov
Epan=function(x) 0.75*(1-x^2)*(abs(x)<1)
x=seq(-3,3,0.001)
plot(x,Epan(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
#Quartic
Quar=function(x) (15/16)*((1-x^2)^2)*(abs(x)<1)
x=seq(-3,3,0.001)
plot(x,Quar(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
#Triweight
TriW=function(x) (35/32)*((1-x^2)^3)*(abs(x)<1)
x=seq(-3,3,0.001)
```

```
plot(x,TriW(x),type="l",lty=1,xlab="x",ylab="K(x)",
ylim=c(0,1.2))
#Tricube
TriC=function(x) (70/81)*((1-(abs(x))^3)^3)*(abs(x)<1)
x=seq(-3,3,0.001)
plot(x,TriC(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
#Cosine
Cosine=function(x) (pi/4)*cos(pi*x/2)*(abs(x)<1)
x=seq(-3,3,0.001)
plot(x,Cosine(x),type="l",lty=1,xlab="x",ylab="K(x)",ylim=c(0,1))
```

کد مربوط به شکل ۴.۲:

```
f<-function(x) 0.75*dnorm(x)+0.25*dnorm(x,1.5,sd=1/3)
f<-Vectorize(f)
simulate<-function(N){
  vec<-numeric(N)
   for(i in 1:N){
     s1<-rbinom(1,1,.75)
     if(s1==1) z=rnorm(1)
       else z=rnorm(1,1.5,1/3)
            vec[i]<-z
                 }
   return(vec)       }
Y<-simulate(1000)
x=seq(-3,3,0.01)
x1=seq(-0.5,0.5,0.01)
plot(x,f(x),type="l",ylim=c(0,0.45))
lines(density(Y,bw=0.06,kernel = "epanechnikov"),col=2)
lines(x1,Epan(x1)-0.56)
abline(h=0)

plot(x,f(x),type="l",ylim=c(0,0.45))
lines(density(Y,bw=0.54,kernel = "gaussian"),col=2)
lines(x1,Gaus(x1)-0.35)
abline(h=0,lty=2)
```

```
plot(x,f(x),type="l",ylim=c(0,0.45))
lines(density(Y,bw=0.18,kernel = "cosine"),col=2)
lines(x1,Cosine(x1)-0.55)
abline(h=0,lty=2)
```

کد مربوط به شکل ۴.۲:

```
f<-function(x) 0.75*dnorm(x)+0.25*dnorm(x,1.5,sd=1/3)
f<-Vectorize(f)
simulate<-function(N){
  vec<-numeric(N)
  for(i in 1:N){
    s1<-rbinom(1,1,.75)
    if(s1==1) z=rnorm(1)
    else z=rnorm(1,1.5,1/3)
    vec[i]<-z
  }
  return(vec)       }
Y<-simulate(1000)
x=seq(-3,3,0.01)
plot(x,f(x),type="l",ylim=c(0,0.45))
lines(density(Y,bw=0.18,kernel = "epanechnikov"),col=2)
lines(density(Y,bw=0.18,kernel = "gaussian"),col=2)
```

کد مربوط به شکل ۱.۳:

```
library(dbEmpLikeGOF)
data(Snow)
snow
plot(density(snow,bw=6),xlab="Snowfall (inches)",
ylab="Density Estimate",main=" ")
plot(density(snow,bw=12),xlab="Snowfall (inches)",
ylab="Density Estimate",main=" ")
```

کد مربوط به شکل ۲.۳:

```
EL<-c(4.37,4.70,1.68,1.75,4.35,1.77,4.25,4.10,4.05,
      1.90,4.00,4.42,1.83,1.83,3.95,4.83,3.87,1.73,
      3.92,3.20,2.33,4.57,3.58,3.70,4.25,3.58,3.67,
      1.90,4.13,4.53,4.10,4.12,4.00,4.93,3.68,1.85,
      3.83,1.85,3.80,3.80,3.33,3.73,1.67,4.63,1.83,
      2.03,2.72,4.03,1.73,3.10,4.62,1.88,3.52,3.77,
      3.43,2.00,3.73,4.60,2.93,4.65,4.18,4.58,3.50,
      4.62,4.03,1.97,4.60,4.00,3.75,4.00,4.33,1.82,
      1.67,3.50,4.20,4.43,1.90,4.08,3.43,1.77,4.50,
      1.80,3.70,2.50,2.27,2.93,4.63,4.00,1.97,3.93,
      4.07,4.50,2.25,4.25,4.08,3.92,4.73,3.72,4.50,
      4.40,4.58,3.50,1.80,4.28,4.33,4.13,1.95)
h_os<-0.467
plot(density(EL,bw=h_os),type="l",xlab="Euroption time (min)",
ylab="Density",main=" ")
plot(density(EL,bw=h_os/2),type="l",xlab="Euroption time (min)",
ylab="Density",main=" ")
plot(density(EL,bw=h_os/4),type="l",xlab="Euroption time (min)",
ylab="Density",main=" ")
plot(density(EL,bw=h_os/5),type="l",xlab="Euroption time (min)",
ylab="Density",main=" ")
```

کد مربوط به شکل ۳.۳

```
Laplace<-function(x) 4*exp(-abs(8*(x-0.5)))
curve(Laplace,-0.2,1.2,ylab=" ",main="Laplace Distribution")

sGamma<-function(x) dgamma(5*x,(1.5)^2,1.5)
curve(sGamma,-0.2,1.2,ylab=" ",main="Simple Gamma Distribution")

mGamma<-function(x) (1/3)*dgamma(8*x,(1.5)^2,1.5)+(1/3)*dgamma(8*x,3^2,3)
+(1/3)*dgamma(8*x,6^2,6)
curve(mGamma,-0.2,1.2,ylab=" ",main="Three Gamma Distribution")

sNorm<-function(x) dnorm(x,mean=0.5,sd=0.2)
curve(sNorm,-0.2,1.2,ylab = " ",main="Simple Normal Distribution")
```

```
m2Norm<-function(x) (1/2)*dnorm(x,0.35,0.1)+(1/2)*dnorm(x,0.65,0.1)
curve(m2Norm,-0.2,1.2,ylab=" ",main="Two Normal Distribution")

m3Norm<-function(x) (1/3)*dnorm(x,0.25,0.075)+(1/3)*dnorm(x,0.5,0.075)
+(1/3)*dnorm(x,0.75,0.075)
curve(m3Norm,-0.2,1.2,ylab=" ",main="Three Normal Distribution")
```

شبیه‌سازی

```
library(kedd)
library(ks)
library(smoothmest)
library(dbEmpLikeGOF)
library(decon)
library(lattice)
set.seed(1234)
Laplace<-function(x) 4*exp(-abs(8*(x-0.5)))
curve(Laplace,-0.2,1.2)


N.sim=250
BW<-matrix(0,nr=N.sim,nc=11)
ISE<-matrix(0,nr=N.sim,nc=11)
Bias.mat<-matrix(0,nr=4,nc=11)
ISE.L2<-matrix(0,nr=4,nc=11)
ISE.L1<-ISE.mean<-ISE.sd<-matrix(0,nr=4,nc=11)
row.names(Bias.mat)<-row.names(ISE.L2)<-row.names(ISE.L1)
<-row.names(ISE.mean)<-row.names(ISE.sd)<-c("25","50","100","200")
colnames(BW)<-colnames(ISE)<-colnames(Bias.mat)<-colnames(ISE.L2)
<-colnames(ISE.L1)<-c("h1","h2","h3","h4","h5","h6","h7",
"h8","h9","h10","h11")

n.vec<-c(25,50,100,200)
for(k in 1:length(n.vec){
for(i in 1:N.sim){
```

```
  x<-rdoublex(k,0.5,1/8)
  BW[i,1]<-h.amise(x)$h
  BW[i,2]<-bw.nrd0(x)
  BW[i,3]<-bw.nrd(x)
  BW[i,4]<-bw.ucv(x)
  BW[i,5]<-bw.bcv(x)
  BW[i,6]<-bw.SJ(x)
  BW[i,7]<-dpik(x)
  BW[i,8]<-bw.dboot2(x,sd(x))
  BW[i,9]<-(BW[i,4]*(BW[i,7]^2))^(1/3)
  BW[i,10]<-(BW[i,7]*(BW[i,4]^2))^(1/3)
  BW[i,11]<-(BW[i,7]*BW[i,4])^(1/2)
  x.grid<-density(x,BW[i,1])$x
  y.grid<-ddoublex(x.grid,0.5,1/8)
  ISE[i,1]<-sum((density(x,BW[i,1])$y-y.grid)^2)
  ISE[i,2]<-sum((density(x,BW[i,2])$y-y.grid)^2)
  ISE[i,3]<-sum((density(x,BW[i,3])$y-y.grid)^2)
  ISE[i,4]<-sum((density(x,BW[i,4])$y-y.grid)^2)
  ISE[i,5]<-sum((density(x,BW[i,5])$y-y.grid)^2)
  ISE[i,6]<-sum((density(x,BW[i,6] )$y-y.grid)^2)
  ISE[i,7]<-sum((density(x,BW[i,7] )$y-y.grid)^2)
  ISE[i,8]<-sum((density(x,BW[i,8] )$y-y.grid)^2)
  ISE[i,9]<-sum((density(x,BW[i,9] )$y-y.grid)^2)
  ISE[i,10]<-sum((density(x,BW[i,10] )$y-y.grid)^2)
  ISE[i,11]<-sum((density(x,BW[i,11] )$y-y.grid)^2)
}

 Bias.mat[k,]<-apply(BW-BW[,1],2,mean)
 boxplot(ISE)
 ISE.mean[k,]<-apply(ISE,2,mean)
 ISE.sd[k,]<-apply(ISE,2,mean)
 ISE.L2[k,]<-apply((ISE-ISE[,1])^2,2,mean)
 ISE.L1[k,]<-apply(abs(ISE-ISE[,1]),2,mean)
}
```

```
 dotplot(t(Bias.mat),auto.key=list(columns = 2),horizontal = FALSE,
  ylab="Bias(h)")
 dotplot(t(ISE.L1),auto.key=list(columns = 2),horizontal = FALSE,
  ylab="L1-distance")
 dotplot(t(ISE.L2),auto.key=list(columns = 2),horizontal = FALSE,
  ylab="L2-distance")

 sGamma<-function(x) dgamma(5*x,(1.5)^2,1.5)
curve(sGamma,-0.2,1.2,ylab=" ",main="Simple Gamma Distribution")
sGamma<-Vectorize(sGamma)

N.sim=250
BW<-matrix(0,nr=N.sim,nc=11)
ISE<-matrix(0,nr=N.sim,nc=11)
Bias.mat<-matrix(0,nr=4,nc=11)
ISE.L2<-matrix(0,nr=4,nc=11)
ISE.L1<-ISE.mean<-ISE.sd<-matrix(0,nr=4,nc=11)
row.names(Bias.mat)<-row.names(ISE.L2)<-row.names(ISE.L1)<-
row.names(ISE.mean)<-row.names(ISE.sd)<-c("25","50","100","200")
colnames(BW)<-colnames(ISE)<-colnames(Bias.mat)<-colnames(ISE.L2)
<-colnames(ISE.L1)<-c("h1","h2","h3","h4","h5","h6","h7",
"h8","h9","h10","h11")

n.vec<-c(25,50,100,200)
for(k in 1:length(n.vec){
for(i in 1:N.sim){
  x<-rgamma(k,(1.5)^2,1.5)
  x<-5*x
  BW[i,1]<-h.amise(x)$h
  BW[i,2]<-bw.nrd0(x)
  BW[i,3]<-bw.nrd(x)
  BW[i,4]<-bw.ucv(x)
  BW[i,5]<-bw.bcv(x)
```

```
  BW[i,6]<-bw.SJ(x)
  BW[i,7]<-dpik(x)
  BW[i,8]<-bw.dboot2(x,sd(x),h0=BW[i,2])
  BW[i,9]<-(BW[i,4]*(BW[i,7]^2))^(1/3)
  BW[i,10]<-(BW[i,7]*(BW[i,4]^2))^(1/3)
  BW[i,11]<-(BW[i,7]*BW[i,4])^(1/2)
  x.grid<-density(x,BW[i,1])$x
  y.grid<-sGamma(x.grid)
  ISE[i,1]<-sum((density(x,BW[i,1])$y-y.grid)^2)
  ISE[i,2]<-sum((density(x,BW[i,2])$y-y.grid)^2)
  ISE[i,3]<-sum((density(x,BW[i,3])$y-y.grid)^2)
  ISE[i,4]<-sum((density(x,BW[i,4])$y-y.grid)^2)
  ISE[i,5]<-sum((density(x,BW[i,5])$y-y.grid)^2)
  ISE[i,6]<-sum((density(x,BW[i,6] )$y-y.grid)^2)
  ISE[i,7]<-sum((density(x,BW[i,7] )$y-y.grid)^2)
  ISE[i,8]<-sum((density(x,BW[i,8] )$y-y.grid)^2)
  ISE[i,9]<-sum((density(x,BW[i,9] )$y-y.grid)^2)
  ISE[i,10]<-sum((density(x,BW[i,10] )$y-y.grid)^2)
  ISE[i,11]<-sum((density(x,BW[i,11] )$y-y.grid)^2)
}

Bias.mat[k,]<-apply(BW-BW[,1],2,mean)
boxplot(ISE)
ISE.mean[k,]<-apply(ISE,2,mean)
ISE.sd[k,]<-apply(ISE,2,mean)
ISE.L2[k,]<-apply((ISE-ISE[,1])^2,2,mean)
ISE.L1[k,]<-apply(abs(ISE-ISE[,1]),2,mean)
}

dotplot(t(Bias.mat),auto.key=list(columns = 2),horizontal = FALSE,
ylab="Bias(h)")
dotplot(t(ISE.L1),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L1-distance")
dotplot(t(ISE.L2),auto.key=list(columns = 2),horizontal = FALSE,
```

```
ylab="L2-distance")

mGamma<-function(x) (1/3)*dgamma(8*x,(1.5)^2,1.5)
+(1/3)*dgamma(8*x,3^2,3)+(1/3)*dgamma(8*x,6^2,6)
curve(mGamma,-0.2,1.2,ylab=" ",main="Three Gamma Distribution")
mGamma<-Vectorize(mGamma)

N.sim=250
BW<-matrix(0,nr=N.sim,nc=11)
ISE<-matrix(0,nr=N.sim,nc=11)
Bias.mat<-matrix(0,nr=4,nc=11)
ISE.L2<-matrix(0,nr=4,nc=11)
ISE.L1<-ISE.mean<-ISE.sd<-matrix(0,nr=4,nc=11)
row.names(Bias.mat)<-row.names(ISE.L2)<-row.names(ISE.L1)
<-row.names(ISE.mean)<-row.names(ISE.sd)<-c("25","50","100","200")
colnames(BW)<-colnames(ISE)<-colnames(Bias.mat)<-colnames(ISE.L2)
<-colnames(ISE.L1)<-c("h1","h2","h3","h4","h5","h6","h7",
"h8","h9","h10","h11")


n.vec<-c(25,50,100,200)
for(k in 1:length(n.vec){
for(i in 1:N.sim){
  y<-rbinom(k,2,1/3)
  x<-numeric(k)
  for(j in 1:k){
    if(y==0) x[j]<-rgamma(1,(1.5)^2,1.5)
    else if(y==1)   x[j]<-rgamma(1,3^2,3)
    else x[j]<-rgamma(1,6^2,6)
  }
  x<-8*x
  BW[i,1]<-h.amise(x)$h
  BW[i,2]<-bw.nrd0(x)
  BW[i,3]<-bw.nrd(x)
```

```
  BW[i,4]<-bw.ucv(x)
  BW[i,5]<-bw.bcv(x)
  BW[i,6]<-bw.SJ(x)
  BW[i,7]<-dpik(x)
  BW[i,8]<-bw.dboot2(x,sd(x),h0=BW[i,2])
  BW[i,9]<-(BW[i,4]*(BW[i,7]^2))^(1/3)
  BW[i,10]<-(BW[i,7]*(BW[i,4]^2))^(1/3)
  BW[i,11]<-(BW[i,7]*BW[i,4])^(1/2)
  x.grid<-density(x,BW[i,1])$x
  y.grid<-mGamma(x.grid)
  ISE[i,1]<-sum((density(x,BW[i,1])$y-y.grid)^2)
  ISE[i,2]<-sum((density(x,BW[i,2])$y-y.grid)^2)
  ISE[i,3]<-sum((density(x,BW[i,3])$y-y.grid)^2)
  ISE[i,4]<-sum((density(x,BW[i,4])$y-y.grid)^2)
  ISE[i,5]<-sum((density(x,BW[i,5])$y-y.grid)^2)
  ISE[i,6]<-sum((density(x,BW[i,6] )$y-y.grid)^2)
  ISE[i,7]<-sum((density(x,BW[i,7] )$y-y.grid)^2)
  ISE[i,8]<-sum((density(x,BW[i,8] )$y-y.grid)^2)
  ISE[i,9]<-sum((density(x,BW[i,9] )$y-y.grid)^2)
  ISE[i,10]<-sum((density(x,BW[i,10] )$y-y.grid)^2)
  ISE[i,11]<-sum((density(x,BW[i,11] )$y-y.grid)^2)
}
Bias.mat[k,]<-apply(BW-BW[,1],2,mean)
boxplot(ISE)
ISE.mean[k,]<-apply(ISE,2,mean)
ISE.sd[k,]<-apply(ISE,2,mean)
ISE.L2[k,]<-apply((ISE-ISE[,1])^2,2,mean)
ISE.L1[k,]<-apply(abs(ISE-ISE[,1]),2,mean)
}
dotplot(t(Bias.mat),auto.key=list(columns = 2),horizontal = FALSE,
ylab="Bias(h)")
dotplot(t(ISE.L1),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L1-distance")
dotplot(t(ISE.L2),auto.key=list(columns = 2),horizontal = FALSE,
```

```
ylab="L2-distance")

sNorm<-function(x) dnorm(x,mean=0.5,sd=0.2)
curve(sNorm,-0.2,1.2,ylab = " ",main="Simple Normal Distribution")
sNorm<-Vectorize(sNorm)


N.sim=250
BW<-matrix(0,nr=N.sim,nc=11)
ISE<-matrix(0,nr=N.sim,nc=11)
Bias.mat<-matrix(0,nr=4,nc=11)
ISE.L2<-matrix(0,nr=4,nc=11)
ISE.L1<-ISE.mean<-ISE.sd<-matrix(0,nr=4,nc=11)
row.names(Bias.mat)<-row.names(ISE.L2)<-row.names(ISE.L1)
<-row.names(ISE.mean)<-row.names(ISE.sd)<-c("25","50","100","200")
colnames(BW)<-colnames(ISE)<-colnames(Bias.mat)<-colnames(ISE.L2)
<-colnames(ISE.L1)<-c("h1","h2","h3","h4","h5","h6","h7",
"h8","h9","h10","h11")


n.vec<-c(25,50,100,200)
for(k in 1:length(n.vec){
for(i in 1:N.sim){
  x<-rnorm(k,mean=0.5,sd=0.2)
  BW[i,1]<-h.amise(x)$h
  BW[i,2]<-bw.nrd0(x)
  BW[i,3]<-bw.nrd(x)
  BW[i,4]<-bw.ucv(x)
  BW[i,5]<-bw.bcv(x)
  BW[i,6]<-bw.SJ(x)
  BW[i,7]<-dpik(x)
  BW[i,8]<-bw.dboot2(x,sd(x),h0=BW[i,2])
  BW[i,9]<-(BW[i,4]*(BW[i,7]^2))^(1/3)
  BW[i,10]<-(BW[i,7]*(BW[i,4]^2))^(1/3)
```

```
  BW[i,11]<-(BW[i,7]*BW[i,4])^(1/2)
  x.grid<-density(x,BW[i,1])$x
  y.grid<-sNorm(x.grid)
  ISE[i,1]<-sum((density(x,BW[i,1])$y-y.grid)^2)
  ISE[i,2]<-sum((density(x,BW[i,2])$y-y.grid)^2)
  ISE[i,3]<-sum((density(x,BW[i,3])$y-y.grid)^2)
  ISE[i,4]<-sum((density(x,BW[i,4])$y-y.grid)^2)
  ISE[i,5]<-sum((density(x,BW[i,5])$y-y.grid)^2)
  ISE[i,6]<-sum((density(x,BW[i,6] )$y-y.grid)^2)
  ISE[i,7]<-sum((density(x,BW[i,7] )$y-y.grid)^2)
  ISE[i,8]<-sum((density(x,BW[i,8] )$y-y.grid)^2)
  ISE[i,9]<-sum((density(x,BW[i,9] )$y-y.grid)^2)
  ISE[i,10]<-sum((density(x,BW[i,10] )$y-y.grid)^2)
  ISE[i,11]<-sum((density(x,BW[i,11] )$y-y.grid)^2)
}
Bias.mat[k,]<-apply(BW-BW[,1],2,mean)
boxplot(ISE)
ISE.mean[k,]<-apply(ISE,2,mean)
ISE.sd[k,]<-apply(ISE,2,mean)
ISE.L2[k,]<-apply((ISE-ISE[,1])^2,2,mean)
ISE.L1[k,]<-apply(abs(ISE-ISE[,1]),2,mean)
}

dotplot(t(Bias.mat),auto.key=list(columns = 2),horizontal = FALSE,
ylab="Bias(h)")
dotplot(t(ISE.L1),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L1-distance")
dotplot(t(ISE.L2),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L2-distance")

m2Norm<-function(x) (1/2)*dnorm(x,0.35,0.1)+(1/2)*dnorm(x,0.65,0.1)
curve(m2Norm,-0.2,1.2,ylab=" ",main="Two Normal Distribution")
m2Norm<-Vectorize(m2Norm)
```

```
N.sim=250
BW<-matrix(0,nr=N.sim,nc=11)
ISE<-matrix(0,nr=N.sim,nc=11)
Bias.mat<-matrix(0,nr=4,nc=11)
ISE.L2<-matrix(0,nr=4,nc=11)
ISE.L1<-ISE.mean<-ISE.sd<-matrix(0,nr=4,nc=11)
row.names(Bias.mat)<-row.names(ISE.L2)<-row.names(ISE.L1)
<-row.names(ISE.mean)<-row.names(ISE.sd)<-c("25","50","100","200")
colnames(BW)<-colnames(ISE)<-colnames(Bias.mat)<-colnames(ISE.L2)
<-colnames(ISE.L1)<-c("h1","h2","h3","h4","h5","h6","h7",
"h8","h9","h10","h11")

n.vec<-c(25,50,100,200)
for(k in 1:length(n.vec){
for(i in 1:N.sim){
  y<-rbinom(k,1,1/2)
  x<-numeric(k)
  for(j in 1:k){
    if(y==0) x[j]<-rnorm(x,0.35,0.1)
    else x[j]<-rnorm(x,0.65,0.1)
  }
  BW[i,1]<-h.amise(x)$h
  BW[i,2]<-bw.nrd0(x)
  BW[i,3]<-bw.nrd(x)
  BW[i,4]<-bw.ucv(x)
  BW[i,5]<-bw.bcv(x)
  BW[i,6]<-bw.SJ(x)
  BW[i,7]<-dpik(x)
  BW[i,8]<-bw.dboot2(x,sd(x),h0=BW[i,2])
  BW[i,9]<-(BW[i,4]*(BW[i,7]^2))^(1/3)
  BW[i,10]<-(BW[i,7]*(BW[i,4]^2))^(1/3)
  BW[i,11]<-(BW[i,7]*BW[i,4])^(1/2)
  x.grid<-density(x,BW[i,1])$x
  y.grid<-m2Norm(x.grid)
```

```
  ISE[i,1]<-sum((density(x,BW[i,1])$y-y.grid)^2)
  ISE[i,2]<-sum((density(x,BW[i,2])$y-y.grid)^2)
  ISE[i,3]<-sum((density(x,BW[i,3])$y-y.grid)^2)
  ISE[i,4]<-sum((density(x,BW[i,4])$y-y.grid)^2)
  ISE[i,5]<-sum((density(x,BW[i,5])$y-y.grid)^2)
  ISE[i,6]<-sum((density(x,BW[i,6])$y-y.grid)^2)
  ISE[i,7]<-sum((density(x,BW[i,7])$y-y.grid)^2)
  ISE[i,8]<-sum((density(x,BW[i,8])$y-y.grid)^2)
  ISE[i,9]<-sum((density(x,BW[i,9])$y-y.grid)^2)
  ISE[i,10]<-sum((density(x,BW[i,10])$y-y.grid)^2)
  ISE[i,11]<-sum((density(x,BW[i,11])$y-y.grid)^2)
}
Bias.mat[k,]<-apply(BW-BW[,1],2,mean)
boxplot(ISE)
ISE.mean[k,]<-apply(ISE,2,mean)
ISE.sd[k,]<-apply(ISE,2,mean)
ISE.L2[k,]<-apply((ISE-ISE[,1])^2,2,mean)
ISE.L1[k,]<-apply(abs(ISE-ISE[,1]),2,mean)
}
dotplot(t(Bias.mat),auto.key=list(columns = 2),horizontal = FALSE,
ylab="Bias(h)")
dotplot(t(ISE.L1),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L1-distance")
dotplot(t(ISE.L2),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L2-distance")

m3Norm<-function(x) (1/3)*dnorm(x,0.25,0.075)
+(1/3)*dnorm(x,0.5,0.075)+(1/3)*dnorm(x,0.75,0.075)
curve(m3Norm,-0.2,1.2,ylab=" ",main="Three Normal Distribution")
m3Norm<-Vectorize(m3Norm)



N.sim=250
BW<-matrix(0,nr=N.sim,nc=11)
```

```
ISE<-matrix(0,nr=N.sim,nc=11)
Bias.mat<-matrix(0,nr=4,nc=11)
ISE.L2<-matrix(0,nr=4,nc=11)
ISE.L1<-ISE.mean<-ISE.sd<-matrix(0,nr=4,nc=11)
row.names(Bias.mat)<-row.names(ISE.L2)<-row.names(ISE.L1)
<-row.names(ISE.mean)<-row.names(ISE.sd)<-c("25","50","100","200")
colnames(BW)<-colnames(ISE)<-colnames(Bias.mat)<-colnames(ISE.L2)
<-colnames(ISE.L1)<-c("h1","h2","h3","h4","h5","h6","h7",
"h8","h9","h10","h11")

n.vec<-c(25,50,100,200)
for(k in 1:length(n.vec){
for(i in 1:N.sim){
  y<-rbinom(k,2,1/3)
  x<-numeric(k)
  for(j in 1:k){
    if(y==0) x[j]<-rnorm(x,0.25,0.075)
    else if(y==1)   x[j]<-rnorm(x,0.5,0.075)
    else x[j]<-rnorm(x,0.75,0.075)
  }
  BW[i,1]<-h.amise(x)$h
  BW[i,2]<-bw.nrd0(x)
  BW[i,3]<-bw.nrd(x)
  BW[i,4]<-bw.ucv(x)
  BW[i,5]<-bw.bcv(x)
  BW[i,6]<-bw.SJ(x)
  BW[i,7]<-dpik(x)
  BW[i,8]<-bw.dboot2(x,sd(x),h0=BW[i,2])
  BW[i,9]<-(BW[i,4]*(BW[i,7]^2))^(1/3)
  BW[i,10]<-(BW[i,7]*(BW[i,4]^2))^(1/3)
  BW[i,11]<-(BW[i,7]*BW[i,4])^(1/2)
  x.grid<-density(x,BW[i,1])$x
  y.grid<-m3Norm(x.grid)
  ISE[i,1]<-sum((density(x,BW[i,1])$y-y.grid)^2)
```

```
  ISE[i,2]<-sum((density(x,BW[i,2])$y-y.grid)^2)
  ISE[i,3]<-sum((density(x,BW[i,3])$y-y.grid)^2)
  ISE[i,4]<-sum((density(x,BW[i,4])$y-y.grid)^2)
  ISE[i,5]<-sum((density(x,BW[i,5])$y-y.grid)^2)
  ISE[i,6]<-sum((density(x,BW[i,6] )$y-y.grid)^2)
  ISE[i,7]<-sum((density(x,BW[i,7] )$y-y.grid)^2)
  ISE[i,8]<-sum((density(x,BW[i,8] )$y-y.grid)^2)
  ISE[i,9]<-sum((density(x,BW[i,9] )$y-y.grid)^2)
  ISE[i,10]<-sum((density(x,BW[i,10] )$y-y.grid)^2)
  ISE[i,11]<-sum((density(x,BW[i,11] )$y-y.grid)^2)
}
Bias.mat[k,]<-apply(BW-BW[,1],2,mean)
boxplot(ISE)
ISE.mean[k,]<-apply(ISE,2,mean)
ISE.sd[k,]<-apply(ISE,2,mean)
ISE.L2[k,]<-apply((ISE-ISE[,1])^2,2,mean)
ISE.L1[k,]<-apply(abs(ISE-ISE[,1]),2,mean)
}
dotplot(t(Bias.mat),auto.key=list(columns = 2),horizontal = FALSE,
ylab="Bias(h)")
dotplot(t(ISE.L1),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L1-distance")
dotplot(t(ISE.L2),auto.key=list(columns = 2),horizontal = FALSE,
ylab="L2-distance")
```

داده‌های ستاره‌شناسی فصل ۴

```
data<-read.table("C:/Users/nhp/Downloads/asu-txt.txt", quote="\"")
x<-data[,4]
x<-x[30000:30300]
hist(x,breaks=100,freq=FALSE,main=" ")
hist(x,freq=FALSE,main=" ")
h.amise(x)$h->h1
lines(density(x,h1),col="red")
bw.nrd0(x)->h2
hist(x,freq=FALSE,main=" ",ylim=c(0,0.025))
```

```
lines(density(x,h2),col="red")
bw.nrd(x)->h3
hist(x,freq=FALSE,main=" ",ylim=c(0,0.025))
lines(density(x,h3),col="red")
bw.ucv(x)->h4
hist(x,freq=FALSE,main=" ",ylim=c(0,0.025))
lines(density(x,h4),col="red")
bw.bcv(x)->h5
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
bw.SJ(x)->h6
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
dpik(x)->h7
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
bw.dboot2(x,sd(x),h0=40)->h8
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
(bw.ucv(x)*(dpik(x)^2))^(1/3)->h9
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
(dpik(x)*(bw.ucv(x)^2))^(1/3)->h10
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
(dpik(x)*bw.ucv(x))^(1/2)->h11
hist(x,freq=FALSE,main=" ",ylim=c(0,0.030))
lines(density(x,h5),col="red")
##Biases
h2-h1
h3-h1
h4-h1
h5-h1
h6-h1
```

```
h7-h1
h8-h1
h9-h1
h10-h1
h11-h1
#ISEs
sum(density(x,h2)$y-density(x,h1)$y)^2
sum(density(x,h3)$y-density(x,h1)$y)^2
sum(density(x,h4)$y-density(x,h1)$y)^2
sum(density(x,h5)$y-density(x,h1)$y)^2
sum(density(x,h6)$y-density(x,h1)$y)^2
sum(density(x,h7)$y-density(x,h1)$y)^2
sum(density(x,h8)$y-density(x,h1)$y)^2
sum(density(x,h9)$y-density(x,h1)$y)^2
sum(density(x,h10)$y-density(x,h1)$y)^2
sum(density(x,h11)$y-density(x,h1)$y)^2
```

مراجع

[1] Ahmad, I. A. and Ran, I. S. (2004) Data based bandwidth selection in kernel density estimation with paramteric start via kernel contrasts. *J. Nonparametr. Stat.*, **16**, 841–-877.

[2] Apostel. T. M. (1974). *Mathematical Analysis*, (2nd Edition), Pearson.

[3] Bean, S .J. and Tsokos, C. P. (1980) Developments in nonparametric density estimation. *Int. Stat. Rev.*, **48**, 267–-287.

[4] Bowman, A. (1984) An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353-–360.

[5] Cao, R. (1993) Bootstrapping the mean integrated squared error. *J. Mult. Anal.*, **45**, 137–160.

[6] Cao, R., Cuevas, A., and Gonzlez Manteiga, W. (1994) A comparative study of several smoothing methods in density estimation. *Comp. Stat. Data Anal.*, **17**, 153–176.

[7] Chacon, J. E., Montanero, J., and Nogales, A.G. (2008) Bootstrap bandwidth selection using an h-dependent pilot bandwidth. *Scand. J. Stat.*, **35**, 139-–157.

[8] Chiu, S. T. (1991a) Some stabilized bandwidth selectors for nonparametric regression. *Ann. Stat.*, **19**, 1528–1546.

[9] Chiu, S. T. (1991b) Bandwidth selection for kernel density estimation. *Ann. Stat.*, **19**, 1883–1905.

[10] Chiu, S. T. (1992) An automatic bandwidth selector for kernel density estimation. *Biometrika*, **79**, 771-–782.

[11] Chiu, S. T. (1996) A comparative review of bandwidth selection for kernel density estimation. *Stat. Sin.*, **6**, 129–145.

[12] Cline, D. B. H. (1988). Admissible kernel estimators of a multivariate density, *Ann. Stat.*, **83**, 596–610.

[13] Devroye, L. and Lugosi, G. (1996) A universal acceptable smoothing factor for kernel density estimation, *Ann. Stat.*, **24**, 2499–2512.

[14] Duin, R. P. W. (1976) On the choice of smoothing parameters of Parzen estimators of probability density functions. *IEEE Trans. Comp.*, **25**, 1175–-1179.

[15] Epanechnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density, *Teor. Veroyatnost. i Primenen.*, **14**(1), 156–161.

[16] Faraway, J. J. and Jhun, M. (1990) Bootstrap choice of bandwidth for density estimation. *J. Am. Stat. Assoc.*, **85**, 1119–1122.

[17] Feluch, W. and Koronacki, J. (1992) A note on modified cross-validation in density estimation. *Comp. Stat. Data Anal.*, **13**, 143-–151.

[18] Fryer, M. J. (1977) A review of some non-parametric methods of density estimation. *J. Appl. Math.*, **20**(3), 335–-354.

[19] Grund, B. and Polzehl, J. (1997). Bias corrected bootstrap bandwidth selection. *J. Nonparametric Stat.*, **8**, 97–126.

[20] Habbema, J. D. F., Hermans, J., and van den Broek, K. (1974) Astepwise discrimination analysis program using density estimation, In: Bruckman, G. (Ed.) COMPSTAT '74. Proceedings in Computational Statistics, pp. 101–110. Physica, Vienna.

[21] Hall, P. (1990) Using the bootstrap to estimate mean square error and select smoothing parameters in nonparametric problems. *J. Mult. Anal.*, **32**, 177–203.

[22] Hall, P. and Johnstone, I. (1992) Empirical functionals and efficient smoothing parameter selection. *J. R. Stat. Soc. Ser. B*, **54**, 475–530.

[23] Hall, P. and Marron, J. S. (1987) Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Prob. Theo. Rel. Fields*, **74**, 567–581.

[24] Hall, P., Marron, J. S. (1991) Lower bounds for bandwidth selection in density estimation. *Prob. Theo. Rel. Fields*, **90**, 149–173.

[25] Hall, P., Marron, J. S., and Park, B. U. (1992) Smoothed cross-validation. *Prob. Theo. Rel. Fields*, **92**, 1–20.

[26] Hall, P., Sheater, S. J., Jones, M.C., and Marron, J. S. (1991) On optimal databased bandwidth selection in kernel density estimation. *Biometrika*, **78**, 263–-269.

[27] Hardle, W., Muller, M., Sperlich, S., and Werwatz, A. (2004) *Nonparametric and Semiparametric Models*, Springer Series in Statistics, Berlin.

[28] Hardle, W. and Vieu, P. (1992) Kernel regression smoothing of time series. *J. Time Ser. Anal.*, **13**, 209–232.

[29] Hastei, T., Tibshirani, R., and Freidman, J. (2009) *The elements of statistical learning*, Springer-Verlag.

[30] Heidenrich, N. B., Schindler, A. and Sperlich, S. (2013) Bandwidth slection for kernel density estimation: a review of fully automatic selectors, *Adv. Stat. Anal.*, **97**(4), 403–433.

[31] Hodges, J. L. and Lehmann, E. L. (1956) The efficiency of some nonparametric competitors to the t-test, *Ann. Math. Statist.*, **13**, 324–35.

[32] Jones, M. C., Marron, J. S., and Park, B. U. (1991) A simple root n bandwidth selector. *Ann. Stat.*, **19**, 1919–1932.

[33] Jones, M. C., Marron, J. S., and Sheather, S. J. (1996a) A brief survey of bandwidth selection for density estimation. *J. Am. Stat. Assoc.*, **91**, 401–407.

[34] Jones, M. C., Marron, J. S., and Sheather, S. J. (1996b) Progress in data-based bandwidth selection for kernel density estimation. *Comp. Stat.*, **11**, 337-–381.

[35] Loader, C. R. (1999) Bandwidth selection: classical or plug-in? *Ann. Stat.*, **27**(2), 415-–438.

[36] Mammen, E., Martínez-Miranda, M. D., Nielsen, J. P., and Sperlich, S. (2011) Do-validation for kernel density estimation. *J. Am. Stat. Assoc.*, **106**, 651-–660.

[37] Marron, J. S. (1988) Automatic smoothing parameter selection: a survey. *Empir. Econ.*, **13**, 187–-208.

[38] Marron, J. S. (1992) Bootstrap bandwidth selection. In: LePage, R., Billard, L. (eds.) Exploring the Limits of Bootstrap, pp. 249–262. Wiley, New York.

[39] Marron, J. S. (1994). Visual understanding of higher order kernels. *J. Comp. Graph. Stat.*, **3**, 447–458.

[40] Marron, J. S. and Nolan, D. (1988) Canonical kernels for density estimation. *Stat. Prob. Lett.*, **7**, 195–199.

[41] Martinez-Miranda, M. D., Nielsen, J.,and Sperlich, S. (2009) One sided cross validation in density estimation. In: Gregoriou, G.N. (ed.) Operational Risk Towards Basel III: Best Practices and Issues in Modeling, Management and Regulation. Wiley, Hoboken.

[42] Nadaraya, E. A. (1974). On the integral mean square error of some nonparametric estimates for the density function, *Theo Probab Appl.*, **19**, 133–141.

[43] Park, B. U. and Marron, J. S. (1990) Comparison of data-driven bandwidth selectors. *J. Am. Stat. Assoc.*, **85**, 66-–72.

[44] Park, B. U. and Turlach, B. A. (1992) Practical performance of several data driven bandwidth selectors, CORE Discussion, Paper 9205.

[45] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode, *Ann. Math. Stat.*, **33**(3), 1065–1076.

[46] Rigollet, P. and Tsybakov, A. (2007) Linear and convex aggregation of density estimators, *Math. Meth. Stat.*, **16**, 260–280.

[47] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function, *Ann. Math. Stat.*, **27**(3), 832–837.

[48] Rudemo, M. (1982) Empirical choice of histograms and kernel density estimators, *Scand. J. Stat.*, **9**, 65–-78.

[49] Rudin , W. (1976). *Principle of Matematical Analysis* (3rd Edition), McGrow-Hill Education.

[50] Ruppert, D. and Cline, B. H. (1994) Bias Reduction in kernel density estimation by smoothed empirical transformations, *Ann. Stat.*, **22**, 185-–210.

[51] Samarov, A. and Tsybakov, A. (2007) Aggregation of density estimators and dimension reduction. In: Nair, V. (ed.) Advances in Statistical Modeling and Inference: essays in honor of Kjell A. Doksum, pp. 233–251.

[52] Savchuk, O. J., Hart, J. D.,and Sheather, S. J. (2010) Indirect cross-validation for density estimation. *J. Am. Stat. Assoc.*, **105**, 415-–423.

[53] Scott, D. W. (1985) Averaged shifted histograms: effective nonparametric density estimators in several dimensions, *Ann. Stat.*, **13**(2), 1024–1040.

[54] Scott, D. W. and Terrell, G. R. (1987) Biased and unbiased cross-validation in density estimation. *J. Am. Stat. Assoc.*, **82**, 1131-–1146.

[55] Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection-method for kernel density estimation. *J. R. Stat. Soc. Ser. B*, **53**, 683-–690.

[56] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[57] Stone, C. J. (1984) An asymptotically optimal window selection rule for kernel density estimates. *Ann. Stat.*, **12**, 1285–1297.

[58] Stute, W. (1992) Modified cross validation in density estimation. *J. Stat. Plan. Inf.*, **30**, 293–305.

[59] Tartar, M. E. and Kronmal, R. A. (1976) An introduction to the implementation and theory of nonparametric density estimation. *Am. Stat.*, **30**, 105-–112.

[60] Taylor, C. C. (1989) Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, **76**, 705–712.

[61] Terrell, G. R. (1990), The Maximal Smoothing Principle in Density Estimation, *J. Amer. Stat. Assoc.*, **85**, 470–477.

[62] Torossett, M. (2011) *An Introduction to Statistical Inference and its Applications with R*, Chapmann & Hall/CRS.

[63] Ullah, A. (1985), Specification Analysis of Econometric Models, *J. Quantitative Econ.*, **2**, 187–209.

[64] van Vliet, P.K. and Gupta, J.M. (1973) Tham-V- sodium bicarbonate in idiopathic respiratory distress syndrome. *Arch. Dis. Child.*, **48**, 249-55.

[65] Wand, M. P. and Jones, M. C. (1994). *Kernel Smoothing*, Capman and Hall, London.

[66] Wand, M. P., Marron, J. S., and Ruppert, D. (1991) Transformations in density estimation. *J. Am. Stat. Assoc.*, **86**, 343–-353.

[67] Watson, G. S. (1969). Density Estimation by Orthogonal Series, *Ann. Math. Stat.*, **40**, 1496–1498.

[68] Wegkamp, M. H., (1999). Quasi universal bandwidth selection for kernel density estimators. *Can. J. Stat.*, **27**, 409–420.

[69] Wegman, E .J. (1972) Nonparametric probability density estimation: I. A summary of available methods. *Technometrics*, **14**, 533-–546 .

[70] Weisberg, S. (1980) *Applied linear regression*, John Wiley & Sons, New York.

[71] Wertz, W. and Schneider, B. (1979) Statistical density estimation: a bibliography. *Int. Stat. Rev.*, **47**, 155-–175.

[72] Yang, Y. (2000) Mixing strategies for density estimation. *Ann. Stat.*, **28**, 75–87.

[73] Yang, L., Marron, S. (1999) Iterated transformation-kernel density estimation. *J. Am. Stat. Assoc.*, **94**, 580–-589.

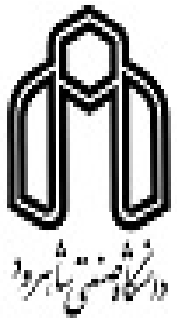
Aabstract

Perhaps the most fundamental concepts in the field of statistic is the porobability function which can be found in the behavior of random variables especially in statistics without knowing the statistical probability of a random sample, ability to perform inference on random samples and generalizing the results to the community does not exist.

On of the methods for estimating density (probability) function is "kernel density estimation", the generalization of naive estimator. In kernel methods, we must specify two parameters: Kernel function and bandwidth. The bandwidth selection plays a key role in kernel methods. When bandwith is small, the curve is nonsmooth. With increasing the bandwith, the smoothness increases. Some of the methods for bandwidth selection is: subjective methds, normal scale rule, least square cross validation etc.

In this disseration, we study density estimation via kernel methods, also, compare some of the bandwidth selection methods and apply them to astrostatistics data.

Keywords: Kernel function, Bandwidth, Cross validation, astrostatistics, Integrated squerred error measure.



Shahrood Univercity of Technology

School Of Mathematical Sciences

Studying some bandwidth methods in kernel estimation for astrostatistics data

Mahboobeh Mirzaei

Supervisor

Prof. M. Arashi

December 2015