



دانشگاه فردوسی مشهد

دانشکده ریاضی و آمار

پایان نامه جهت اخذ درجه کارشناسی ارشد آمار اقتصادی و اجتماعی

عنوان:

معرفی روشی ساده و مؤثر برای اصلاح کران در  
برآورد ناپارامتری تابع چگالی و رگرسیون هسته

استاد راهنما:

دکتر مهدی عمادی

استاد مشاور:

دکتر وحید فکور

نگارش:

الهه آخوندزاده کاشانی

خرداد ماه ۹۱

# فهرست مطالب

فهرست علائم و اختصارات ۵

پیشگفتار ۸

۱ مقدمه ای بر برآوردگر ناپارامتری تابع چگالی ۱۰

۱.۱ برآوردگر هیستوگرام ..... ۱۰

۲.۱ برآوردگر هسته ای ..... ۱۲

۳.۱ اثر پهنای باند ..... ۱۶

۴.۱ اثر توابع هسته ..... ۱۷

۵.۱ نمایش گرافیکی برآورد هسته ای چگالی ..... ۱۹

۶.۱ معیارهای ارزیابی ..... ۲۰

۱.۶.۱ آریبی ..... ۲۱

۲.۶.۱ واریانس ..... ۲۲

|    |  |       |
|----|--|-------|
| ۲۳ | ..... میانگین مربع خطا                       | ۳.۶.۱ |
| ۲۶ | ..... میانگین انتگرال مربع خطا               | ۴.۶.۱ |
| ۲۷ | ..... انتخاب پارامتر هموارساز                | ۷.۱   |
| ۲۸ | ..... قاعده تخمینی سیلورمن                   | ۱.۷.۱ |
| ۳۱ | ..... روش جایگذاری تصحیح شده                 | ۲.۷.۱ |
| ۳۲ | ..... انتخاب هسته                            | ۸.۱   |
| ۳۵ | ..... تعدیل پهنای باند برای هسته های متفاوت. | ۹.۱   |
| ۳۶ | ..... بهینه سازی هسته                        | ۱۰.۱  |
| ۳۷ | ..... فواصل اطمینان و نوارهای اطمینان        | ۱۱.۱  |
| ۳۹ | ..... برآورد هسته ای چگالی چند متغیره        | ۱۲.۱  |

## ۲ رگرسیون ناپارامتری ۴۲

|    |                                  |       |
|----|----------------------------------|-------|
| ۴۲ | ..... رگرسیون هسته ای یک متغیره  | ۱.۲   |
| ۴۵ | ..... امید شرطی                  | ۲.۲   |
| ۴۶ | ..... طرح های تصادفی و تثبیت شده | ۳.۲   |
| ۴۷ | ..... رگرسیون هسته ای            | ۴.۲   |
| ۴۷ | ..... طرح تصادفی                 | ۱.۴.۲ |
| ۵۰ | ..... طرح تثبیت شده              | ۲.۴.۲ |
| ۵۱ | ..... ویژگی های آماری            | ۵.۲   |

## ۳ روش های اصلاح کران ۵۴

|    |     |                          |
|----|-----|--------------------------|
| ۵۴ | ۱.۳ | اثرات کران               |
| ۶۱ | ۲.۳ | روش جونز                 |
| ۶۲ | ۳.۳ | روش تبدیل داده ها        |
| ۶۳ | ۴.۳ | روش انعکاس داده ها       |
| ۶۳ | ۵.۳ | روش کارونامونی و آلبرت   |
| ۶۴ | ۶.۳ | روش شبه داده             |
| ۶۵ | ۷.۳ | روش هسته های کرانی       |
| ۶۶ | ۸.۳ | روش پهنای باند محلی      |
| ۶۷ | ۹.۳ | روش پهنای باند تعدیل شده |

#### ۴ مطالعه شبیه سازی ۷۱

|    |       |  |
|----|-------|--|
| ۷۱ | ۱.۴   | شبیه سازی داده ها                      |
| ۷۲ | ۲.۴   | برآورد چگالی                           |
| ۷۹ | ۳.۴   | برآورد رگرسیونی                        |
| ۸۳ | ۴.۴   | دو کاربرد روش های اصلاح کران در اقتصاد |
| ۸۴ | ۱.۴.۴ | توزیع درآمد جهانی                      |
| ۸۶ | ۲.۴.۴ | برآورد منحنی انگل                      |
| ۸۸ | ۵.۴   | بحث و نتیجه گیری                       |

#### ۹۰ مراجع

#### ۹۵ واژه نامه

---

|     |         |
|-----|---------|
| ۹۹  | ضمیمه ۱ |
| ۱۰۱ | ضمیمه ۲ |
| ۱۰۳ | ضمیمه ۳ |

# فهرست علائم و اختصارات

$K$  ..... تابع هسته (یک متغیره)

$\kappa$  ..... تابع هسته (چند متغیره)

$h$  ..... پهنای باند

$\|K\|_2^2 = \int \{K(u)\}^2 du$  ..... مربع نرم  $L_2$

$Bias$  ..... اریبی

$Var$  ..... واریانس

$MSE$  ..... میانگین مربع خطا

$MISE$  ..... میانگین انتگرال مربع خطا

$AMISE$  ..... تقریب میانگین انتگرال مربع خطا

$\#$  ..... تعداد اعضای یک مجموعه

$s.t.$  ..... به طوری که

$as$  ..... چنانکه

$o.w$  ..... جای دیگر

$\mu_r(K) = \int u^r K(u) du$  ..... گشتاور دوم  $K$

$I(\cdot)$  ..... تابع نشانگر

$diag(a_1, a_2, \dots, a_n)$  ..... ماتریس قطری با عناصر  $a_1, a_2, \dots, a_n$

$\xrightarrow{P}$  ..... همگرایی در احتمال

$\xrightarrow{L}$  ..... همگرایی در توزیع

# چکیده

در برآورد چگالی و رگرسیون ناپارامتری افزایش واریانس و اریبی در کران ها - که اصطلاحاً اثرات کران نامیده می شوند - به علت اطلاعات یک طرفه ای که از داده ها داریم، می تواند کاملاً مشکل ساز باشد. این مشکل برای برآوردهایی که روی متغیرهای تبدیل یافته اجرا می شود به سادگی زیاد شده و حتی ممکن است برآوردهای نهایی و در نهایت نتایج را اساساً تغییر دهد. در این پایان نامه پس از مرور مختصری در مورد روش های موجود، یک روش اصلاح کران جدید سرراست و بسیار ساده پیشنهاد می شود که از روش تغییر پهنای باند محلی در کران ها استفاده می کند. سپس رفتار آماری آن بررسی شده و کارایی برآورد چگالی و رگرسیون برای حجم نمونه های کوچک و متوسط مطالعه خواهد شد. کارایی این روش در مطالعه ی شبیه سازی بسیار خوب است، همچنین برای برآورد توزیع درآمد جهانی و منحنی انگل در اقتصاد بسیار خوب عمل می کند.



# پیشگفتار

یک مسئله معروف در برآورد ناپارامتری (رگرسیون یا برآورد چگالی) اثرات کران است. در بسیاری از مطالعات کاربردی کران ها مورد توجه خاص هستند، برای مثال در تحلیل فقر، لازم است که برآوردهای قابل اعتمادی از توزیع درآمد در سمت چپ یعنی "نزدیک" به کران طبیعی صفر داشته باشیم، همچنین اگر علاقه مند به بررسی ریسک، فقر و نابرابری باشید، باید نگاه ویژه ای به گروه های جوان یا پیر، درجه ی تحصیلی افراد، شرکت های بزرگ یا کوچک و غیره داشته باشید. به طور مشابه وقتی از رگرسیون ناپارامتری در اقتصادسنجی استفاده می شود، اثرات کران نه تنها در کران ها بلکه در نزدیک آن ها هم دیده می شوند. با توجه به این مثال ها می توان مشاهده کرد که استفاده از روش های اصلاح کران در هر دو مورد برآورد هسته ای چگالی و رگرسیونی ضرورت دارد.

تا کنون روش های متعددی برای اصلاح اثرات کران معرفی شده اند، به طور کلی روش های موجود می تواند به گروه های زیر تقسیم شود:

۱. روش انعکاس داده ها (اسکاتر (۱۹۸۵)، سیلورمن (۱۹۸۶)، کلاین و هارت (۱۹۹۱))
۲. روش تبدیل داده ها (واند و همکاران (۱۹۹۱)، مارون و راپرت (۱۹۹۴))
۳. روش شبه داده (کولینگ و هال (۱۹۹۶))
۴. روش هسته های کرانی (گسر و مولر (۱۹۷۹)، گسر و همکاران (۱۹۸۳)، جونز (۱۹۹۳))
۵. روش پهنای باند محلی (رایس (۱۹۸۴)، مولر (۱۹۹۱)، هال و ورلی (۱۹۹۱))

جالب توجه است که علیرغم اهمیت روش های اصلاح کران در عمل و مقدار قابل توجه (اما نه خیلی زیاد) مطالعات نظری این روش ها هم در برآورد چگالی و هم در رگرسیون به ندرت استفاده می شوند. یک دلیل واضح این امر، فقدان برنامه های اجرایی در نرم افزارهای آماری و اقتصادی است. در کنار فقدان نرم افزار دلیل دیگر برای استفاده ی اندک از این روش ها می تواند پیچیدگی آن در مقایسه با بهبود مشهود ایجاد شده در برآورد نهایی باشد. به همین دلیل در این پایان نامه به منظور اصلاح اثرات کران روشی پیشنهاد شده که ساده است و به ندرت نیاز به محاسبات بیشتری نسبت به روش بدون اصلاح کران دارد.

به طور خلاصه در جستجوی یک روش اصلاح کران ساده و سریع هستیم تا بتواند در نهایت با روش جونز ۱۹۹۳ و چندجمله ای محلی در هر دو مسئله رگرسیون و چگالی رقابت کند. این روش بر پایه ی ایده ی کاهش قابل توجه اریبی به دست آمده است (هال و پارک ۲۰۰۲) را ببینید). هرچند سادگی این روش باعث افزایش جدی واریانس می شود، اما در مجموع، برآوردها در میانگین مربع خطا بهبود می یابند.

در این پایان نامه بررسی نشده است که کدام روش به طور یکنواخت بهترین جرم احتمال را در کران ها و یا نزدیک به آن ها به کار می برد، بلکه روشی ساده و کاربردی معرفی شده است و با استفاده از یک مطالعه جامع شبیه سازی روش پیشنهاد شده با روش های موجود مقایسه شده و دو کاربرد ارائه شده است. فصل اول به معرفی برآورد گر هسته ای چگالی می پردازد. در فصل دوم رگرسیون هسته ای معرفی خواهد شد. فصل سوم مربوط به معرفی روش های اصلاح اثرات کران و پیشنهاد یک روش ساده و مؤثر به منظور کاهش اثرات کران است. در فصل چهارم چند روش اصلاح کران در داده های شبیه سازی شده برای برآورد چگالی و رگرسیون هسته ای مقایسه شده است و همچنین داده های مربوط به توزیع درآمد جهانی برای برآورد چگالی و برآورد منحنی انگل در رگرسیون هسته ای نیز بررسی شده است.

# فصل ۱

## مقدمه ای بر برآوردگر ناپارامتری تابع چگالی

### ۱.۱ برآوردگر هیستوگرام

ساده ترین برآوردگر تابع چگالی برآوردگر هیستوگرام است. فرض کنید نمونه تصادفی  $X_1, X_2, \dots, X_n$  را به حجم  $n$  از توزیعی پیوسته و نامعلوم داریم. می دانیم برای رسم هیستوگرام داده ها به  $j$  طبقه با طول  $h$  - که اصطلاحاً پهنای بانده نامیده می شود - طبقه بندی می شوند. در این صورت هر طبقه توسط بازه  $B_j$  نمایش داده می شود که:

$$B_j = [x_0 + (j - 1)h, x_0 + jh], \quad j \in \mathbb{Z}$$

و  $x_0$  مبدأ اولین طبقه است.

از اینرو می توان برآوردگر هیستوگرام تابع چگالی را به صورت زیر تعریف کرد:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \sum_j I(X_i \in B_j) I(x \in B_j), \quad (1.1)$$

که

$$I(X_i \in B_j) = \begin{cases} 1, & X_i \in B_j \\ 0, & \text{o.w.} \end{cases}$$

اگر مرکز طبقه  $B_j$  با  $m_j$  نمایش داده شود، از فرمول (۱.۱) به سادگی می توان دریافت که برای هر مشاهده  $x$  که در بازه  $B_j = [m_j - \frac{h}{4}, m_j + \frac{h}{4})$  قرار گیرد هیستوگرام برآورد مشابهی برای  $f$  - که با  $\hat{f}_h(m_j)$  نمایش داده می شود - محاسبه می کند. می توان ثابت کرد که

$$\hat{f}_h(m_j) = \frac{1}{nh} \#\{X_i \in [m_j - \frac{h}{4}, m_j + \frac{h}{4}]\}$$

نمایش دهنده ی برآوردگر هیستوگرام تابع چگالی است و  $\#$  تعداد اعضای مجموعه را نمایش می دهد که در فاصله ی دلخواهی از  $f$  تعریف شده باشند. بدون از دست دادن کلیت فرض می کنیم که  $f$  روی فاصله ی  $[0, 1]$  تعریف شده باشد. اگر  $m$  یک عدد صحیح باشد بازه هایی را به صورت زیر تعریف می کنیم:

$$B_1 = \left[0, \frac{1}{m}\right], B_2 = \left[\frac{1}{m}, \frac{2}{m}\right], \dots, B_m = \left[\frac{m-1}{m}, 1\right]$$

همانطور که ملاحظه می شود طول هر یک از بازه ها برابر  $\frac{1}{m}$  است که اصطلاحاً پهناى باندها نامیده می شوند و آن را به طور کلی با نماد  $h$  نمایش می دهیم. فرض کنیم  $Y_j$  تعداد مشاهداتی است که در بازه ی  $B_j$  قرار می گیرند، قرار می دهیم  $P_j = \int_{B_j} f(u) du$  و  $\hat{p}_j = \frac{Y_j}{n}$ ، از اینرو برآوردگر هیستوگرام تابع چگالی عبارت است از:

$$\hat{f}_h(x) = \sum_{j=1}^m \frac{\hat{p}_j}{h} I(x \in B_j).$$

## ۲.۱ برآوردگر هسته ای

می دانیم یک روش برآورد  $f(x)$  محاسبه ی عبارت زیر است:

$$\hat{f}(x) = \frac{1}{n \times \text{طول فاصله}}$$

$$\# \left\{ \text{مشاهداتی که در یک فاصله ی کوچک حول } x \text{ می افتند} \right\} \quad (۲.۱)$$

در اینجا تابع  $f$  در فاصله ی  $[x - h, x + h]$  به طول  $2h$  برآورد شده است. از اینرو می توان نوشت:

$$\hat{f}_h(x) = \frac{1}{2nh} \# \{X_i \in [x - h, x + h]\} \quad (۳.۱)$$

اگر در این عبارت از تابع وزنی (۴.۱) که تابع هسته ای یکنواخت نام دارد استفاده شود، می توان عبارت (۳.۱) را به صورت عبارت (۵.۱) بازنویسی کرد:

$$K(u) = \frac{1}{2} I(|u| \leq 1) \quad (۴.۱)$$

که  $u = (x - X_i)/h$

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (۵.۱)$$

$$= \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} I \left( \left| \frac{x - X_i}{h} \right| \leq 1 \right) \quad (6.1)$$

تابع هسته ای یکنواخت به هر مشاهده  $X_i$  که فاصله آن از  $x$  (نقطه ای که می خواهیم تابع چگالی احتمال را برآورد کنیم) کمتر از  $h$  است وزن  $1/2$  را تخصیص می دهد و مشاهداتی که فاصله آن ها از  $x$  بیشتر از  $h$  است وزن صفر می گیرند زیرا تابع نشانگر  $I(\cdot)$  طبق تعریف برای تمام مقادیر  $u = (x - X_i)/h$  که بزرگتر از ۱ هستند برابر صفر است.

از فرمول (۵.۱) مشهود است که تمام کاری که تاکنون انجام شد فرمولبندی عبارت (۲.۱) بوده است. با توجه به عبارت (۶.۱) می توان دریافت که برای هر مشاهده در فاصله  $[x - h, x + h]$  مقدار تابع نشانگر برابر یک خواهد بود و یک سهم برای مقدار فراوانی آن در نظر گرفته می شود. اما هر سهم به طور مساوی وزن دهی شده و مهم نیست مشاهده  $X_i$  چقدر به  $x$  نزدیک است (مشروط به اینکه در فاصله  $h$  از  $x$  باشد).

شاید بهتر باشد سهم بیشتری از وزن برای مشاهداتی که بسیار نزدیک  $x$  هستند، نسبت به آن هایی که فاصله ی زیادی از آن دارند در نظر گرفت، در این صورت برای مثال فرمول زیر را در نظر بگیرید:

$$\hat{f}_h(x) = \frac{1}{2nh} \sum_{i=1}^n \frac{3}{2} \left\{ 1 - \left( \frac{x - X_i}{h} \right)^2 \right\} I \left( \left| \frac{x - X_i}{h} \right| \leq 1 \right) \quad (7.1)$$

$$= \frac{1}{nh} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right), \quad i = 1, \dots, n \quad (8.1)$$

که  $K$  مختصرنویسی تابع وزنی دیگری به نام هسته ای اپنیچ نیکوف است:

$$K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1).$$

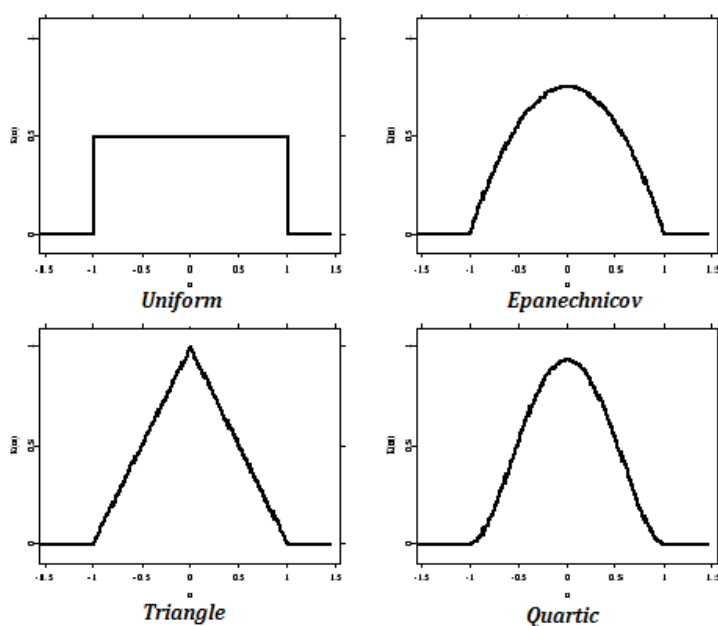
با دقت در عبارت (۷.۱) می توان دریافت که این رویه یک روش ساده برای شمارش تعداد مشاهداتی است که در فاصله حول  $X$  هستند، به طوری که  $X_i$  هایی که به  $X$  نزدیکترند سهم بیشتری از وزن نسبت به مشاهداتی که دورترند دارند.

| <i>Kernel</i>            | هسته                   | $K(u)$   |
|--------------------------|------------------------|--|
| <i>Uniform</i>           | یکنواخت                | $\frac{1}{2}I( u  \leq 1)$                                   |
| <i>Triangle</i>          | مثلث                   | $(1 -  u )I( u  \leq 1)$                                     |
| <i>Epanechnikov</i>      | اپنیچ نیکوف            | $\frac{3}{4}(1 - u^2)I( u  \leq 1)$                          |
| <i>Quartic(Biweight)</i> | درجه ی چهارم (دو وزنی) | $\frac{15}{16}(1 - u^2)^2I( u  \leq 1)$                      |
| <i>Triweight</i>         | سه وزنی                | $\frac{35}{32}(1 - u^2)^3I( u  \leq 1)$                      |
| <i>Gaussian</i>          | گاوسی                  | $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$     |
| <i>Cosine</i>            | کسینوس                 | $\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right)I( u  \leq 1)$ |

جدول ۱.۱: توابع هسته مختلف

ویژگی اخیر را هسته ی اپنیچ نیکوف و بسیاری دیگر از هسته ها که برخی از آن ها در جدول ۱.۱ معرفی شده اند، دارند.

توجه کنید که توابع هسته حول صفر متقارن هستند. در شکل (۱.۱) برخی از توابع هسته نمایش داده شده است.



شکل ۱.۱: نمایش چند تابع هسته: یکنواخت (بالا چپ)، اپنیچ نیکوف (بالا راست)

مثلث (پایین چپ)، درجه ی چهارم (پایین راست)

اکنون می توان فرم کلی برآوردگر هسته ای چگالی از یک چگالی احتمال  $X$  را ارائه داد، بر

اساس نمونه تصادفی  $X_1, X_2, \dots, X_n$  از  $f$  داریم:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i), \quad (9.1)$$

که

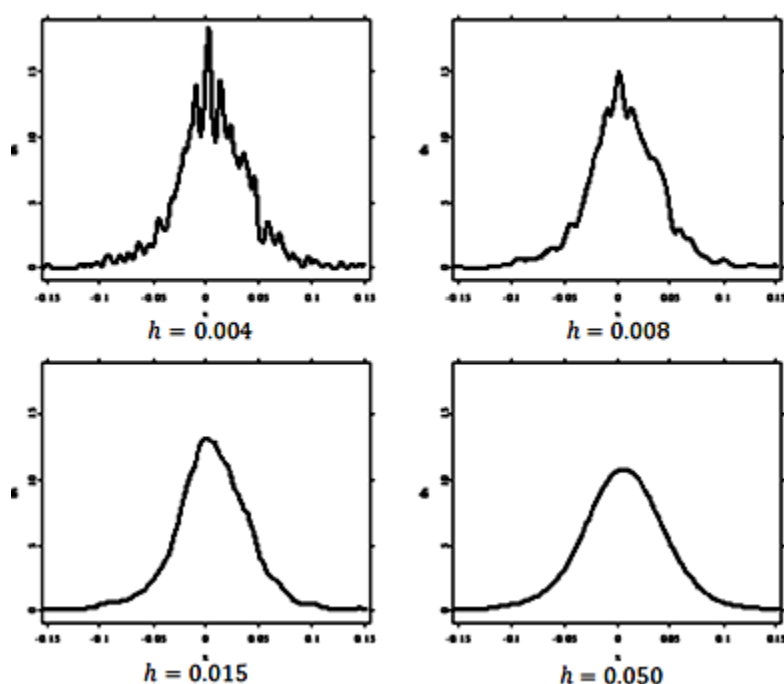
$$K_h(\bullet) = \frac{1}{h} K(\bullet/h) \quad (10.1)$$



$K(\bullet)$  یک تابع هسته مشابه آنچه در جدول ۱.۱ نمایش داده شده است، می باشد و  $h$  پهنای باند<sup>۱</sup> را مشخص می کند. توجه کنید که عبارت "تابع هسته" اشاره به تابع وزنی  $K$  (که تابعی متقارن است) دارد، در حالی که برآوردگر هسته ای چگالی اشاره به فرمول (۹.۱) دارد.

### ۳.۱ اثر پهنای باند

در مسئله برآورد تابع چگالی به روش هسته ای باید توجه داشت که پهنای باند همواری برآورد را کنترل می کند و انتخاب پهنای باند مناسب یک مسئله ی بسیار مهم است. شکل ۲.۱ برآوردهای چگالی را با استفاده از هسته ای دو وزنی و پهنای باند مختلف نمایش می دهد.



شکل ۲.۱: برآورد چگالی هسته ای با استفاده از چهار پهنای باند

$$h = 0.004, h = 0.008, h = 0.015, h = 0.050$$

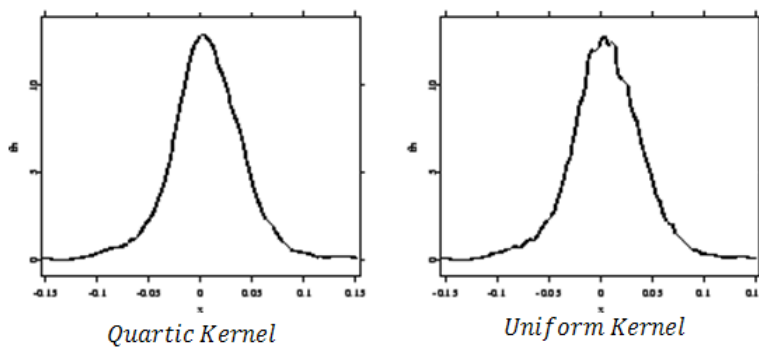
<sup>۱</sup> Bandwidth

تعیین مقداری از  $h$  که درجه ی همواری بهینه را بدون برخی محک های قراردادی در دسترس تعیین کند، دشوار است. این مسئله در بخش ۷.۱ بیشتر بحث خواهد شد.

## ۴.۱ اثر توابع هسته

توابع هسته معمولاً توابع چگالی احتمال هستند، یعنی انتگرال آن ها برابر ۱ است و برای تمام  $u$  هایی که در دامنه  $K$  هستند،  $K(u) \geq 0$ ، یک نتیجه ی فوری که از  $\int K(u)du = 1$  به دست می آید عبارت است از  $\int \hat{f}_h(x)dx = 1$ ، یعنی برآوردگر هسته ای چگالی نیز یک تابع چگالی احتمال است. بعلاوه، ویژگی های پیوستگی و مشتق پذیری  $K$  به  $\hat{f}_h$  منتقل می شود. برای مثال، اگر  $K$ ،  $u$  مرتبه پیوسته ی مشتق پذیر باشد،  $\hat{f}_h$  نیز این ویژگی را خواهد داشت.

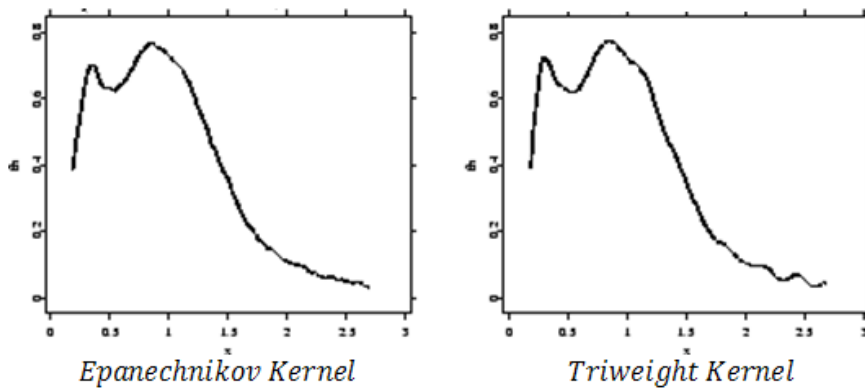
در یک سطح شهودی تر "ویژگی وراثت"  $\hat{f}_h$  در همواری نمودارش منعکس می شود. در شکل ۳.۱ برای مقدار معلوم  $h = 0.018$ ، برآوردهای هسته ای چگالی با استفاده از توابع هسته ی مختلف ترسیم شده اند.



شکل ۳.۱: برآورد هسته ای چگالی با استفاده از توابع هسته ی دو وزنی (چپ)

و یکنواخت (راست) با پهنای باند  $h = 0.018$

توجه کنید که چگونه برآورد برپایه ی هسته ی یکنواخت (راست) انعکاس دهنده ی شکل جعبه ای تابع هسته ی تحت آن با رفتار ناهموار آن است، در حالی که برآوردی که تابع هسته ی دوزنی هموار را بکار برده است (چپ)، دارای یک شکل پیوسته و هموار است.



شکل ۴.۱: برآورد هسته ای چگالی با استفاده از توابع هسته ای اپنیچ نیکوف

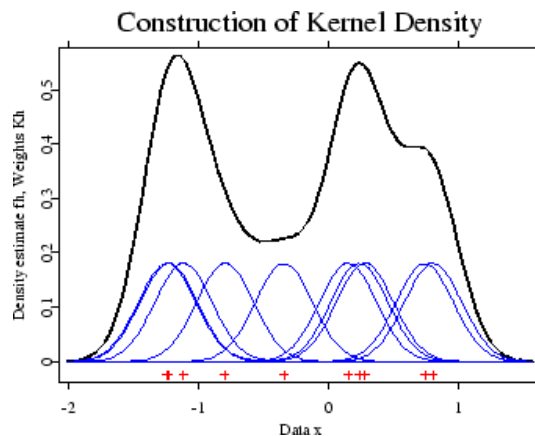
(چپ) و سه وزنی (راست) با پهنای باند  $h = ۰/۰۱۸$

این تفاوت تنها به برآوردهایی که به وسیله ی توابع هسته ی پیوسته و ناپیوسته صورت گرفته اند محدود نمی شود، حتی بین برآوردهای توابع هسته ی پیوسته تفاوت های قابل توجهی در همواری وجود دارد (برای مقادیر مشابهی از  $h$ )، این مسئله با توجه به شکل ۴.۱ قابل مشاهده است. در این شکل برای  $h = ۰/۰۱۸$  برآوردهای چگالی با استفاده از هسته ی اپنیچ نیکوف (چپ) و هسته ی سه وزنی (راست) به ترتیب، ترسیم شده است و تفاوت کاملی در همواری نمودارهای دو برآورد وجود دارد.

## ۵.۱ نمایش گرافیکی برآورد هسته ای چگالی

قبل از آنکه به ویژگی های آماری برآوردگرهای هسته ای چگالی بپردازیم اجازه دهید دیدگاه دیگری از برآورد هسته ای چگالی ارائه دهیم که انگیزه و بینش بیشتری در مورد اینکه رویه چگونه عمل می کند به ما می دهد.

شکل ۵.۱ برآورد هسته ای چگالی را برای یک مجموعه داده ساختگی به همراه توابع هسته ای متناسب نمایش می دهد.



شکل ۵.۱: برآورد هسته ای چگالی به عنوان مجموع هسته ها

تابع هسته ای متناسب به صورت زیر نمایش داده می شود:

$$\frac{1}{nh} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} K_h(x - X_i).$$

دقت کنید که مساحت زیر برآورد چگالی برابر ۱ است، بنابراین مساحت زیر هر تابع هسته ای

متناسب (با استفاده از انتگرال گیری به وسیله ی تغییر متغیر) برابر است با:

$$\int \frac{1}{nh} K\left(\frac{x - X_i}{h}\right) dx = \frac{1}{nh} \int K(u) h du = \frac{1}{nh} h \int K(u) du = \frac{1}{n}.$$