Izglađivanje empirijskih krivulja

Radašinović, Nevena

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO–MATEMATIČKI FAKULTET
MATEMATIČKI ODSJEK

Nevena Radašinović

IZGLAĐIVANJE EMPIRIJSKIH KRIVULJA

Diplomski rad

Voditelj rada:
izv. prof. dr. sc. Miljenko Huzak

Zagreb, veljača 2019.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Uvod

Svrha ovog diplomskog rada je dati pregled linearnih procjenitelja regresijskih funkcija vezanih uz problem neparametarske regresije. Regresijski problem je pronaći funkciju r na osnovi podataka $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ uzimajući da vrijedi relacija

$$Y_i = r(x_i) + \epsilon_i, \tag{1}$$

gdje su $(\epsilon_i)_{i=1,\ldots,n}$ nezavisne slučajne varijable s očekivanjem 0 i konačnom varijancom σ^2. Ponekad se varijanca od ϵ_i gleda kao funkcija od x. Također, u literaturi je najčešće zadovoljeno da su $(\epsilon_i)_{i=1,\ldots,n}$ nezavisne i jednako distribuirane slučajne varijable. Slučajnu varijablu Y nazivamo varijablom odaziva (eng. response variable), a x kovarijatom, varijablom poticaja ili prediktorom (eng. covariate, feature). Procjenitelj $\hat{r}_n(x)$ od $r(x)$ naziva se izglađivač (eng. smoother). Pojam neparametarske regresije odnosi se na takve metode pronalaska izglađivača koje zadovoljavaju minimalne pretpostavke o regresijskoj funkciji r i uglavnom se odnose na glatkoću. Vrijednosti kovarijate x se mogu tretirati kao fiksne, tj. deterministički, ili kao realizacije slučajne varijable X i u tom slučaju slučajni uzorak zapisujemo kao $(X_1, Y_1), \ldots, (X_n, Y_n)$ te regresijsku funkciju interpretiramo na sljedeći način:

$$r(x) = \mathbb{E}(Y|X = x). \tag{2}$$

U ovom radu u fokusu će biti fiksni dizajn.

Poglavlje 1

Osnovni pojmovi

1.1 Regresijski model

Neka je $\hat{r}_n(x)$ procjenitelj funkcije $r(x)$ iz regresijskog modela (1). Jasno je da moramo odabrati kriterije koje ćemo koristiti za penalizaciju pogreške procjene regresijske funkcije i predikcije varijable odaziva. U ovom su poglavlju navedeni pojmovi vezani uz razne pogreške izglađivača koji se koriste dalje u radu.
Kao funkciju gubitka koristimo kvadratnu pogrešku, u oznaci SE:

$$\mathrm{SE}(x) = (\hat{r}_n(x) - r(x))^2. \tag{1.1}$$

Očekivanje kvadratne pogreške ili srednju kvadratnu pogrešku označavamo s

$$\mathrm{MSE}(x) = \mathbb{E}(\mathrm{SE}(x)) = \mathbb{E}((\hat{r}_n(x) - r(x))^2). \tag{1.2}$$

Lako se pokaže da vrijedi

$$\begin{aligned} \mathrm{MSE}(x) &= (\mathbb{E}(\hat{r}_n(x)))^2 - 2r(x)\mathbb{E}(\hat{r}_n(x)) + r(x)^2 + \mathbb{E}(\hat{r}_n(x)^2) - (\mathbb{E}(\hat{r}_n(x)))^2 \\ &= (\mathbb{E}(\hat{r}_n(x)) - r(x))^2 + \mathbb{V}\mathrm{ar}(\hat{r}_n(x)) \end{aligned} \tag{1.3}$$

U prvom sumandu u gornjem izrazu treba prepoznati kvadrat pristranosti procjenitelja $\hat{r}_n(x)$, pri čemu je pristranost definirana kao $\mathbb{E}(\hat{r}_n(x)) - r(x)$, a u drugom sumandu varijancu istog procjenitelja. Jednadžba (1.3) ukazuje na glavni izazov u izglađivanju, a to je postići ravnotežu između pristranosti i varijance procjenitelja (eng. bias-variance tradeoff). Kažemo da smo podatke *previše izgladili* ako je pristranost velika, a varijanca mala (*velika* i *mala* za neki odabrani kriterij). Ako je pristranost mala, a varijanca velika, kažemo da smo podatke *premalo izgladili*.
SE i MSE odnose se na greške za fiksnu vrijednost x. Ako želimo dobiti uvid o tome kako se izglađivač ponaša globalno, za sve vrijednosti kovarijate, možemo koristiti

jedan od sljedeća dva pojma.
Integrirana kvadratna greška definira se kao broj

$$\text{ISE} = \int (\hat{r}_n(x) - r(x))^2 f(x) dx \tag{1.4}$$

te njezinu očekivanu vrijednost ili srednju integriranu kvadratnu grešku označavamo s

$$\text{MISE} = \mathbb{E}(\text{ISE}) = \mathbb{E} \int (\hat{r}_n(x) - r(x))^2 f(x) dx, \tag{1.5}$$

gdje je $f(x)$ gustoća točaka x, a u slučaju fiksnog dizajna $f(x) \equiv \text{const}$. Primjetimo da primjenom Fubinijevog teorema dobijemo:

$$\text{MISE} = \mathbb{E}(\text{ISE}) = \int \mathbb{E}(\text{SE}(x)) f(x) dx = \int \text{MSE}(x) f(x) dx. \tag{1.6}$$

Označimo s ARSS prosječnu sumu kvadrata reziduala vrijednosti prave funkcije i njene procjene u svim točkama x_i danih podataka:

$$\text{ARSS}(\hat{r}_n) = \frac{1}{n} \sum_{i=1}^{n} (\hat{r}_n(x_i) - r(x_i))^2. \tag{1.7}$$

Definirajmo rizik kao

$$\text{R}(\hat{r}_n) = \mathbb{E}(\text{ARSS}) = \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} (\hat{r}_n(x_i)) - r(x_i))^2. \tag{1.8}$$

Intuitivno, u slučaju fiksnog ekvidistantnog dizajna kada su x_i međusobno udaljeni za $\frac{1}{n}$, ARSS se može shvatiti kao diskretna aproksimacija ISE, a R kao diskretna aproksimacija od MISE.
Uzmimo novu observaciju $Y_i^* = r(x_i) + \epsilon_i^*$ u svakom x_i, $i = 1, \ldots, n$ tako da su $(\epsilon_i^*)_{i=1,\ldots,n}$ nezavisne s $(\epsilon_i)_{i=1,\ldots,n}$. Izglađivač nekad želimo vrednovati i kao prediktora vrijednosti varijable odaziva. Neka je $\hat{r}_n(x_i)$ predikcija od Y_i^*. Analogno, kvadratna prediktivna pogreška u x_i definira se kao

$$PRSE(Y_i^*, \hat{r}_n(x_i)) = (Y_i^* - \hat{r}_n(x_i))^2 = (r(x_i) + \epsilon_i^* - \hat{r}_n(x_i))^2. \tag{1.9}$$

Nadalje, definiramo prosječnu kvadratnu prediktivnu pogrešku i prediktivni rizik kao njeno očekivanje, odnosno:

$$\begin{aligned} PRSS(\hat{r}_n) &= \frac{1}{n} \sum_{i=1}^{n} (Y_i^* - \hat{r}_n(x_i))^2 \\ PR(\hat{r}_n) &= \mathbb{E}(\frac{1}{n} \sum_{i=1}^{n} (Y_i^* - \hat{r}_n(x_i))^2). \end{aligned} \tag{1.10}$$

Vrijedi

$$\begin{aligned}
\mathrm{PR}(\hat{r}_n) &= \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} Y_i^* - r(x_i) + r(x_i) - \hat{r}_n(x_i))^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}[\mathbb{E}(Y_i^* - r(x_i))^2 + \mathbb{E}(r(x_i) - \hat{r}_n(x_i))^2 \\
&+ 2\mathbb{E}[(Y_i^* - r(x_i))(r(x_i) - \hat{r}_n(x_i))]] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\epsilon_i^*)^2 + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\hat{r}_n(x_i)) - r(x_i))^2 + 2\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\epsilon_i^*)(r(x_i) - \hat{r}_n(x_i))]] \\
&= \sigma^2 + \mathrm{R}(\hat{r}_n).
\end{aligned} \tag{1.11}$$

Zadnja jednakost vrijedi jer su ϵ_i^* nekorelirani s $\hat{r}_n(x_i)$ pa je treći sumand jednak 0. Vidimo da su rizik i prediktivni rizik jednaki do na konstantu σ^2, a to znači da izglađivač koji minimizira rizik također minimizira i prediktivni rizik te obratno. Drugim riječima, ako nam je kriterij za odabir $\hat{r}_n$ minimizacija nekog rizika, dobar izglađivač je i dobar prediktor te obratno.
Sve gore navedene veličine odgovaraju našoj predodžbi o kvaliteti procjene, ali ih ne možemo izračunati jer bi to zahtjevalo poznavanje funkcije r (koju pokušavamo procijeniti) ili dodatno uzorkovanje varijable odaziva. Dodajmo još jedan pojam kojeg ćemo nazvati prosječna suma kvadrata:

$$\mathrm{LS} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{r}_n(x_i))^2. \tag{1.12}$$

Ovaj izraz može se dobiti iz podataka. Često se minimizacija tog izraza koristi kao kriterij odabira parametara modela i naziva se procjena metodom najmanjih kvadrata (eng. least squares estimate). Ako se koristi i kao kriterij vrednovanja modela, treba napomenuti da je LS kao procjenitelj prediktivnog rizika pristran, odnosno podcjenjuje ga. Uzrok leži u tome što se isti podaci koriste za procjenu parametara modela i za vrednovanje kvalitete procjene. Metode navedene u sljedećem odlomku pokušavaju zaobići taj nedostatak.

1.2 Unakrsno vrednovanje

Neformalno rečeno, unakrsno vrednovanje (eng. cross-validation) je metoda ocjenjivanja modela. Rezultat unakrsnog vrednovanja je procjena prediktivnog rizika.

Zbog toga se upotrebljava kao usporedni kriterij između različitih modela, ali i za određivanje hiperparametara modela. Naime, svi izglađivači u ovom radu ovisit će o tzv. parametru izglađivanja koji se zapravo tretira kao hiperparametar i često odabire tom metodom.
Kao što je spomenuto ranije, ova metoda procjene prediktivnog rizika pokušava ukloniti pristranost, barem "prema dolje", tako da podijeli uzorak na dio koji se koristi za procjenjivanje parametara modela i dio koji se koristi za vrednovanje. Postoji više vrsta unakrsnog vrednovanja, ovisno o načinu na koji se dijeli uzorak. Ovdje opisano je tzv. *k-terostruko* unakrsno vrednovanje:
Neka su $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ podaci iz modela (1). Za neki prirodni broj $k \leq n$ podatke particioniramo na k podjednakih dijelova. Neka je $\kappa : \{1, \ldots, n\} \to \{1, \ldots, k\}$ funkcija koja određuje tu particiju. Za svaki $l \in \{1, \ldots, k\}$ s $\hat{r}^{-l}$ označimo izglađivač dobiven iz podataka bez l-tog elementa particije. Tada je vrijednost koju računamo (eng. *CV-score*)

$$CV(\hat{r}) = \frac{1}{n} \sum_{i=1}^{n} L(Y_i, \hat{r}^{-\kappa(i)}(x_i)) \tag{1.13}$$

gdje je L odabrana funkcija gubitka, većinom kvadratna pogreška pa, u skladu s dosadašnjim pojmovima, uzimamo $L(Y_i, \hat{r}^{-i}(x_i)) = (Y_i - \hat{r}^{-i}(x_i))^2$. Drugim riječima, svaki element particije smo jednom koristili za evaluaciju greške izglađivača kojeg smo procjenili na podacima iz preostalih elemenata particije. U slučaju $k = n$ svaki element particije sadrži točno jedan podatak i riječ je o *"leave-one-out"* unakrsnom vrednovanju (dalje u tekstu "LOO"). U tom slučaju možemo pisati

$$CV(\hat{r}) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{r}^{-i}(x_i))^2. \tag{1.14}$$

Pogledajmo kako funkcionira LOO unakrsno vrednovanje kao procjenitelj prediktivnog rizika:

$$\begin{aligned}
\mathbb{E}(Y_i - \hat{r}^{-i}(x_i))^2 &= \mathbb{E}(Y_i - r(x_i) + r(x_i) - \hat{r}^{-i}(x_i))^2 \\
&= \mathbb{E}(\epsilon_i)^2 + 2\mathbb{E}(\epsilon_i(r(x_i) - \hat{r}^{-i}(x_i))) + \mathbb{E}(r(x_i) - \hat{r}^{-i}(x_i))^2 \\
&= \sigma^2 + \mathbb{E}(r(x_i) - \hat{r}^{-i}(x_i))^2 \\
&\approx \sigma^2 + \mathbb{E}(r(x_i) - \hat{r}_n(x_i))^2 \\
&= \sigma^2 + \mathrm{MSE}(x_i).
\end{aligned} \tag{1.15}$$

Treća jednakost slijedi iz nekoreliranosti ϵ_i s $\hat{r}^{-i}(x_i)$, a aproksimacija iz pretpostavke da je $\hat{r}^{-i}$ koji je zapravo $\hat{r}_{n-1}$ dovoljno sličan $\hat{r}_n$ jer su dobiveni na gotovo jednakim uzorcima. Sumacijom po i lako dobijemo

$$\mathbb{E}(CV(\hat{r}_n)) \approx \sigma^2 + \mathrm{R}(\hat{r}_n) = \mathrm{PR}(\hat{r}_n). \tag{1.16}$$

Iz ovoga vidimo da je LOO unakrsno vrednovanje "skoro" nepristran procjenitelj prediktivnog rizika, a točnost procjene, kao što se vidi iz raspisa, ovisi o ponašanju preciznosti izglađivača s obzirom na promjene u veličini uzorka.

1.3 Asimpotska notacija

Neka su $f : \mathbb{R} \to \mathbb{R}$ i $g : \mathbb{R} \to \mathbb{R}$ proizvoljne funkcije.
Kažemo da je f reda g i pišemo $f(x) = O(g(x))$ ako postoji $C > 0$ takav da vrijedi

$$(\forall x \in \mathbb{R})\ |f(x)| \leq C|g(x)| \tag{1.17}$$

Notacija

$$f(x) = O(g(x)),\ x \to a \tag{1.18}$$

za neki $a \in \mathbb{R}$ znači da postoje $c, C > 0$ takvi da vrijedi

$$|x - a| < c \implies |f(x)| \leq C|g(x)|. \tag{1.19}$$

Ekvivalentno je

$$f(x) = O(g(x)),\ x \to a \iff \limsup_{x \to a} \left| \frac{f(x)}{g(x)} \right| < \infty. \tag{1.20}$$

Analogno, notacija

$$f(x) = O(g(x)),\ x \to \infty \tag{1.21}$$

podrazumijeva da postoje $c, C > 0$ takvi da

$$x > c \implies |f(x)| \leq C|g(x)|. \tag{1.22}$$

Kažemo da je f malog reda g kad $x \to a$ i pišemo $f(x) = o(g(x))$ ako za svaki $\epsilon > 0$ postoji $c > 0$ takav da vrijedi

$$|x - a| < c \implies |f(x)| \leq \epsilon|g(x)|. \tag{1.23}$$

To je ekvivalentno s

$$f(x) = o(g(x)),\ x \to a \iff \lim_{x \to a} \left| \frac{f(x)}{g(x)} \right| = 0. \tag{1.24}$$

Analogna tvrdnja vrijedi za granično ponašanje u ∞.
Kažemo da je f asimptotski ekvivalentno g kad $x \to a$ i pišemo $f(x) \sim g(x),\ x \to a$ ako vrijedi

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 1. \tag{1.25}$$

Analogna tvrdnja vrijedi za granično ponašanje u ∞. Ovdje navodimo neka pravila za manipulaciju s O-notacijom, uz sve oznake kao prije.

$$f(x) = O(f(x)) \tag{1.26}$$
$$cO(f(x)) = O(f(x)) \tag{1.27}$$
$$O(O(f(x))) = O(f(x)) \tag{1.28}$$
$$O(f(x))O(g(x)) = O(f(x)g(x)) \tag{1.29}$$
$$O(f(x)g(x)) = f(x)O(g(x)). \tag{1.30}$$

1.4 Jezgre

Definicija 1.4.1. *Jezgra je funkcija $K : \mathbb{R} \to \mathbb{R}$ koja zadovoljava sljedeće uvjete:*

$$K(x) \geq 0, \tag{1.31}$$
$$\int K(x)dx = 1, \tag{1.32}$$
$$\int xK(x)dx = 0, \tag{1.33}$$
$$0 < \int x^2K(x)dx = M_2 < \infty. \tag{1.34}$$

Drugim riječima, jezgra je funkcija gustoće neke slučajne varijable koja nije konstanta i ima očekivanje 0. Ponekad se u definiciji jezgre zahtjeva još i simetričnost oko 0, odnosno parnost te dodatni uvjet

$$\int K(x)^2dx = V < \infty. \tag{1.35}$$

Definirajmo pomoćnu funkciju $I(x)$ kao

$$I(x) = \begin{cases} 1, & |x| \leq 1 \\ 0, & |x| > 1. \end{cases} \tag{1.36}$$

Sada možemo navesti neke često korištene jezgre:

Jezgra	$K(x)$	$\int x^2K(x)dx$	$\int K(x)^2dx$
Uniformna	$K(x) = \frac{1}{2}I(x)$	$\frac{1}{3}$	$\frac{1}{2}$
Gaussova	$K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	1	$\frac{1}{2\sqrt{\pi}}$
Epanechnikova	$K(x) = \frac{3}{4}(1-x^2)I(x)$	$\frac{1}{5}$	$\frac{3}{5}$
Tricube	$K(x) = \frac{70}{81}(1-\lvert x\rvert^3)^3I(x)$	$\frac{35}{243}$	$\frac{175}{247}$
Triangularna	$K(x) = (1-\lvert x\rvert)I(x)$	$\frac{1}{6}$	$\frac{2}{3}$
Biweight	$K(x) = \frac{15}{16}(1-x^2)^2I(x)$	$\frac{1}{7}$	$\frac{5}{7}$
Triweight	$K(x) = \frac{35}{32}(1-x^2)^3I(x)$	$\frac{1}{9}$	$\frac{350}{429}$

Jezgre se često koriste kod procjenjivanja funkcija gustoće kada imamo uzorak iz nepoznate neprekidne razdiobe. Navodimo kratku motivaciju. Kao što znamo,

$$f(x) = \frac{d}{dx}F(x) = \lim_{h\to 0}\frac{F(x+h) - F(x-h)}{2h}, \tag{1.37}$$

za neki $h > 0$ i gotovo svaki $x \in \mathbb{R}$. Procjena koja se koristi za funkciju distribucije je empirijska funkcija distribucije $\hat{F}$ za koju po Glivenko-Cantellijevom teoremu znamo da uniformno konvergira ka F gotovo sigurno. Ako uzmemo dovoljno mali h i u gornju jednadžbu umjesto F uvrstimo empirijsku funkciju distribucije

$$\hat{F}(x) = \frac{\#\{x_i : x_i \leq x\}}{n}, \tag{1.38}$$

dobijemo

$$\hat{f}(x) = \frac{\#\{x_i \in \big\langle x-h, x+h\big]\}}{2nh}. \tag{1.39}$$

Ako nam K ovdje označava uniformnu jezgru, tada gornji izraz možemo zapisati kao

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h}) \tag{1.40}$$

i to nazivamo procjeniteljem funkcije gustoće jezgrama. Svi procjenitelji funkcija gustoće koji koriste jezgre izgledaju kao (1.40). Gornju formulu možemo interpretirati na način da jezgre shvatimo kao težine i tada vidimo da procjenitelj svakom x_i pridruži "centar mase" jer sve jezgre navedene na početku imaju maksimum u 0 i svima osim Gaussove je nosač $\big[-1, 1\big]$. Tada se transformacijom $\frac{x-x_i}{h}$ za neki x_i pojedinom x

pridaje važnost prema tome koliko je udaljen od x_i, a parametar h određuje širinu utjecaja jezgri. Ako je nosač jezgre $\big[-1, 1\big]$, tada kompozicijom s $\frac{x-x_i}{h}$ dobijemo nosač $\big[x_i - h, x_i + h\big]$. Vrijednost procjenitelja $\hat{f}(x)$ je prosjek svih n težina dodijeljenih tom x u odnosu na sve vrijednosti kovarijate, a sve je skalirano s h da vrijednost integrala od $\hat{f}$ bude 1. Koja će se jezgra točno koristiti ovisi o problemu koji imamo i tome koje svojstvo procjenitelja želimo optimizirati.

Poglavlje 2

Linearni izglađivači

Definicija 2.0.1. *Procjenitelj $\hat{r}_n(x)$ od $r(x)$ je linearni izglađivač ako za svaki x postoji vektor $l(x) = (l_1(x), l_2(x), \ldots, l_n(x))^T$ takav da je*

$$\hat{r}_n(x) = \sum_{i=1}^{n} l_i(x) Y_i. \tag{2.1}$$

Ako definiramo $\hat{r}_n = (\hat{r}_n(x_1), \ldots, \hat{r}_n(x_n))^T$, $Y = (Y_1, \ldots, Y_n)^T$ i $L = (l(x_1), \ldots, l(x_n))^T$, tada je

$$\hat{r}_n = LY. \tag{2.2}$$

Definicija 2.0.2. *Matrica L zove se matrica izglađivanja (eng. smoothing matrix, hat matrix). i-ti redak od L zove se efektivna jezgra za procjenu $r(x_i)$, a efektivne stupnjeve slobode definiramo kao*

$$\nu = tr(L). \tag{2.3}$$

i-ti redak od L, odnosno $l(x_i)$, interpretira se kao vektor težina koje se dodjeljuju svakom Y_i pri procjeni $r(x_i)$. Nadalje, efektivni stupnjevi slobode mogu se shvatiti kao generalizacija stupnjeva slobode u parametarskoj regresiji, gdje su oni često odgovarali broju parametara koji se procjenjuju. Primjer je linearna regresija, gdje su stupnjevi slobode jednaki broju regresijskih koeficijenata. No, oni imaju možda jasniju interpretaciju ako ih definiramo na sljedeći način. Neka je $\hat{r}(Y)$ izglađivač funkcije r. Ovdje se želi naglasiti ovisnost izglađivača o realizaciji varijable odaziva, a ovisnost o realizaciji kovarijate se podrazumijeva implicitno. Neka je $M : \mathbb{R}^n \to \mathbb{R}^n$ preslikavanje takvo da $M(Y) = \hat{r}$. Definiramo generalizirane stupnjeve slobode kao

$$gdf(M) = \sum_{i=1}^{n} \frac{\partial \mathbb{E}(\hat{r}_i(Y))}{\partial r_i} \tag{2.4}$$

Dakle, generalizirani stupnjevi slobode su suma osjetljivosti očekivanih vrijednosti procjenitelja u točkama dizajna na male promjene u vrijednostima varijable odaziva. Može se reći da mjere fleksibilnost modela. Ako je model jako fleksibilan, procijenjene i stvarne vrijednosti su dosta blizu pa je osjetljiv na male promjene te su i stupnjevi slobode veliki. U slučaju linearnih izglađivača

$$\begin{aligned} gdf(M) &= \sum_{i=1}^{n} \frac{\partial \mathbb{E}(\hat{r}_i(Y))}{\partial r_i} = \sum_{i=1}^{n} \frac{\partial \mathbb{E}(\sum_{j=1}^{n} L_{ij} Y_j)}{\partial r_i} \\ &= \sum_{i=1}^{n} \frac{\partial (\sum_{j=1}^{n} L_{ij} r_j)}{\partial r_i} = \sum_{i=1}^{n} L_{ii} \\ &= tr(L). \end{aligned} \tag{2.5}$$

Vidimo da dobijemo točno definiciju efektivnih stupnjeva slobode pa interes za njih ima smisla.

Primjer 2.0.3. *(Regresogram)*
Neka su a, b takvi da je $a \leq x_i \leq b$, za $i = 1, \ldots, n$. Uzmimo ekvidistantnu subdiviziju segmenta $[a, b]$ koja određuje njegovu m-članu particiju $B_1, \ldots, B_m$ ($m-1$ poluotvorenih intervala i jedan rubni zatvoreni). Neka je s k_j označen broj točaka x_i koji pripada intervalu B_j i uzmimo da je $k_j > 0, \forall j$. Tada definiramo regresogram kao izglađivač:

$$\hat{r}_n(x) = \sum_{j=1}^{m} \mathbb{1}_{B_j}(x) \frac{1}{k_j} \sum_{i=1}^{n} \mathbb{1}_{B_j}(x_i) Y_i \tag{2.6}$$

Vidimo da je $\hat{r}_n$ step funkcija koja na svakom B_j procjenjuje r s prosjekom Y_i takvih da $x_i \in B_j$. Također, za dani $x \in B_j$ vidimo da je $\hat{r}_n$ linearna funkcija po svim Y_i s težinama $l_i(x) = \frac{1}{k_j}$ za $x_i \in B_j$ i $l_i(x) = 0$ inače. Ako pretpostavimo da su točke $x_1, \ldots, x_n$ sortirane, tj. $x_1 \leq x_2 \leq \ldots \leq x_n$ tada redak matrice izglađivanja izgleda poput

$$L_i = l(x_i)^T = (0, 0, \ldots, 0, \frac{1}{k_j}, \ldots, \frac{1}{k_j}, 0, \ldots, 0). \tag{2.7}$$

Za konkretan slučaj, npr. $m = 3$, $k_1 = 2$, $k_2 = 3$, $k_3 = 2$ to je:

$$L = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{2.8}$$

Primjetimo da regresogram ovisi o parametru $h = \frac{1}{m}$, odnosno o širini intervala B_j koju odaberemo. Taj parametar izglađivanja (tzv. širina prozora u ovom slučaju) i njegove razne inačice u primjerima koje ćemo tek navesti utječe na to koliko ćemo zagladiti procjenu - optimalno, previše ili premalo u smislu ravnoteže pristranosti i varijance izglađivača kao u (1.3). Lako se vidi da su efektivni stupnjevi slobode jednaki broju intervala B_j, $\nu = tr(L) = m$.

Jedna varijacija ovog izglađivača je izglađivač prozorima, detaljniji opis može se naći u [8]. Tamo se intervali B_j, odnosno prozori, definiraju tako da svi osim možda zadnjeg (B_m) sadrže jednak broj točaka x_i. Za odabrani $w \in [0,1]$, intervali sadrže $[wn]$ točaka (zadnji može i manje) pa je u tom slučaju w parametar zaglađivanja. Kod ove varijacije se ne moramo brinuti imamo li prazan interval B_j kao u regresogramu.

Primjer 2.0.4. *(Lokalni prosjeci)*
Neka je $h > 0$ i $B_x = \{i : |x_i - x| \leq h\}$. Označimo s n_x broj elemenata u B_x. Izglađivač lokalnim prosjecima u nekom $x \in \mathbb{R}$ definiramo kao:

$$\hat{r}_n(x) = \begin{cases} \frac{1}{n_x} \sum_{i \in B_x} Y_i, \ n_x > 0 \\ 0, \ n_x = 0. \end{cases} \tag{2.9}$$

Vidimo da $\hat{r}_n$ procjenjuje $r(x)$ tako da uzima prosjek svih Y_i za koje su x_i u h-okolini od x i da je to linearni izglađivač s težinama $l_i(x) = \frac{1}{n_x}$ za $|x_i - x| < h$ i $l_i(x) = 0$ inače.
Uzmimo za primjer jednostavan slučaj $n = 7$, $x_i = \frac{i}{7}$, $h = \frac{1}{7}$ i pretpostavimo da su x_i sortirani. Tada matrica izglađivanja izgleda ovako:

$$L = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{2.10}$$

U [8] nalazimo sličan izglađivač, tzv. izglađivač pomičnim sredinama. Za $w \in [0,1]$ takav da je $[wn]$ neparan broj te $i \in \{1, \ldots, n\}$ definiramo susjedstvo kao skup indeksa

$$B_i = \{\max(i - \frac{[wn]-1}{2}, 1), \ldots, i-1, i, i+1, \ldots, \min(i + \frac{[wn]-1}{2}, n)\}, \tag{2.11}$$

za sortirane vrijednosti varijable poticaja. To se naziva simetrično najbliže susjedstvo. Tada je procjena u x_i definirana s

$$\hat{r}_n(x_i) = \frac{1}{[wn]} \sum_{j \in B_i} Y_j. \tag{2.12}$$

Vidimo da je procjena u x_i prosjek $\frac{[wn]-1}{2}$ susjednih Y_j slijeva, $\frac{[wn]-1}{2}$ zdesna i samog Y_i. Dakle, izglađivač lokalnim prosjecima uzima u obzir susjedne x_i do neke udaljenosti, a izglađivač pomičnim sredinama susjedne x_i do nekog indeksa.

Primjer 2.0.5. *(Linearna regresija, jednostavna i polinomna)*
Sljedeći model, iako ima nešto jače pretpostavke na oblik regresijske funkcije, jedan je od najpoznatijih modela u statistici općenito i, kao što ćemo pokazati, zadovoljava definiciju linearnog izglađivača. Počinjemo od modela čija dimenzija kovarijate može biti i veća od 1. Neka je zadan model (1) i neka je $x_i = (1, x_{i1}, \ldots, x_{ip})$ za $p \geq 1$. Jednostavni linearni regresijski model pretpostavlja da je $r(x) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$, što znači da realizacija slučajnog uzorka zadovoljava

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i, \tag{2.13}$$

pri čemu su ϵ_i nezavisne slučajne varijable s homogenom varijancom. Taj model uvodi pretpostavku o linearnoj ovisnosti varijable odaziva o varijabli poticaja i naziva se linearnom regresijom. Koeficijenti se biraju procjenom najmanjih kvadrata, odnosno, procjenitelj je rješenje minimizacijskog problema navedenog prije kao izraz LS*:*

$$\mathrm{LS}(\beta) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2. \tag{2.14}$$

Ako uvedemo vektorski zapis $Y = (Y_1, \ldots, Y_n)^T$, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ i matricu dizajna

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix}, \tag{2.15}$$

model možemo zapisati kao

$$Y = X\beta + \epsilon, \tag{2.16}$$

a minimizacijski problem kao

$$\min_{\beta} ||Y - X\beta|| \tag{2.17}$$

pri čemu je $||\cdot||$ *euklidska norma na* $\mathbb{R}^n$*. Rješenje postoji jer ga iz zadnjeg zapisa možemo prepoznati kao projekciju na potprostor razapet stupcima od* X *u Hilbertovom prostoru* $\mathbb{R}^n$ *s euklidskim skalarnim produktom. Ako je* X^TX *invertibilna matrica, procjenitelj koeficijenata* β *glasi:*

$$\hat{\beta} = (X^TX)^{-1}X^TY, \tag{2.18}$$

a procjenitelj regresijske funkcije $r(x)$ *u točki* $x = (1, x_1, \ldots, x_p)$*:*

$$\hat{r}_n(x) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_x = x^T\hat{\beta}. \tag{2.19}$$

Ako uvrstimo

$$\hat{r}_n(x) = x^T(X^TX)^{-1}X^TY, \tag{2.20}$$

jasno je da je $\hat{r}_n$ *linearna funkcija od* $(Y_1, \ldots, Y_n)$*, a matrica izglađivanja*

$$L = X^T(X^TX)^{-1}X^T \tag{2.21}$$

je naravno projektor na potprostor razapet stupcima od X*. Budući da je trag projektora jednak njegovom rangu,* $p + 1 = tr(L)$*, vidimo da su ovdje efektivni stupnjevi slobode jednaki broju parametara modela.*

Ovog procjenitelja možemo primijeniti na naš početni univarijatni model (1) na više načina. Najjednostavnije je da uvrstimo $p = 1$*,* $\beta = (\beta_0, \beta_1)$ *i matricu dizajna*

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}. \tag{2.22}$$

Tu metodu procjene nazivamo jednostavnom linearnom regresijom.
Nadalje, kao što smo pretpostavili linearnu ovisnost Y *od* x*, možemo pretpostaviti polinomnu ovisnost nekog stupnja* p *i to nazivamo polinomnom regresijom. Tada model glasi:*

$$Y_i = \sum_{j=0}^{p} \beta_j x_i^j + \epsilon_i. \tag{2.23}$$

U tom slučaju tretiramo vektor $(1, x_i, x_i^2, \ldots, x_i^p)$ *kao vektor poticaja pa matrica dizajna izgleda kao*

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p \\ 1 & x_2 & x_2^2 & \ldots & x_2^p \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & x_n & x_n^2 & \ldots & x_n^p \end{pmatrix} \tag{2.24}$$

i vektor $\hat{\beta}$ je zapravo procjena koeficijenata regresijskog polinoma. Ovdje se p može shvatiti kao hiperparametar zaglađivanja koji se isto može procjenjivati.

Sada ćemo se fokusirati na 3 vrste izglađivača koje ćemo formalno podijeliti na 2 veće skupine po tome koliki utjecaj pri formiranju procjene $\hat{r}_n(x)$ za neki x imaju pojedini x_i. Uzmimo za primjer lokalne prosjeke i linearnu regresiju. Pri određivanju procjene za $r(x)$ kod lokalnih prosjeka gleda se samo prozor kojem x pripada, tj. uža okolina od x i ostali elementi slučajnog uzorka nemaju utjecaj na to. Takve metode koje pridaju različite težine pojedinim x_i pri procjenjivanju regresijske funkcije za neki x nazivamo metodama lokalne regresije. S druge strane, linearna regresija koeficijente modela određuje globalno, minimizacijom izraza koji pridaje jednake težine svim x_i pa u određivanju procjene za neki x jednaku važnost imaju svi elementi uzorka. To bismo mogli nazvati metodama globalne regresije, a njen najvažniji primjer, osim već spomenute linearne regresije, bit će regresija splajnovima.

Poglavlje 3

Metode lokalne regresije

Neka je zadan regresijski model (1). U ovom poglavlju navedeni su linearni procjenitelji od $r(x)$ koji daju veće težine onim Y_i za koje su x_i blizu x.

3.1 Lokalni procjenitelji jezgrama

Za početak možemo na jednostavan način iskoristiti ideju procjenitelja funkcija gustoće kako bismo konstruirali regresijski izglađivač. Pretpostavimo da su $x_i = \frac{i}{n}$, $i = 0, 1, \ldots, n$, dakle točke x_i su ekvidistantne i $x_i \in \big[0, 1\big]$, $i = 1, \ldots, n$, a radi lakše notacije kasnije uzimamo dodatnu točku $x_0 = 0$. Prethodno smo za vrijednost gustoće u nekom x uzimali prosjek težina dodijeljenih svim x_i pa tako ovdje za procjenu regresijske funkcije u nekom x možemo uzeti težinski prosjek Y_i koji je opet određen s položajem x_i u odnosu na x:

$$\hat{r}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h}) Y_i \tag{3.1}$$

Ako izbacimo pretpostavku da su x_i ekvidistantne, i radi lakšeg zapisa podrazumijevamo da su x_i uređeni, imamo sljedeću modifikaciju gornje definicije.

Definicija 3.1.1. *Neka je $h > 0$ širina pojasa (eng. bandwidth). Priestley-Chao procjenitelj jezgrama je linearni procjenitelj definiran relacijom*

$$\hat{r}_n(x) = \frac{1}{h} \sum_{i=1}^{n} (x_i - x_{i-1}) K(\frac{x - x_i}{h}) Y_i. \tag{3.2}$$

Širina pojasa h je hiperparametar kojeg dalje radi jednostavnosti zovemo samo parametar izglađivanja. Pokažimo kako bismo generalizirali Priestley-Chao procjenitelja u slučaju $\big[x_0, x_n\big] = \big[a, b\big]$. Kako bismo sveli taj slučaj na gornju definiciju,

napravimo linearnu transformaciju realizacije slučajnog uzorka $\widetilde{x}_i = \frac{1}{b-a}x_i - \frac{a}{b-a}$ pa smo preslikali $\big[a,b\big]$ u $\big[0,1\big]$ i na taj način smo dobili novu realizaciju slučajnog uzorka $(\widetilde{x}_1, Y_1), \ldots, (\widetilde{x}_n, Y_n)$. Dakle, argumentu x za kojeg tražimo procjenu pridružimo $\widetilde{x} = \frac{1}{b-a}x - \frac{a}{b-a}$. Neka je $\hat{r}_n(x) = \frac{1}{\widetilde{h}}\sum_{i=1}^{n}(\widetilde{x}_i - \widetilde{x}_{i-1})K(\frac{\widetilde{x}-\widetilde{x}_i}{\widetilde{h}})Y_i$ odabrani procjenitelj na transformiranoj realizaciji slučajnog uzorka. Pogledajmo što nam to znači u terminima x_i:

$$\begin{aligned}\hat{r}_n(x) &= \frac{1}{\widetilde{h}}\sum_{i=1}^{n}(\widetilde{x}_i - \widetilde{x}_{i-1})K(\frac{\widetilde{x}-\widetilde{x}_i}{\widetilde{h}})Y_i \\ &= \frac{1}{\widetilde{h}}\sum_{i=1}^{n}(\frac{1}{b-a}(x_i - x_{i-1}))K(\frac{\frac{1}{b-a}(x_i - x_{i-1})}{\widetilde{h}})Y_i \\ &= \{h = \widetilde{h}(b-a)\} = \frac{1}{h}\sum_{i=1}^{n}(x_i - x_{i-1})K(\frac{x-x_i}{h})Y_i.\end{aligned} \tag{3.3}$$

Zaključak je da, ako imamo proizvoljan uzorak, izvršimo navedenu linearnu transformaciju na njemu te na novonastalom uzorku odaberemo izglađivač, to je ekvivalentno tome da smo na početnom uzorku izabrali izglađivač sa širinom pojasa onoliko puta većom koliki je raspon kovarijate početnog uzorka. Drugim riječima, $\widetilde{h}$ je optimalan za prvi izglađivač ako i samo ako je h optimalan za drugi izglađivač. Zato možemo Priestley-Chao procjenitelja koristiti u općenitom slučaju.

Definicija 3.1.2. *Neka je $h > 0$ širina pojasa. Nadaraya-Watson procjenitelj jezgrama je linearni procjenitelj definiran relacijom*

$$\hat{r}_n(x) = \sum_{i=1}^{n}\frac{K(\frac{x-x_i}{h})}{\sum_{j=1}^{n}K(\frac{x-x_j}{h})}Y_i. \tag{3.4}$$

Izbor jezgre K nije toliko bitan jer se teoretski može pokazati da je rizik neosjetljiv na njega i procjene koje se dobiju korištenjem različitih jezgri su numerički obično dosta slične. S druge strane, izbor h je važan. Širina pojasa h je ovdje parametar izglađivanja, a to je širina okoline od x kojoj pridajemo veću težinu pri formiranju procjene i o njemu ovisi koliko ćemo izgladiti podatke - što je pojas širi podaci su izglađeniji. Općenito, h se bira u ovisnosti o veličini uzorka pa se nekad koristi u oznaci $h = h_n$. Sljedeći rezultat pokazuje kako procjenitelj ovisi o izboru h u smislu kvalitete koju mjerimo integriranim rizikom definiranim u 1.5. U tu svrhu promatrat ćemo vrijednosti kovarijate $x_1, \ldots, x_n$ kao realizaciju slučajnog uzorka iz razdiobe s gustoćom f jer nam trebaju pretpostavke o ponašanju kovarijate kada se veličina uzorka povećava, odnosno o distribuciji njenih vrijednosti. Vrijedi sljedeći teorem (vidjeti [1], str. 73, teorem 5.44):

Teorem 3.1.3. *Integrirana srednja kvadratna pogeška Nadaraya-Watson procjenitelja jezgrama je*

$$\begin{aligned}\text{MISE} = \frac{h_n^4}{4}(\int x^2 K(x)dx)^2 \int (r''(x) + 2r'(x)\frac{f'(x)}{f(x)})^2 dx + \\ + \frac{\sigma^2 \int K^2(x)dx}{nh_n} \int \frac{1}{f(x)}dx + o((nh_n)^{-1}) + o(h_n^4)\end{aligned} \tag{3.5}$$

kada $h_n \to 0$ *i* $nh_n \to \infty$.

Prvi sumand je kvadrirana pristranost, a drugi varijanca procjenitelja kao u (1.3). Primjetimo da pristranost ovisi o distribuciji x_i preko izraza

$$2r'(x)\frac{f'(x)}{f(x)} \tag{3.6}$$

i zbog toga (3.6) nazivamo pristranost dizajna. Nadalje, pokazuje se da procjenitelji jezgrama imaju veliku pristranost blizu rubova raspona vrijednosti kovarijate i ta pojava zove se granična pristranost. Deriviranjem (3.5) po h_n može se dobiti optimalna širina pojasa, ali ona ovisi o nepoznatoj regresijskoj funkciji r pa se u praksi bira LOO unakrsnim vrednovanjem.

Definicija 3.1.4. *Neka je* $h > 0$ *širina pojasa. Gasser-Müller procjenitelj jezgrama je linearni izglađivač definiran relacijom:*

$$\hat{r}_n(x) = \frac{1}{h}\sum_{i=1}^{n} Y_i \int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds, \tag{3.7}$$

pri čemu su

$$s_0 = 0, s_{i-1} \leq x_i \leq s_i, i = 1, \ldots, n, s_n = 1. \tag{3.8}$$

Vidimo da Gasser-Müller procjenitelj pojedinim Y_i pridaje težine koje su srednja vrijednost integrala funkcije K u okolini od x_i. Ako na trenutak zanemarimo da je dizajn fiksan i uzmemo u obzir da je K funkcija gustoće, možemo to interpretirati i kao da težina od Y_i ovisi o vjerojatnosti okoline od x_i s obzirom na to da je "centar vjerojatnosne gustoće" od K smješten u x. Nadalje, možemo vidjeti da je on konvolucija funkcije K i step funkcije $\sum_{i=1}^{n} \mathbb{1}_{\{x\in[s_{i-1},s_i]\}} Y_i$.
Jedna modifikacija Gasser-Müller izglađivača umjesto step funkcije u konvoluciji koristi koristi po dijelovima linearnu funkciju koja nastaje povlačenjem pravca između svake dvije susjedne točke Y_i.

Definicija 3.1.5. *Neka je $h > 0$ širina pojasa. Clarkov izglađivač jezgrama je linarni izglađivač definiran relacijom*

$$\hat{r}_n(x) = \frac{1}{h}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})p_Y(s)ds, \tag{3.9}$$

pri čemu je

$$p_Y(s) = \begin{cases} Y_1, \ s \leq x_1 \\ Y_i\frac{x_{i+1}-s}{x_{i+1}-x_i} + Y_{i+1}\frac{s-x_i}{x_{i+1}-x_i}, \ x_{i-1} < s \leq x_i \\ Y_n, \ x_n < s. \end{cases} \tag{3.10}$$

Iako za sve izglađivače jezgrama vrijede slični rezultati, ovdje ćemo se usredotočiti na one vezane uz Gasser-Müllerov izglađivač.

Konzistentnost Gasser-Müllerovog izglađivača

Želimo provesti analizu pogreške Gasser-Müllerovog izglađivača u smislu ocjene njegove srednje kvadratne pogreške i rizika i njihove ovisnosti o veličini uzorka i parametru zaglađivanja. U tu svrhu uvodimo neke dodatne pretpostavke. Neka je dizajn ekvidistantan, odnosno

$$x_i = \frac{2i-1}{2n}, \ i = 1, \ldots, n \tag{3.11}$$

$$s_0 = 0, s_n = 1, s_i = \frac{x_{i+1}+x_i}{2}, \ i = 1, \ldots, n \tag{3.12}$$

i neka je $K \in C^1(\,[-1,1]\,)$ jezgra. Dakle, jezgra zadovoljava kao i prije

$$K(x) \geq 0, \ \int_{-1}^{1} K(x)dx = 1, \ \int_{-1}^{1} xK(x)dx = 0, \ 0 < \int_{-1}^{1} x^2K(x)dx = M_2 < \infty \tag{3.13}$$

uz jednu dodatnu pretpostavku

$$\int_{-1}^{1} K(x)^2dx = V < \infty. \tag{3.14}$$

Za početak računamo srednju kvadratnu pogrešku Gasser-Müllerovog izglađivača za neki $x \in \langle 0, 1\rangle$, odnosno njen asimptotski izraz. Kao što je pokazano, znamo da je

$$\mathrm{MSE}(x) = (\mathbb{E}(\hat{r}(x)) - r(x))^2 + \mathbb{V}\mathrm{ar}(\hat{r}(x)) \tag{3.15}$$

pa srednju kvadratnu pogrešku možemo dobiti proučavajući pristranost i varijancu izglađivača zasebno.

Lema 3.1.6.

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \frac{\sigma^2}{nh^2}\int_0^1 K^2(\frac{x-u}{h})du + O((nh)^{-2}) \tag{3.16}$$

Dokaz.

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \mathbb{E}(\hat{r}(x)^2) - (\mathbb{E}(\hat{r}(x)))^2 \tag{3.17}$$

$$= \mathbb{E}\left[\left(\frac{1}{h}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\; Y_i\right)^2\right] - \left[\frac{1}{h}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\;\mathbb{E}(Y_i)\right]^2. \tag{3.18}$$

Prvi izraz u (3.18) je jednak:

$$\frac{1}{h^2}\sum_{i,j=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\int_{s_{j-1}}^{s_j} K(\frac{x-u}{h})du\;\mathbb{E}(Y_iY_j). \tag{3.19}$$

Znamo da je

$$\mathbb{E}(Y_iY_j) = r(x_i)r(x_j) + \mathbb{E}(\epsilon_i\epsilon_j) = \begin{cases} r(x_i)r(x_j),\ i \neq j \\ r(x_i)r(x_j) + \sigma^2,\ i = j \end{cases} \tag{3.20}$$

pa iz toga slijedi da je izraz u (3.19) jednak

$$\frac{1}{h^2}\sum_{i,j=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\int_{s_{j-1}}^{s_j} K(\frac{x-u}{h})du\; r(x_i)r(x_j) \tag{3.21}$$

$$+\frac{\sigma^2}{h^2}\sum_{i=1}^{n}\left(\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right)^2. \tag{3.22}$$

Drugi izraz u (3.18) je

$$\frac{1}{h^2}\sum_{i,j=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\int_{s_{j-1}}^{s_j} K(\frac{x-u}{h})du\; r(x_i)r(x_j). \tag{3.23}$$

(3.21) i (3.23) zajedno daju

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \frac{\sigma^2}{h^2}\sum_{i=1}^{n}\left(\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right)^2. \tag{3.24}$$

Pogledajmo razliku između (3.24) i tvrdnje Leme u (3.16).

$$\frac{\sigma^2}{h^2}\sum_{i=1}^{n}\left(\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right)^2 - \frac{\sigma^2}{nh^2}\int_0^1 K^2(\frac{x-u}{h})du \tag{3.25}$$

$$= \frac{\sigma^2}{h^2}\sum_{i=1}^{n}\left[\left(\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right)^2 - \frac{1}{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})^2 ds\right] \tag{3.26}$$

Po Teoremu srednje vrijednosti postoje $\theta_i \in [\, s_{i-1}, s_i]$ i $\xi_i \in [\, s_{i-1}, s_i]$, $i = 1, \ldots, n$ takvi da vrijedi

$$\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds = (s_i - s_{i-1})K(\frac{x-\theta_i}{h}) \tag{3.27}$$

$$\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})^2 ds = (s_i - s_{i-1})K(\frac{x-\xi_i}{h})^2 \tag{3.28}$$

za $i = 1, \ldots, n$. Ako to uvrstimo, (3.26) postaje

$$\frac{\sigma^2}{h^2}\sum_{i=1}^{n}(s_i - s_{i-1})\left[(s_i - s_{i-1})K(\frac{x-\theta_i}{h})^2 - \frac{1}{n}K(\frac{x-\xi_i}{h})^2\right] \tag{3.29}$$

$$= \frac{\sigma^2}{(hn)^2}\sum_{i=1}^{n}\left[K(\frac{x-\theta_i}{h})^2 - K(\frac{x-\xi_i}{h})^2\right] \tag{3.30}$$

Budući da je $K \in C^1([\,-1,1]\,)$, K^2 je posebno Lipshitz neprekidna pa postoji konstanta $C > 0$ takva da za $u, v \in [\,-1,1]$ vrijedi $|K(u)^2 - K(v)^2| \leq C|u - v|$. Uvedemo li supstituciju $u_i = \frac{x-\theta_i}{h}$ i $v_i = \frac{x-\xi_i}{h}$ za $i = 1, \ldots, n$, imamo $|u_i - v_i| \leq \frac{1}{nh}$ pa je

$$\frac{\sigma^2}{(nh)^2}\left|\sum_{i=1}^{n}\left[K(\frac{x-\theta_i}{h})^2 - K(\frac{x-\xi_i}{h})^2\right]\right| \tag{3.31}$$

$$\leq \frac{\sigma^2}{(nh)^2}\sum_{i=1}^{n}\left|K(u_i)^2 - K(v_i)^2\right| \leq \frac{C}{n^2}\sum_{u_i, v_i \in [\,-1,1]}|u_i - v_i| \tag{3.32}$$

$$\leq \frac{C\sigma^2}{(nh)^2}\frac{1}{nh}O(nh) = O((nh)^{-2})O(1) = O((nh)^{-2}) \tag{3.33}$$

jer je kardinalnost skupa $\{i : u_i, v_i \in [\,-1,1]\,\}$ reda $O(nh)$. ■

Za dovoljno mali h, izraz za varijancu iz Leme (3.1.6) se može zapisati kao

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \frac{\sigma^2 V}{nh} + O((nh)^{-2}) \tag{3.34}$$

pri čemu je $V = \int_{-1}^{1} K^2(u)du$.

Lema 3.1.7. *Neka je* $r \in C^1([\,0,1\,])$. *Tada je*

$$\mathbb{E}(\hat{r}(x)) = \frac{1}{h}\int_0^1 K(\frac{x-s}{h})r(s)ds + O(n^{-1}). \tag{3.35}$$

Dokaz. Znamo da je

$$\mathbb{E}(\hat{r}(x)) = \mathbb{E}\left[\frac{1}{h}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\; Y_i\right] \tag{3.36}$$

$$= \frac{1}{h}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\; r(x_i). \tag{3.37}$$

Pogledajmo razliku između izraza u (3.37) i tvrdnje Leme u (3.35).

$$\left|\frac{1}{h}\sum_{i=1}^{n}\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\; r(x_i) - \frac{1}{h}\int_0^1 K(\frac{x-s}{h})r(s)ds\right| \tag{3.38}$$

$$= \frac{1}{h}\left|\sum_{i=1}^{n}(r(x_i) - r(\xi_i))\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right|, \tag{3.39}$$

pri čemu su $x_{i-1} \le \xi_i \le x_i$. Po Teoremu srednje vrijednosti znamo da postoji $\theta_i \in [\,s_{i-1}, s_i]$ (odnosno $\theta_i \in [\,\min(x_i,\xi_i), \max(x_i,\xi_i)]\,$) takav da vrijedi

$$r(x_i) - r(\xi_i) = r'(\theta_i)(x_i - \xi_i) \tag{3.40}$$

za $i = 1, \ldots, n$. Zbog $|x_i - \xi_i| \le \frac{1}{n}$ imamo

$$|r(x_i) - r(\xi_i)| \le \max_{\theta\in[\,0,1]} |r'(\theta)|\frac{1}{n} \tag{3.41}$$

za $i = 1, \ldots, n$. Iz toga slijedi da je izraz u (3.39) omeđen s

$$\frac{1}{h}\sum_{i=1}^{n}|(r(x_i) - r(\xi_i))|\left|\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right| \tag{3.42}$$

$$\le \frac{1}{hn}\max_{\theta\in[\,0,1]} |r'(\theta)|\sum_{i=1}^{n}\left|\int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds\right| \tag{3.43}$$

Supstitucija $u = \frac{x-s}{h}$ uz oznake $u_{i-1} = \frac{x-s_{i-1}}{h}$ i $u_i = \frac{x-s_i}{h}$ daje gornju ogradu za (3.43).

$$\frac{1}{n} \max_{\theta \in [\,0,1]} |r'(\theta)| \sum_{i=1}^{n} \left| \int_{u_{i-1}}^{u_i} K(u)du \right| \tag{3.44}$$

$$\leq \frac{1}{n} \max_{\theta \in [\,0,1]} |r'(\theta)| \max_{u \in [\,0,1]} |K(u)| \sum_{u_i \in [\,-1,1]} |u_i - u_{i-1}|. \tag{3.45}$$

Kardinalnosti skupa $u_i : u_i \in [\,-1,1\,]$ je $O(nh)$ pa je

$$\sum_{u_i \in [\,-1,1]} |u_i - u_{i-1}| = \frac{1}{nh} O(nh) = O(1). \tag{3.46}$$

Iz toga slijedi da je izraz u (3.45) $O(n^{-1})$. ■

Korolar 3.1.8. *Ako je $r \in C^2([\,0,1\,])$, tada je pristranost od $\hat{r}(x)$ jednaka*

$$\mathbb{E}(\hat{r}(x)) - r(x) = \frac{h^2}{2} r''(x) M_2 + o(h^2) + O(n^{-1}) \tag{3.47}$$

pri čemu je $M_2 = \int_{-1}^{1} u^2 K(u)du$.

Dokaz. Lema (3.1.7) uz supstituciju $u = \frac{x-s}{h}$ daje

$$\frac{1}{h} \int_0^1 K(\frac{x-s}{h}) r(s)ds = \int_{\frac{x-1}{h}}^{\frac{x}{h}} K(u) r(x-hu)du. \tag{3.48}$$

Taylorov razvoj funkcije r u okolini od x je

$$r(x-hu) = r(x) - hr'(x)u + \frac{h^2}{2} r''(x)u^2 + o(h^2) \tag{3.49}$$

pri čemu koristimo Peanov oblik ostatka. u kojem za funkciju p vrijedi

$$\lim_{(x-hu) \to x} p(x-hu) = 0 \Rightarrow \lim_{h \to 0} p(x-hu) = 0. \tag{3.50}$$

Iz toga slijedi

$$\int_{\frac{x-1}{h}}^{\frac{x}{h}} K(u) r(x-hu)du = r(x) \int_{\frac{x-1}{h}}^{\frac{x}{h}} K(u)du - hr'(x) \int_{\frac{x-1}{h}}^{\frac{x}{h}} uK(u)du \tag{3.51}$$

$$+ \frac{h^2}{2} r'(x) \int_{\frac{x-1}{h}}^{\frac{x}{h}} u^2 K(u)du + h^2 \int_{\frac{x-1}{h}}^{\frac{x}{h}} u^2 K(u) p(x-hu)du. \tag{3.52}$$

Budući da nas zanima granično ponašanje kada $h \to 0$, ako uzmemo h_1 dovoljno mali, imamo $[-1,1] \subseteq [\frac{x-1}{h}, \frac{x}{h}]$ za svaki $h \leq h_1$ pa iz definicije od K u (1.4.1) slijedi

$$\int_{\frac{x-1}{h}}^{\frac{x}{h}} K(u)du = 1, \int_{\frac{x-1}{h}}^{\frac{x}{h}} uK(u)du = 0, \int_{\frac{x-1}{h}}^{\frac{x}{h}} u^2K(u)du = M_2 < \infty \tag{3.53}$$

Nadalje, za iste uvjete na h

$$h^2 \int_{\frac{x-1}{h}}^{\frac{x}{h}} u^2K(u)p(x-hu)du = h^2 \int_{-1}^{1} u^2K(u)p(x-hu)du. \tag{3.54}$$

Budući da je $\lim_{h\to 0} p(x-hu) = 0$, za proizvoljan $\epsilon > 0$ postoji h_2 takav je $|p(x-hu)| < \epsilon$ za svaki $h \leq h_2$. Iz toga za $h \leq \min(h_1, h_2)$ slijedi

$$h^2 \int_{-1}^{1} u^2K(u)p(x-hu)du \leq \epsilon h^2 \int_{-1}^{1} u^2K(u)du = \epsilon h^2 M_2 = o(h^2), \tag{3.55}$$

jer je ϵ proizvoljno mali. Konačno, za $h \leq \min(h_1, h_2)$ uvrštavanjem (3.53) dobijemo

$$\frac{1}{h} \int_0^1 K(\frac{x-s}{h})r(s) = r(x) + \frac{h^2}{2} r''(x)M_2 + o(h^2). \tag{3.56}$$

Iz (3.56) i Leme (3.1.7) imamo tvrdnju korolara. ■

Korolar (3.1.8) daje indikaciju da će pristranost izglađivača biti najveća tamo gdje je vrijednost $r''(x)$ najveća po apsolutnoj vrijednosti. Dakle, $\hat{r}$ će biti najviše pristran tamo gdje se nagib od r mijenja brzo, tj. tamo gdje je regresijska funkcija jako nelinearna.
Nadalje, tamo gdje je $r'' > 0$ izglađivač će imati tendenciju precjenjivanja regresijske funkcije, dok će ju za $r'' < 0$ podcjenjivati. Drugi riječima, izglađivač će precjenjivati konveksne dijelove i podcjenjivati konkavne, odnosno "izravnavati udubljenja i izbočenja" regresijske funkcije. Iz toga vizualnom inspekcijom podataka možemo dobiti ideju o tome za koje se vrijednosti kovarijate izglađivač neće ponašati najbolje u smislu pristranosti i u kojem smjeru će griješiti u procjeni.

Konačno, dolazimo do sljedećeg teorema.

Teorem 3.1.9. *Neka je $r \in C^2([0,1])$. Ako $n \to \infty$ i $h \to 0$ tako da $nh \to \infty$, tada je*

$$\text{MSE}(x) \sim \frac{\sigma^2 V}{nh} + \frac{h^4}{4} r''(x)^2 M_2^2, \tag{3.57}$$

pri čemu je V definiran u (3.34), a M_2 u (3.53). Nadalje, za $r''(x) \neq 0$ asimptotski optimalna širina pojasa je

$$h_{opt}(x) = n^{-\frac{1}{5}} \left(\frac{\sigma^2 V}{r''(x)^2 M_2^2} \right)^{\frac{1}{5}} \tag{3.58}$$

i MSE *optimiziran po h je*

$$\mathrm{MSE}_{opt}(x) \sim \frac{1.25}{n^{\frac{4}{5}}} \big(|r''(x)| \sigma^4 M_2 V^2 \big)^{\frac{2}{5}}. \tag{3.59}$$

Dokaz. Izraz za MSE(x) u (3.57) slijedi uvrštavanjem (3.34) i (3.47). Imamo

$$\begin{aligned}
(\mathbb{E}(\hat{r}(x)) - r(x))^2 &= (\frac{h^2}{2} r''(x) M_2 + o(h^2) + O(n^{-1}))^2 && (3.60)\\
&= (\frac{h^2}{2} r''(x) M_2 + O(h^2 + n^{-1}))^2 && (3.61)\\
&= \frac{h^4}{4} r''(x)^2 M_2^2 + M_2 \max_{x \in [0,1]} r''(x)^2 O(h^2) O(h^2 + n^{-1}) && (3.62)\\
&= \frac{h^4}{4} r''(x)^2 M_2^2 + O(h^2) O(h^2 + n^{-1}) && (3.63)
\end{aligned}$$

pri čemu se u zadnjoj jednakosti koristi neprekidnost od r''. Sada imamo

$$\begin{aligned}
\mathrm{MSE}(x) &= (\mathbb{E}(\hat{r}(x)) - r(x))^2 + \mathbb{V}\mathrm{ar}(\hat{r}(x)) && (3.64)\\
&= \frac{h^4}{4} r''(x)^2 M_2^2 + O(h^2) O(h^2 + n^{-1}) + \frac{\sigma^2 V}{nh} + O((nh)^{-2}), && (3.65)
\end{aligned}$$

iz čega slijedi izraz za MSE. Optimalna širina prozora dobije se deriviranjem (3.57) po h, a (3.59) uvrštavanjem (3.58) u (3.57). ■

Teorem (3.1.9) ima za posljedicu da je optimalni rizik veći tamo gdje je $|r''(x)|$ veće. Intuitivno, jasno je da je regresijsku funkciju teže procijeniti u točkama u čijoj okolino njezino ponašanje jako varira, tj. tamo gdje je dosta "krivudava". S druge strane, optimalna širina prozora je veća tamo gdje je $|r''(x)|$ manje. Uzmimo prvo da je $r''(x)$ približno 0 za neki x. To znači da je r skoro linearna u okolini od x i da sve točke u okolini od x daju dobru informaciju o ponašanju r. Iz tog razloga h treba proširiti da bismo jače "uprosječili" procjenu i tako smanjili varijancu izglađivača (jer time nećemo puno povećati pristranost). Obrnuto, ako je $|r''(x)|$ veliko, to znači da se ponašanje funkcije u okolini x intenzivno mijenja i samo točke jako blizu x sadržavaju korisne informacije o r. Iz tog razloga treba smanjiti širinu pojasa, ali ovaj put kako

bismo smanjili pristranost izglađivača.

Teorem (3.1.9) nam govori o optimalnom MSE(x) i optimalnoj širini pojasa za neki fiksni x. Vidimo da je optimalna brzina opadanja srednje kvadratne pogreške $O(n^{-\frac{4}{5}})$. No, taj teorem ne možemo direktno primijeniti ako bismo htjeli proučiti $\mathrm{R}(\hat{r}) = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(\hat{r}_n(x_i)) - r(x_i))^2$, tj. asimptotsko ponašanje rizika izglađivača. Razlog leži u tome što se u Korolaru (3.1.8) i posljedici Leme (3.1.6) za fiksni x zahtjeva dovoljno mali h da bi vrijedile jednakosti u (3.53) i izraz za varijancu u (3.34), odnosno potrebno je $[\,-1,1\,] \subseteq [\,\frac{x-1}{h}, \frac{x}{h}\,]$. Drugim riječima, mora biti $x \in [\,h, 1-h\,]$ i takvu točku nazivamo unutrašnjom točkom. Ako računamo MSE(x_i) za $i = 1, \ldots, n$ i gledamo asimptotsko ponašanje po $n \to \infty$ i $h \to 0$ tako da $nh \to \infty$, ne možemo postići da sve realizacije kovarijate budu unutrašnje točke. Naš dizajn je ekvidistantan s razmakom širine $\frac{1}{n}$ među točkama x_i, a zbog $nh \to \infty$ vidimo da $\frac{1}{n} \to 0$ brže nego $h \to 0$ pa nam neizbježno ostaju neki $x_i \in [\,0, h\rangle \cup \langle 1-h, 1\,]$ i njih nazivamo graničnim točkama. U graničnim točkama imamo problem veće pristranosti koji nazivamo granična pristranost. Jedan pokušaj da se riješi taj problem je modifikacija procjenitelja

$$\hat{r}_c(x) = \sum_{i=1}^{n} Y_i \int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds \Big/ \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K(\frac{x-s}{h})ds. \tag{3.66}$$

To je tzv. *cut-and-normalize* izglađivač i ima svojstvo da je $\hat{r}_c(x) = \hat{r}(x)$ za $x \in [\,h, 1-h\,]$, no u graničnim točkama $\hat{r}_c$ ima manju pristranost. Može se pokazati da je u tom slučaju optimalna brzina opadanja rizika $O(n^{-\frac{3}{4}})$, a ne $O(n^{-\frac{4}{5}})$ kao u slučaju unutrašnjih točaka što bismo htjeli. Drugi pristup smanjivanju granične pristranosti je korištenje tzv. graničnih jezgri. Za početak uzmimo proizvoljan $x \in [\,0, h\rangle$, odnosno donju graničnu točku. Tada je $x = qh$ za neki $q \in [\,0, 1\rangle$. Pretpostavimo da za svaku takvu točku možemo konstruirati modifikaciju jezgre K, odnosno graničnu jezgru K_q koja zadovoljava

$$K_q(x) \geq 0, \tag{3.67}$$

$$\int_{-1}^{q} K_q(x)dx = 1, \tag{3.68}$$

$$\int_{-1}^{q} xK_q(x)dx = 0, \tag{3.69}$$

$$0 < \int_{-1}^{q} x^2 K_q(x)dx = M_{2q} \tag{3.70}$$

$$\int_{-1}^{q} K_q(x)^2 dx = V_q < \infty \tag{3.71}$$

te da je $K_q(x)$ diferencijabilna po x i neprekidna po q i da vrijedi

$$q \to 1 \implies K_q \to K. \tag{3.72}$$

Ako imamo takvu graničnu jezgru i nju koristimo u izglađivaču za neku graničnu točku $x = qh$, analogno kao u Lemi (3.1.6) pokaže se da je

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \frac{\sigma^2}{nh} V_q + O((nh)^{-2}), \tag{3.73}$$

a jer je K_q ograničena i neprekidna po q, također je po teoremu o dominiranoj konvergenciji V_q neprekidna po q pa je

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) \leq \frac{\sigma^2}{nh} \max_{q\in[\,0,1]} V_q + O((nh)^{-2}) \tag{3.74}$$

$$\implies \mathbb{V}\mathrm{ar}(\hat{r}(x)) = O((nh)^{-1}) + O((nh)^{-2}) = O((nh)^{-1}). \tag{3.75}$$

Također, jednakosti (3.53) u (3.1.8) su ispunjene jer je $[\,-1, q\,] \subseteq [\,\frac{x-1}{h}, \frac{x}{h}\,]$ za $x = qh$ pa je

$$\mathbb{E}(\hat{r}(x)) - r(x) = \frac{h^2}{2} r''(x) M_{2q} + o(h^2) + O(n^{-1}) \tag{3.76}$$

$$\leq \frac{h^2}{2} \max_{x\in[\,0,1]} |r''(x)| M_{2q} + o(h^2) + O(n^{-1}) \tag{3.77}$$

$$= O(h^2) + O(n^{-1}) = O(h^2 + n^{-1}) \tag{3.78}$$

Iz (3.75) i (3.78) imamo ocjenu za MSE donjih graničnih točaka kad $n \to \infty$ i $h \to 0$ tako da $nh \to \infty$:

$$\mathrm{MSE}(qh) = (\mathbb{E}(\hat{r}(qh)) - r(qh))^2 + \mathbb{V}\mathrm{ar}(\hat{r}(qh)) \tag{3.79}$$

$$= O((nh)^{-1}) + O((h^2 + n^{-1})^2) \tag{3.80}$$

$$= O((nh)^{-1}) + O(h^4 + h^2 n^{-1} + n^{-2}) \tag{3.81}$$

$$= O((nh)^{-1}) + O(h^4 + n^{-2}) \tag{3.82}$$

$$= O((nh)^{-1} + h^4), \tag{3.83}$$

pri čemu sve gornje ocjene vrijede uniformno po q.
Analogno se pokaže da jednaka ocjena za MSE vrijedi i za gornje granične točke, odnosno točke oblika $x = 1 - qh$ za $q \in [\,0, 1\rangle$ ako za izglađivač u tom slučaju

koristimo granične jezgre sa svojstima

$$K_q(x) \geq 0, \tag{3.84}$$

$$\int_{-q}^{1} K_q(x)dx = 1, \tag{3.85}$$

$$\int_{-q}^{1} xK_q(x)dx = 0, \tag{3.86}$$

$$0 < \int_{-q}^{1} x^2 K_q(x)dx = M_{2q} \tag{3.87}$$

$$\int_{-q}^{1} K_q(x)^2 dx = V_q < \infty \tag{3.88}$$

te kao i prije da je $K_q(x)$ diferencijabilna po x i neprekidna po q i $\lim_{q\to\infty} K_q = K$. Primjetimo da ako imamo donju graničnu jezgru $K_q(x)$, gornju graničnu jezgru možemo dobiti kompozicijom $K_q(-x)$. U tablici ispod navedeni su neki primjeri donjih graničnih jezgri.

Jezgra K	$K_q(x)$
Uniformna	$K_q(x) = \frac{2}{(1+q)^3}[\,3(1-q)x + 2(1-q+q^2)]\,I_{[\,-1,q]}\,(x)$
Epanechnikova	$K_q(x) = \frac{12(x+1)}{(1+q)^4}[\,(1-2q)x + \frac{1}{2}(3q^2-2q+1)]\,I_{[\,-1,q]}\,(x)$
Biweight	$K_q(x) =$ $\frac{15(1+x)^2(q-x)}{(1+q)^5}[\,2x(5\frac{1-q}{1+q}-1) + (3q-1) + 5\frac{(1-q)^2}{1+q}]\,I_{[\,-1,q]}\,(x)$

Sada imamo sve potrebno da odredimo asimptotski rizik Gasser-Müllerovog izglađivača te kao za MSE pronađemo h koji ga minimizira, odnosno nađemo globalno optimalan parametar izglađivanja i optimalan rizik.

Teorem 3.1.10. *Neka je $r \in C^2(\,[\,0,1]\,)$. Ako u svim graničnim točkama u izglađivaču koristimo pripadajuće granične jezgre, tada za $n \to \infty$ i $h \to 0$ tako da $nh \to \infty$, vrijedi*

$$\mathrm{R}(\hat{r}) \sim \frac{\sigma^2 V}{nh} + \frac{h^4}{4} M_2^2 J_2(r), \tag{3.89}$$

pri čemu je $J_2(r) = \int_0^1 r''(x)^2 dx$. Nadalje, asimptotski globalno optimalna širina pojasa je

$$h_{opt} = n^{-\frac{1}{5}} \left(\frac{\sigma^2 V}{J_2(r) M_2^2} \right)^{\frac{1}{5}}, \tag{3.90}$$

a asimptotski optimalan rizik

$$\mathrm{R}(\hat{r})_{opt} \sim \frac{1.25}{n^{\frac{4}{5}}} J_2(r)^{\frac{1}{5}} (\sigma^4 M_2 V^2)^{\frac{2}{5}}. \tag{3.91}$$

Dokaz.
Neka je $B = \{x_i : x_i < h \vee x_i > 1 - h\}$. Vrijedi

$$\mathrm{R}(\hat{r}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(\hat{r}(x_i) - r(x_i))^2 \tag{3.92}$$

$$= \frac{1}{n} \sum_{x_i \notin B} \mathbb{E}(\hat{r}(x_i) - r(x_i))^2 + \frac{1}{n} \sum_{x_i \in B} \mathbb{E}(\hat{r}(x_i) - r(x_i))^2 \tag{3.93}$$

Uvrštavanjem rezultata iz (3.57) i (3.83) dobivamo

$$\mathrm{R}(\hat{r}) \sim \frac{1}{n} \sum_{x_i \notin B} \left(\frac{\sigma^2 V}{nh} + \frac{h^4}{4} r''(x_i)^2 M_2^2 + \right) + \frac{1}{n} O(nh) O((nh)^{-1} + h^4) \tag{3.94}$$

$$= \frac{\sigma^2 V}{nh} + \frac{h^4}{4} M_2^2 \frac{1}{n} \sum_{x_i \notin B} r''(x_i)^2 + O(h) O((nh)^{-1} + h^4). \tag{3.95}$$

Budući da je r'' neprekidna,

$$\lim_{n \to \infty, h \to 0} \frac{1}{n} \sum_{x_i \notin B} r''(x_i)^2 = \int_0^1 r''(x)^2 dx = J_2(r) \tag{3.96}$$

pa imamo

$$\mathrm{R}(\hat{r}) \sim \frac{\sigma^2 V}{nh} + \frac{h^4}{4} M_2^2 J_2(r). \tag{3.97}$$

Isto kao prije, izraz za h_{opt} dobije se deriviranjem (3.89) po h, a izraz za $\mathrm{R}(\hat{r})_{opt}$ uvrštavanjem h_{opt} u (3.89). ■

Teorem (3.1.10) kaže da ako prilagodimo izglađivač u graničnim točkama, možemo sa individualnih točaka prenijeti globalno red konvergencije $O(n^{-\frac{4}{5}})$. Nažalost, Teorem (3.1.10) ima slabu praktičnu korist jer je za određivanje optimalne širine pojasa potrebno poznavanje σ i $J_2(r)$. No, moguće je aproksimirati $J_2(r)$, a za σ postoje konzistentni procjenitelji tako da se teoretski može doći do procjene za h_{opt}.

Svi rezultati za konzistentnost Gasser-Müllerovog izglađivača dobiveni su pod pretpostavkom uniformnog dizajna. Oni se mogu poopćiti tako da se dizajn generira iz pozitivne i diferencijabilne funkcije gustoće w tako da je zadovoljena relacija

$$\int_0^{x_i} w(x) dx = \frac{2i - 1}{2n}, \ i = 1, \ldots, n. \tag{3.98}$$

Može se pokazati da je izraz za pristranost izglađivača za neku fiksnu unutarnju točku x u takvom dizajnu jednak kao u Korolaru (3.1.8), dok izraz za varijancu postaje

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \frac{\sigma^2 V}{w(x)nh} + O((nh)^{-2}). \tag{3.99}$$

U tom slučaju optimalna širina pojasa za fiksni x je

$$h_{opt} = n^{-\frac{1}{5}} \left(\frac{\sigma^2 V}{w(x) r''(x)^2 M_2^2} \right)^{\frac{1}{5}}, \tag{3.100}$$

a optimalni rizik

$$\mathrm{R}(\hat{r})_{opt} = \frac{1.25}{n^{\frac{4}{5}}} \left(\frac{\sigma^2 V}{w(x)} \right)^{\frac{4}{5}} (r''(x)^2 M_2^2)^{\frac{1}{5}}. \tag{3.101}$$

Uz modifikaciju izglađivača graničnim jezgrama u graničnim točkama, dobijemo globalno optimalnu širinu pojasa

$$h_{opt} = n^{-\frac{1}{5}} \left(\frac{\sigma^2 V}{J_2(r) M_2^2} \right)^{\frac{1}{5}} \tag{3.102}$$

i optimalan rizik

$$\mathrm{R}(\hat{r}) = \frac{1.25}{n^{\frac{4}{5}}} (\sigma^2 V)^{\frac{4}{5}} (J_2(r) M_2^2)^{\frac{1}{5}}, \tag{3.103}$$

gdje je sada

$$J_2(r) = \int_0^1 r''(x)^2 w(x) dx. \tag{3.104}$$

3.2 Lokalna polinomna regresija

Za motivaciju sljedećeg izglađivača promotrit ćemo iz druge perspektive već spomenutog Nadaraya-Watson procjenitelja jezgrama. On ima sljedeće svojstvo. Recimo da za neki fiksni x regresijsku funkciju želimo aproksimirati konstantom u okolini x, odnosno $a_0 = \hat{r}_n(x)$, a za minimizacijski kriterij odaberemo težinsku sumu kvadrata

$$\sum_{i=1}^{n} K(\frac{x_i - x}{h})(Y_i - a_0)^2, \tag{3.105}$$

pri čemu je K jezgra. Traženjem minimuma (3.105) po a_0 odnosno parcijalnom derivacijom tog izraza lako se dobije

$$a_0 = \sum_{i=1}^{n} \frac{K(\frac{x-x_i}{h})}{\sum_{j=1}^{n} K(\frac{x-x_j}{h})} Y_i, \tag{3.106}$$

što je točno Nadaraya-Watson procjenitelj jezgrama. Budući da se a_0 određuje za svaki pojedini x, tu ovisnost navodimo eksplicitno pa pišemo $a_0 = a_0(x)$. Primjetimo da smo ovdje umjesto $K(\frac{x-x_i}{h})$ uzimali $K(\frac{x_i-x}{h})$ da naglasimo centriranost oko točke x. Zato u ovom dijelu radi dosljednosti zahtijevamo dodatni uvjet na K, a to je simetričnost oko 0.
Težinskom sumom kvadrata pri formiranju procjene $\hat{r}(x)$ najveću ulogu dobivaju oni x_i koji su blizu x i za koje bi a_0 trebala i sama dobro odgovarati kao procjena regresijske funkcije $r(x_i)$. Argument za to je i Taylorov teorem koji kaže

$$r(x_i) = r(x) + O(|x_i - x|) \tag{3.107}$$

pa je prirodno najveću težinu pridavati onim x_i u kojima je vrijednost regresijske funkcije blizu $r(x)$.
Želimo napraviti generalizaciju kriterija u (3.105). Opet u duhu Taylorovog teorema, mogli bismo r u okolini nekog fiksnog x umjesto konstantom a_0 aproksimirati polinomom stupnja p. Za neki u u okolini od x definiramo polinom

$$p_x(u) = a_0 + a_1(u-x) + \frac{a_2}{2!}(u-x)^2 + \ldots + \frac{a_p}{p!}(u-x)^p. \tag{3.108}$$

Regresijsku funkciju $r(u)$ u okolini od x aproksimiramo polinomom

$$r(u) = p_x(u). \tag{3.109}$$

Analogno, koeficijente $a = (a_0, \ldots, a_n)$ biramo tako da minimiziramo

$$\sum_{i=1}^{n} K(\frac{x-x_i}{h})(Y_i - p_x(x_i))^2. \tag{3.110}$$

Dakle, procjena r u okolini od x glasi

$$\hat{r}_n(u) = \hat{a}_0 + \hat{a}_1(u-x) + \frac{\hat{a}_2}{2!}(u-x)^2 + \ldots + \frac{\hat{a}_p}{p!}(u-x)^p. \tag{3.111}$$

U samoj točki $u = x$ to je

$$\hat{r}_n(x) = \hat{a}_0(x). \tag{3.112}$$

Za $p = 0$ izglađivač kojeg ova metoda daje je, kao što je već pokazano, Nadaraya-Watson izglađivač jezgrama, a u slučaju $p = 1$ metoda se naziva lokalna linearna regresija.
Da bismo lakše došli do izraza za koeficijente $\hat{a}$ uvedimo matrične oznake

$$X_x = \begin{bmatrix} 1 & x_1 - x & \ldots & \frac{(x_1-x)^p}{p!} \\ 1 & x_2 - x & \ldots & \frac{(x_2-x)^p}{p!} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x & \ldots & \frac{(x_n-x)^p}{p!} \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} \tag{3.113}$$

i neka je W_x $n \times n$ dijagonalna matrica čiji je element na mjestu (i,i) jednak $K(\frac{x-x_i}{h})$, odnosno $W_x = diag(K(\frac{x_1-x}{h}), K(\frac{x_2-x}{h}), \ldots, K(\frac{x_n-x}{h}))$. Tada težinsku sumu (3.110) možemo zapisati kao

$$(Y - X_x a)^T W_x (Y - X_x a). \tag{3.114}$$

Imamo

$$\begin{aligned} f(a) &= (Y - X_x a)^T W_x (Y - X_x a) = (Y^T - a^T X_x^T) W_x (Y - X_x a) && (3.115)\\ &= Y^T W_x Y - Y^T W_x X_X a - a^T X_x^T W_x Y + a^T X_x^T W_x X_x a && (3.116)\\ &= Y^T W_x Y - 2Y^T W_x X_X a + a^T X_x^T W_x X_x a. && (3.117) \end{aligned}$$

Želimo minimizirati izraz (3.117) po a, odnosno dobiti diferencijal funkcije f. Prvi dio gornjeg izraza je afina funkcija, a za općenitu kvadratnu formu

$$g(a) = a^T S a = \sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij} a_i a_j, \tag{3.118}$$

pri čemu je S proizvoljna kvadratna matrica, lako dobijemo

$$\begin{aligned} \frac{\partial g}{\partial a_k}(a) &= \sum_{j \neq k} S_{kj} a_j + 2S_{kk} a_k + \sum_{i \neq k} S_{ik} a_i && (3.119)\\ &= \sum_{i=1}^{n} S_{ik} a_i + \sum_{j=1}^{n} S_{kj} a_j = \sum_{i=1}^{n} (S_{ik} + S_{ki}) a_i = [\, a^T (S^T + S)]_{\,k} && (3.120)\\ \implies Dg(a) &= a^T (S^T + S). && (3.121) \end{aligned}$$

Izjednačavanjem $Df(a) = 0$ imamo

$$\begin{aligned} 0 &= -2Y^T W_x X_X + a^T ((X_x^T W_x X_x)^T + X_x^T W_x X_x) && (3.122)\\ &= -2Y^T W_x X_X + 2a^T (X_x^T W_x X_x) && (3.123)\\ \implies a^T (X_x^T W_x X_x) &= Y^T W_x X_X. && (3.124) \end{aligned}$$

Pod pretpostavkom invertibilnosti matrice $X_x^T W_x X_x$ slijedi

$$\begin{aligned} a^T &= Y^T W_x X_X (X_x^T W_x X_x)^{-1} && (3.125)\\ a &= (X_x^T W_x X_x)^{-1} X_x^T W_x Y. && (3.126) \end{aligned}$$

Budući da je naša procjena u točki x jednaka slobodnom članu, ako s e_1 označimo $(p+1) \times 1$ vektor s prvim elementom 1 i ostalim 0, odnosno $e_1 = (1, 0, \ldots, 0)^T$, možemo pisati

$$\hat{r}(x) = a_0(x) = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x Y \tag{3.127}$$

Osim za slučaj $p = 0$, jednostavnija eksplicitna formula postoji i za slučaj $p = 1$. Matrica X_x je tada

$$X_x = \begin{bmatrix} 1 & x_1 - x \\ 1 & x_2 - x \\ \vdots & \\ 1 & x_n - x \end{bmatrix}. \tag{3.128}$$

Imamo

$$X_x^T W_x X_x = \begin{bmatrix} 1 & \dots & 1 \\ x_1 - x & \dots & x_n - x \end{bmatrix} \begin{bmatrix} K(\frac{x_1-x}{h}) & & \\ & \ddots & \\ & & K(\frac{x_n-x}{h}) \end{bmatrix} \begin{bmatrix} 1 & x_1 - x \\ 1 & x_2 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix} \tag{3.129}$$

$$= \begin{bmatrix} K(\frac{x_1-x}{h}) & \dots & K(\frac{x_n-x}{h}) \\ (x_1 - x)K(\frac{x_1-x}{h}) & \dots & (x_n - x)K(\frac{x_n-x}{h}) \end{bmatrix} \begin{bmatrix} 1 & x_1 - x \\ 1 & x_2 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix} \tag{3.130}$$

$$= \begin{bmatrix} \sum_{i=1}^n K(\frac{x_i-x}{h}) & \sum_{i=1}^n (x_i - x)K(\frac{x_i-x}{h}) \\ \sum_{i=1}^n (x_i - x)K(\frac{x_i-x}{h}) & \sum_{i=1}^n (x_i - x)^2 K(\frac{x_i-x}{h}). \end{bmatrix} \tag{3.131}$$

Uvedemo li oznake

$$s_r(x) = \frac{1}{n} \sum_{i=1}^n (x_i - x)^r K(\frac{x_i - x}{h}), \tag{3.132}$$

izraz (3.131) postaje

$$X_x^T W_x X_x = n \begin{bmatrix} s_0(x) & s_1(x) \\ s_1(x) & s_2(x) \end{bmatrix}. \tag{3.133}$$

Sada imamo

$$n(X_x^T W_x X_x)^{-1} = \frac{1}{s_0(x)s_2(x) - s_1(x)^2} \begin{bmatrix} s_2(x) & -s_1(x) \\ -s_1(x) & s_0(x) \end{bmatrix}. \tag{3.134}$$

Nadalje,

$$X_x^T W_x Y = \begin{bmatrix} \sum_{i=1}^n K(\frac{x_i-x}{h}) Y_i \\ \sum_{i=1}^n (x_i - x)K(\frac{x_i-x}{h}) Y_i \end{bmatrix}. \tag{3.135}$$

Iz (3.134) i (3.135) dobijemo

$$\hat{r}(x) = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x Y \tag{3.136}$$

$$= \frac{\frac{1}{n}}{s_0(x)s_2(x) - s_1(x)^2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s_2(x) & -s_1(x) \\ -s_1(x) & s_0(x) \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n K(\frac{x_i - x}{h}) Y_i \\ \sum_{i=1}^n (x_i - x) K(\frac{x_i - x}{h}) Y_i \end{bmatrix} \tag{3.137}$$

$$= \frac{\frac{1}{n}}{s_0(x)s_2(x) - s_1(x)^2} \begin{bmatrix} s_2(x) & -s_1(x) \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n K(\frac{x_i - x}{h}) Y_i \\ \sum_{i=1}^n (x_i - x) K(\frac{x_i - x}{h}) Y_i \end{bmatrix} \tag{3.138}$$

$$= \frac{\frac{1}{n}}{s_0(x)s_2(x) - s_1(x)^2} \left[s_2(x) \sum_{i=1}^n K(\frac{x_i - x}{h}) Y_i - s_1(x) \sum_{i=1}^n (x_i - x) K(\frac{x_i - x}{h}) Y_i \right] \tag{3.139}$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{[\, s_2(x) - s_1(x)(x_i - x)] \, K(\frac{x_i - x}{h}) Y_i}{s_0(x)s_2(x) - s_1(x)^2}. \tag{3.140}$$

Pretpostavimo da je K jezgra s nosačem $[\,-1, 1]$, simetrična oko 0 i $\int K(x)^2 dx = V < \infty$. Nadalje, neka je dizajn $x_i = \frac{2i-1}{2n}$, $i = 1, \ldots, n$, regresijska funkcija $r \in C^2([\,0, 1]\,)$ te $x \in [\,h, 1-h]$ unutarnja točka. Želimo izračunati MSE(x) za lokalno linearno izglađivanje i koristimo sličnu tehniku kao za Gasser-Müller izglađivač jezgrama.

Prvo slijedi pomoćni rezultat.

Lema 3.2.1. *Neka je $s_r(x)$ definiran kao u (3.132). Tada je*

$$s_r(x) = \int_0^1 (s - x)^r K(\frac{s - x}{h}) ds + O(n^{-1}). \tag{3.141}$$

Dokaz. Neka su $s_0 = 0, s_n = 1, s_i = \frac{x_{i+1} + x_i}{2}$ slično kao prije. Po Teoremu srednje vrijednosti postoje $\epsilon_i \in [\, s_{i-1}, s_i]\,, i = 1, \ldots, n$ takvi da vrijedi

$$I(x) = \int_0^1 (s - x)^r K(\frac{s - x}{h}) ds = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} (s - x)^r K(\frac{s - x}{h}) ds \tag{3.142}$$

$$= \sum_{i=1}^n (s_{i-1} - s_i) K(\frac{\epsilon_i - x}{h}) (\epsilon_i - x)^r \tag{3.143}$$

$$= \frac{1}{n} \sum_{i=1}^n K(\frac{\epsilon_i - x}{h}) (\epsilon_i - x)^r \tag{3.144}$$

Nadalje, zbog Lipschitz neprekidnosti funkcije $(s-x)^r K(\frac{s-x}{h})$ po s imamo

$$|s_r(x) - I(x)| = \frac{1}{n}\left|\sum_{i=1}^{n}\left(K(\frac{x_i-x}{h})(x_i-x)^r - K(\frac{\epsilon_i-x}{h})(\epsilon_i-x)^r\right)\right| \tag{3.145}$$

$$\leq \frac{1}{n}\sum_{x_i\in[\,x-h,x+h]} C\left|\frac{x_i-x}{h} - \frac{\epsilon_i-x}{h}\right| \tag{3.146}$$

$$= \frac{1}{n}\sum_{x_i\in[\,x-h,x+h]} C\left|\frac{x_i-\epsilon_i}{h}\right| \tag{3.147}$$

za neku konstantu $C > 0$. Budući da je kardinalnost skupa $\{x_i \in [\,x-h, x+h\,]\}$ jednaka $O(nh)$ i $\left|\frac{x_i-\epsilon_i}{h}\right| < \frac{1}{nh}$, izraz u (3.147) je omeđen s

$$\frac{1}{n}\sum_{x_i\in[\,x-h,x+h]} C\left|\frac{x_i-\epsilon_i}{h}\right| \leq \frac{1}{n}O(nh)\frac{1}{nh} = O(n^{-1}), \tag{3.148}$$

što je tvrdnja Leme. ■

Lema 3.2.2. *Pod navedenim pretpostavkama,*

$$\mathbb{E}(\hat{r}(x)) - r(x) = \frac{h^2}{2}r''(x)M_2 + o(h^2) + O(n^{-1}), \tag{3.149}$$

pri čemu je $M_2 = \int_{-1}^{1} x^2K(x)dx$ *kao i prije.*

Skica dokaza. Uz oznaku $R = \begin{bmatrix} r(x_1) \\ \vdots \\ r(x_n) \end{bmatrix}$, znamo da je

$$\mathbb{E}(\hat{r}(x)) = e_1^T(X_x^T W_x X_x)^{-1} X_x^T W_x R. \tag{3.150}$$

Taylorov razvoj funkcije r u okolini od x za pojedini x_i glasi

$$r(x_i) = r(x) + r'(x)(x_i-x) + \frac{r''(x)}{2}(x_i-x)^2 + o(|x_i-x|^2). \tag{3.151}$$

Zato možemo pisati

$$R = X_x\begin{bmatrix} r(x) \\ r'(x) \end{bmatrix} + \frac{r''(x)}{2}\begin{bmatrix} (x_1-x)^2 \\ \vdots \\ (x_n-x)^2 \end{bmatrix} + \begin{bmatrix} o(|x_1-x|^2) \\ \vdots \\ o(|x_n-x|^2) \end{bmatrix}. \tag{3.152}$$

Ako prvi dio izraza (3.152) ubacimo u formulu za očekivanje procjene u x, dobijemo

$$e_1^T(X_x^T W_x X_x)^{-1} X_x^T W_x X_x \begin{bmatrix} r(x) \\ r'(x) \end{bmatrix} = r(x). \tag{3.153}$$

Zato je

$$\mathbb{E}(\hat{r}(x)) - r(x) = e_1^T(X_x^T W_x X_x)^{-1} X_x^T W_x \left(\frac{r''(x)}{2} \begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{bmatrix} + \begin{bmatrix} o(|x_1 - x|^2) \\ \vdots \\ o(|x_n - x|^2) \end{bmatrix} \right). \tag{3.154}$$

Otprije znamo da je

$$\frac{1}{n} X_x^T W_x X_x = \begin{bmatrix} s_0(x) & s_1(x) \\ s_1(x) & s_2(x) \end{bmatrix}, \tag{3.155}$$

a sada imamo još

$$\frac{1}{n} X_x^T W_x \begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{bmatrix} = \begin{bmatrix} s_2(x) \\ s_3(x) \end{bmatrix}. \tag{3.156}$$

Za unutarnju točku x lema (3.2.1) uz supstituciju $u = \frac{s-x}{h}$ daje

$$s_r(x) = \int_0^1 (s - x)^r K(\frac{s - x}{h}) ds + O(n^{-1}) \tag{3.157}$$

$$= \int_{-\frac{x}{h}}^{\frac{1-x}{h}} h^r u^r K(u) du + O(n^{-1}) = h^r \int_{-1}^{1} u^r K(u) du + O(n^{-1}). \tag{3.158}$$

Za simetričnu jezgru K funkcija $u \mapsto u^r K(u)$ je parna za paran r i neparna za neparni r pa uz svojstva jezgre K imamo

$$s_0(x) = 1 + O(n^{-1}) \tag{3.159}$$

$$s_1(x) = O(n^{-1}) \tag{3.160}$$

$$s_2(x) = h^2 M_2 + O(n^{-1}) \tag{3.161}$$

$$s_3(x) = O(n^{-1}). \tag{3.162}$$

Uvrštavanjem (3.159) u (3.155) i (3.156) dobijemo

$$\frac{1}{n} X_x^T W_x X_x = \begin{bmatrix} 1 + O(n^{-1}) & O(n^{-1}) \\ O(n^{-1}) & h^2 M_2 + O(n^{-1}) \end{bmatrix} \tag{3.163}$$

$$\frac{1}{n} X_x^T W_x \begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{bmatrix} = \begin{bmatrix} h^2 M_2 + O(n^{-1}) \\ O(n^{-1}) \end{bmatrix}. \tag{3.164}$$

Kada sve uvrstimo u (3.154) dobijemo tvrdnju leme

$$\mathbb{E}(\hat{r}(x)) - r(x) = \frac{h^2}{2} r''(x) M_2 + o(h^2) + O(n^{-1}). \blacksquare \tag{3.165}$$

Lema 3.2.3. *Pod navedenim pretpostavkama,*

$$\mathbb{V}\mathrm{ar}(\hat{r}(x)) = \frac{\sigma^2 V}{nh} + o((nh)^{-1}), \tag{3.166}$$

gdje je $V = \int_{-1}^{1} K(x)^2 dx$ *kao i prije.*

Skica dokaza. Ako s L označimo matricu $e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x$, lako dobijemo izraz za varijancu procjene u nekom fiksnom x.

$$\hat{r}(x) = LY = \sum_{i=1}^{n} L_i(x) Y_i \tag{3.167}$$

$$\mathbb{E}(\hat{r}(x)^2) = \mathbb{E}\Big(\sum_{i=1}^{n} \sum_{j=1}^{n} L_i(x) L_j(x) Y_i Y_j \Big) \tag{3.168}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} L_i(x) L_j(x) r(x_i) r(x_j) + \sum_{i=1}^{n} \sigma^2 L_i(x)^2 \tag{3.169}$$

$$\mathbb{E}(\hat{r}(x))^2 = \Big(\sum_{i=1}^{n} L_i(x) r(x_i) \Big)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} L_i(x) L_j(x) r(x_i) r(x_j) \tag{3.170}$$

$$\implies \mathbb{V}\mathrm{ar}(\hat{r}(x)) = \sigma^2 \sum_{i=1}^{n} L_i(x)^2 = \sigma^2 L L^T \tag{3.171}$$

$$= \sigma^2 e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x^2 X_x (X_x^T W_x X_x)^{-1} e_1 \tag{3.172}$$

Izraz (3.172) vrijedi u općenitom slučaju. Ako matricu $X_x^T W_x^2 X_x$ raspišemo za lokalni linearni izglađivač, imamo

$$\frac{\sigma^2}{n} X_x^T W_x^2 X_x = \frac{\sigma^2}{n} X_x^T W_x W_x X_x \tag{3.173}$$

$$= \frac{\sigma^2}{n} \begin{bmatrix} K(\frac{x_1 - x}{h}) & \dots & K(\frac{x_n - x}{h}) \\ (x_1 - x) K(\frac{x_1 - x}{h}) & \dots & (x_n - x) K(\frac{x_n - x}{h}) \end{bmatrix} \begin{bmatrix} K(\frac{x_1 - x}{h}) & (x_1 - x) K(\frac{x_1 - x}{h}) \\ \vdots & \vdots \\ K(\frac{x_n - x}{h}) & (x_n - x) K(\frac{x_n - x}{h}) \end{bmatrix} \tag{3.174}$$

$$= \frac{\sigma^2}{n} \begin{bmatrix} \sum_{i=1}^{n} K(\frac{x_i - x}{h})^2 & \sum_{i=1}^{n} (x_i - x) K(\frac{x_i - x}{h})^2 \\ \sum_{i=1}^{n} (x_i - x) K(\frac{x_i - x}{h})^2 & \sum_{i=1}^{n} (x_i - x)^2 K(\frac{x_i - x}{h})^2 \end{bmatrix}. \tag{3.175}$$

Aproksimacijama analognim onima u Lemi (3.2.1) i (3.159) može se pokazati

$$\frac{\sigma^2}{n} X_x^T W_x^2 X_x = \begin{bmatrix} h^{-1}\sigma^2 V + o(h^{-1}) & O(n^{-1}) \\ O(n^{-1}) & h\sigma^2 J_2(K^2) + O(n^{-1}) \end{bmatrix}, \tag{3.176}$$

pri čemu je $J_2(f) = \int_{-1}^{1} x^2 f(x)dx$ kao i prije. Uvrštavanjem tog rezultata i prethodno dobivenog izraza za $X_x^T W_x X_x$ slijedi tvrdnja Leme. ■

Promotrimo li rezultate (3.2.2) i (3.2.3), vidimo da su oni jednaki onima za lokalne izglađivače jezgrama, u smislu jednakog izraza za pristranost i varijancu izglađivača i jednakog reda konvergencije u unutarnjim točkama.
Navedeni rezultati odnose se na lokalni linearni izglađivač u unutarnjim točkama. Općenito, može se pokazati da lokalni polinomni izglađivači neparnog stupnja p imaju pristranost reda $O(h^{p+1})$ za razliku od onih parnog stupnja čija je pristranost reda $O(h^{p+2})$. Nadalje, pristranost izglađivača parnog stupnja ovisi o dizajnu, odnosno o gustoći f točaka kovarijate, dok pristranost izglađivača neparnog stupnja ne ovisi i u tom smislu su prilagodljiviji dizajnu. Varijanca je u oba slučaja reda $O((nh)^{-1})$. Nadalje, u slučaju graničnih točaka lokalno polinomno izglađivanje neparnog stupnja automatski eliminira graničnu pristranost u smislu da je red pristranosti graničnih točaka jednak unutrašnjim. To je velika prednosti u odnosu na izglađivače jezgrama kod kojih smo to morali korigirati posebno graničnim jezgrama.

Poglavlje 4

Metode globalne regresije - splajnovi

Neka je zadan regresijski model (1). Kod lokalnog polinomnog izglađivanja i ranije vidjeli smo da je izglađivač određen optimizacijskim problemom kojeg odaberemo i klasom funkcija na kojoj tražimo njegovo rješenje. Odaberimo nešto općenitiji skup funkcija, tzv. Sobolovljev prostor funkcija na segmentu $[\,a,b]$ u oznaci $W_2^m[\,a,b]$:

$$W_2^m[\,a,b] = \{f \text{ } m\text{-puta derivabilne funkcije} : f, f', \ldots, f^{(m-1)} \in C([\,a,b]\,), \tag{4.1}$$

$$\int_a^b (f^{(m)}(x))^2 dx < \infty\}. \tag{4.2}$$

Definirajmo izglađivač $\hat{r}$ kao rješenje minimizacijskog problema

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - r(x_i))^2 + \lambda \int_a^b (r^{(m)}(x))^2 dx \tag{4.3}$$

po $r \in W_2^m[\,a,b]$. Takav izglađivač koji minimizira (4.3) nazivamo polinomni splajn. Kriterij (4.3) predstavlja ravnotežu između vjerne procjene uzorka i glatkoće funkcije. Ona je regulirana parametrom λ. Prvi dio izraza je već spomenut kao LS u (1.12), a drugi dio je prirodna mjera glatkoće za funkcije iz $W_2^m[\,a,b]$. Primjetimo da kada je $\lambda = 0$ izglađivač koji se dobije interpolira točke uzorka jer je prostor $W_2^m[\,a,b]$ beskonačnodimenzionalan. S druge strane, kada $\lambda \to \infty$ naglasak je na glatkoći funkcije i svi izglađivači kojima $\int_a^b \hat{r}^{(m)}(x)^2 dx$ poprima veliku vrijednost se jako "kažnjavaju". Zato je rješenje u tom slučaju izglađivač linearne polinomne regresije stupnja $m-1$. Još jedna motivacija za ovako odabran regresijski problem je sljedeća. Za svaki $x \in [\,a,b]$ po Taylorovom teoremu je

$$r(x) = \sum_{k=1}^{m-1} \frac{r^{(k)}(a)}{k!}(x-a)^k + \int_a^x \frac{(x-u)^{m-1}}{(m-1)!} r^{(m)}(u)du. \tag{4.4}$$

Sjetimo se da kod modela polinomne regresije u (2.23) pretpostavljamo da je $\text{Rem}(x) = \int_a^x \frac{(x-u)^{m-1}}{(m-1)!} r^{(m)}(u)du$ zanemariv. Ovdje je ideja iščitati iz podataka koliko velik $\text{Rem}(x)$ smije biti. Naime, može se pokazati da je pod određenim uvjetima $J_m(r) = \int_a^b r^{(m)}(x)^2 dx$ jedna moguća mjera udaljenosti između izglađivača i polinoma stupnja $m-1$, odnosno mjera odstupanja izglađivača od polinomnog modela. To se može vidjeti iz relacije

$$\text{Rem}(x) = \int_a^x \frac{(x-u)^{m-1}}{(m-1)!} r^{(m)}(u)du = \int_a^b \frac{(x-u)_+^{m-1}}{(m-1)!} r^{(m)}(u)du \tag{4.5}$$

$$\leq \int_a^b (r^{(m)}(u))^2 du \int_a^x \frac{(x-u)^{2(m-1)}}{((m-1)!)^2} du \leq \frac{(b-a)^{2m-1} J_m(r)}{(2m-1)((m-1)!)^2} \tag{4.6}$$

$$\implies \max_{a \leq x \leq b} \text{Rem}(x) \leq C J_m(r), \tag{4.7}$$

pri čemu prva nejednakost slijedi iz Cauchy-Schwarzove nejednakosti. Ako bismo željeli minimizirati LS izraz $\sum_{i=1}^n (Y_i - r(x))^2$ pod uvjetom

$$\int_a^b (r^{(m)}(u))^2 du \leq \rho, \tag{4.8}$$

za neku odabranu konstantu $\rho \geq 0$, uvođenjem Lagrangeovog multiplikatora λ vidimo da je to ekvivalentno minimizaciji $\text{LS} + \lambda(J_m(r) - \rho)$, što je zapravo ekvivalentno minimizaciji (4.3).
Iz toga da polinomni splajn minimizira kriterij (4.3) nije odmah jasno kako on izgleda. Zato ćemo definirati splajn i dokazati da je on rješenje zadanog problema.

Definicija 4.0.1. *Polinomni splajn reda r na segmentu $[a, b]$ s čvorovima $\epsilon_1, \ldots, \epsilon_k$ takvih da $a \leq \epsilon_1 \leq \ldots \leq \epsilon_k \leq b$ je realna funkcija s takva da vrijedi:*

(i) s je polinom reda r na svakom intervalu $[\epsilon_i, \epsilon_{i+1}\rangle$ (odnosno s je po dijelovima polinomna funkcija)

(ii) s je klase C^{r-2}, odnosno $r-2$ puta neprekidno diferencijabilna

(iii) s ima derivaciju reda $r-1$ koja ima prekide u točkama $\epsilon_1, \ldots, \epsilon_k$ (to je step funkcija sa skokovima u čvorovima).

Dakle, polinomni splajn je po dijelovima polinomna funkcija koja je po segmentima "glatko slijepljena". To je ekvivalentno tome da za $x \in [a, b]$ s možemo prikazati kao

$$s(x) = \sum_{j=0}^{r-1} \theta_j x^j + \sum_{j=1}^{k} \eta_j (x - \epsilon_j)_+^{r-1} \tag{4.9}$$

za neke koeficijente $\theta_0, \ldots, \theta_{r-1}, \eta_1, \ldots, \eta_k$, pri čemu je $x_+ = max\{x, 0\}$ i notacija $(x-\epsilon_j)_+^{r-1}$ podrazumijeva $((x-\epsilon_j)_+)^{r-1}$. Primjetimo da funkcija s iz (4.9) zadovoljava definiciju (4.0.1). Nadalje, označimo sa $S^r(\epsilon_1, \ldots, \epsilon_k)$ sve polinomne splajnove oblika (4.9). Tada je $S^r(\epsilon_1, \ldots, \epsilon_k)$ vektorski prostor, a jer su funkcije $1, x, \ldots, x^{r-1}, (x - \epsilon_1)_+^{r-1}, \ldots, (x - \epsilon_k)_+^{r-1}$ nezavisne, slijedi da taj prostor ima dimenziju $k+r$. Također, iz (4.0.1) možemo vidjeti da za jednoznačno određen polinomni splajn treba $r(k+1)$ koeficijenata od kojih je $k(r-1)$ određeno s uvjetima neprekidnosti, dakle prostor svih splajnova stvarno ima dimenziju $k+r$ pa su definicija (4.0.1) i izraz (4.9) konzistenti. Međutim, rješenje (4.3) je posebna vrsta polinomnog splajna, to je tzv. prirodni splajn.

Definicija 4.0.2. *Prirodni splajn je polinomni splajn reda $r = 2m$, s čvorovima $x_1, \ldots, x_n$ koji zadovoljava dodatni uvjet*

(iv) s je polinom reda m izvan $[x_1, x_n]$.

Da bismo jednoznačno odredili polinomni splajn potrebno je $2m(n-2)$ koeficijenata za unutrašnje segmente i $2m$ za rubne, dakle ukupno $2mn$ koeficijenata. Od toga je iz uvjeta neprekidnosti određeno $n(2m-1)$, što znači da prostor prirodnih splajnova ima dimenziju n. Označimo ga s $NS^{2m}(x_1, \ldots, x_n)$. Jasno je da je to potprostor od $S^{2m}(x_1, \ldots, x_n)$. Također, odmah vidimo da prirodni splajn $s(x) = \sum_{j=0}^{r-1} \theta_j x^j + \sum_{j=1}^{k} \eta_j (x - x_j)_+^{r-1}$ mora zadovoljavati

$$\theta_m = \ldots = \theta_{2m-1} = 0. \tag{4.10}$$

Od sada nadalje pretpostavljamo da je $a = 0, b = 1$, odnosno promatramo splajnove i prirodne splajnove na $[0, 1]$ i svi rezultati su rađeni pod tim uvjetima. Sljedeća lema daje jedno korisno svojstvo prirodnih splajnova.

Lema 4.0.3. *Neka je $s_1, \ldots, s_n$ baza za $NS^{2m}(x_1, \ldots, x_n)$. Tada postoje koeficijenti $\theta_{0j}, \ldots, \theta_{(m-1)j}, \eta_{1j}, \ldots, \eta_{nj}$ za $j = 1, \ldots, n$ takvi da vrijedi*

$$s_j(x) = \sum_{i=0}^{m-1} \theta_{ij} x^i + \sum_{i=1}^{n} \eta_{ij} (x - x_i)_+^{2m-1}, \tag{4.11}$$

za svaki $j = 1, \ldots, n$. Ako je $s(x) = \sum_{j=1}^{n} b_j s_j(x)$ prirodni splajn i $f \in W_2^m[a, b]$, tada je

$$\int_0^1 f^{(m)}(x) s^{(m)}(x) dx = (-1)^m (2m-1)! \sum_{i=1}^{n} f(x_i) \sum_{j=1}^{n} b_j \eta_{ij}. \tag{4.12}$$

Dokaz.
Prva tvrdnja leme slijedi iz dviju već pokazanih činjenica. Prvo, $NS^{2m}(x_1,\ldots,x_n) \leq S^{2m}(x_1,\ldots,x_n)$ pa se svaki prirodni splajn može napisati kao linearna kombinacija baze za polinomne splajnove. Drugo, (4.10) daje nužne uvjete za neke od koeficijenata u tom raspisu. Ostaje još pokazati jednakost (4.12). Za početak, primijetimo da je $s^{(m+j)}$ jednaka 0 izvan $[\,x_1, x_n]$ za $j = 0,\ldots,m-1$. Parcijalnom integracijom dobijemo:

$$\int_0^1 f^{(m)}(x)s^{(m)}(x)dx = f^{(m-1)}(x)s^{(m)}(x)\Big|_0^1 - \int_0^1 f^{(m-1)}(x)s^{(m+1)}(x)dx \tag{4.13}$$

$$= (-1)^1 \int_0^1 f^{(m-1)}(x)s^{(m+1)}(x)dx = \ldots = (-1)^{m-1}\int_0^1 f'(x)s^{(2m-1)}(x)dx \tag{4.14}$$

$$= (-1)^{m-1}\sum_{i=1}^{n-1}\int_{x_i}^{x_{i+1}} f'(x)s^{(2m-1)}(x)dx \tag{4.15}$$

Nadalje, $s(x) = \sum_{j=1}^n b_j s_j(x)$ pa je

$$s^{(2m-1)}(x) = (2m-1)!\sum_{j=1}^n b_j \sum_{k=1}^n \eta_{kj}\mathbb{1}_{[}x_k, 1](x). \tag{4.16}$$

Uvrštavanjem u (4.15) imamo:

$$\int_0^1 f^{(m)}(x)s^{(m)}(x)dx = (-1)^{m-1}(2m-1)!\sum_{i=1}^{n-1}\int_{x_i}^{x_{i+1}} f'(x)\Big[\sum_{j=1}^n b_j\sum_{k=1}^i \eta_{kj}\Big]dx \tag{4.17}$$

$$= (-1)^{m-1}(2m-1)!\sum_{i=1}^{n-1}(f(x_{i+1}) - f(x_i))\sum_{j=1}^n b_j\sum_{k=1}^i \eta_{kj} \tag{4.18}$$

$$= (-1)^{m-1}(2m-1)!\sum_{j=1}^n b_j\Big[\sum_{i=1}^{n-1}(f(x_{i+1}) - f(x_i))\sum_{k=1}^i \eta_{kj}\Big] \tag{4.19}$$

$$= (-1)^{m-1}(2m-1)!\sum_{j=1}^n b_j\Big[\sum_{i=1}^{n-1} f(x_{i+1})\sum_{k=1}^i \eta_{kj} - \sum_{i=1}^{n-1} f(x_i)\sum_{k=1}^i \eta_{kj}\Big] \tag{4.20}$$

$$= (-1)^{m-1}(2m-1)!\sum_{j=1}^{n} b_j\Big[\sum_{i=2}^{n} f(x_i)\sum_{k=1}^{i-1}\eta_{kj} - \sum_{i=1}^{n-1} f(x_i)\sum_{k=1}^{i}\eta_{kj}\Big] \tag{4.21}$$

$$= (-1)^{m-1}(2m-1)!\sum_{j=1}^{n} b_j\Big[\sum_{i=2}^{n-1} f(x_i)\Big(\sum_{k=1}^{i-1}\eta_{kj} - \sum_{k=1}^{i}\eta_{kj}\Big) \tag{4.22}$$

$$+ f(x_n)\sum_{k=1}^{n-1}\eta_{kj} - f(x_1)\eta_{1j}\Big] \tag{4.23}$$

$$= (-1)^{m}(2m-1)!\sum_{j=1}^{n} b_j\Big[\sum_{i=1}^{n-1} f(x_i)\eta_{ij} - f(x_n)\sum_{k=1}^{n-1}\eta_{kj}\Big]. \tag{4.24}$$

Uzmimo $x > x_n$. Kao što je rečeno, za $j = 1, \ldots, n$ je $s_j^{(2m-1)}(x) = 0$, a zbog $s_j^{(2m-1)}(x) = \sum_{k=1}^{n}\eta_{kj}$ imamo

$$\sum_{k=1}^{n}\eta_{kj} = 0 \implies \sum_{k=1}^{n-1}\eta_{kj} = -\eta_{nj}. \tag{4.25}$$

Uvrštavanjem (4.25) u (4.24) dobijemo

$$\int_0^1 f^{(m)}(x)s^{(m)}(x)dx = (-1)^m(2m-1)!\sum_{j=1}^{n} b_j\sum_{i=1}^{n} f(x_i)\eta_{ij}, \tag{4.26}$$

što zamjenom poretka sumacije daje tvrdnju leme. ■

Sljedeća lema govori o svojstvu interpolacijske optimalnosti prirodnih splajnova.

Lema 4.0.4. *Neka je $s_1, \ldots, s_n$ baza za $NS^{2m}(x_1, \ldots, x_n)$ kojoj pridružujemo matricu dizajna $X = [\,s_j(x_i)\,]_{\,i,j=1,\ldots,n}$ i neka je $a = (a_1, \ldots, a_n)^T$ vektor konstanti. Tada, ako je $n \geq m$, jedinstvena funkcija koja minimizira $J_m(f)$ na skupu funkcija iz $W_2^m[\,0,1\,]$, a koje zadovoljavaju $f(x_i) = a_i$, je prirodni splajn $s(x) = \sum_{j=1}^{n} b_j s_j(x)$, pri čemu je $b = (b_1, \ldots, b_n)^T$ rješenje sustava $Xc = a$. Posebno, matrica X je punog ranga n.*

Dokaz.
Pokažimo prvo da je X punog ranga. Za to je dovoljno dokazati da je $Xc = 0$ ako i samo ako je $c = 0$. Neka je $Xc = 0$ i pridružimo vektoru c prirodni splajn $s = \sum_{j=1}^{n} c_j s_j$. Tada je

$$\begin{bmatrix} s(x_1) \\ \vdots \\ s(x_n) \end{bmatrix} = Xc = 0. \tag{4.27}$$

Po lemi (4.0.3) vrijedi

$$\int_0^1 s^{(m)}(x)^2 dx = (-1)^m (2m-1)! \sum_{i=1}^{n} s(x_i) \sum_{j=1}^{n} b_j \eta_{ij} = 0. \tag{4.28}$$

Iz toga slijedi da je s polinom reda m na $[\,0,1\,]$ koji ima vrijednost 0 u $n \geq m$ točaka. To znači da je $s = \sum_{j=1}^{n} c_j s_j = 0$. Budući da je $\{s_1, \ldots, s_n\}$ baza, mora biti $c = 0$. Neka je sada $s(x) = \sum_{j=1}^{n} b_j s_j(x)$ pri čemu je $b = X^{-1}a$. Primijetimo da vrijedi $\sum_{j=1}^{n} s_j(x_i) = a_i$, odnosno s interpolira točke a_i. Uzmimo proizvoljnu funkciju $g \in W_2^m[\,0,1\,]$ koja zadovoljava $g(x_i) = a_i$ za $i = 1, \ldots, n$. Tada vrijedi:

$$J_m(g-s) = \int_0^1 (g^{(m)}(x) - s^{(m)}(x))^2 dx \tag{4.29}$$

$$= J_m(g) + J_m(s) - 2\int_0^1 g^{(m)}(x) s^{(m)}(x) dx \tag{4.30}$$

$$\implies J_m(g) = J_m(s) - 2J_m(s) + J_m(g-s) + 2\int_0^1 g^{(m)}(x) s^{(m)}(x) dx \tag{4.31}$$

$$= J_m(s) + 2\int_0^1 s^{(m)}(x)[\,g^{(m)}(x) - s^{(m)}(x)\,]\,dx + J_m(g-s). \tag{4.32}$$

Po lemi (4.0.3) i zbog $g(x_i) = s(x_i)$ vrijedi

$$\int_0^1 s^{(m)}(x)[\,g^{(m)}(x) - s^{(m)}(x)\,]\,dx \tag{4.33}$$

$$= (-1)^m (2m-1)! \sum_{i=1}^{n} [\,g(x_i) - s(x_i)\,] \sum_{j=1}^{n} b_j \eta_{ij} = 0. \tag{4.34}$$

Izraz (4.32) daje:

$$J_m(g) = J_m(s) + J_m(g-s). \tag{4.35}$$

Iz toga slijedi da je $J_m(g) > J_m(s)$ osim ako je $J_m(g-s) = 0$. Dakle, $\int_0^1 (g^{(m)}(x) - s^{(m)}(x))^2 dx = 0$, iz čega slijedi da je $g^{(m)} - s^{(m)}$ jednako 0 na intervalu $[\,0,1\,]$. No, tada $g - s$ mora biti polinom reda m koji ima $n \geq m$ nultočaka, ali tada je $g - s \equiv 0$. Dakle, vrijedi tvrdnja leme. ■

Koristeći lemu (4.0.4) sljedeći teorem pokazuje da je rješenje problema (4.3) prirodni splajn i daje eksplicitni oblik izglađivača.

Teorem 4.0.5. *Neka je $s_1, \ldots, s_n$ baza za $NS^{2m}(x_1, \ldots, x_n)$ kojoj pridružujemo matricu dizajna $X = [\, s_j(x_i)]_{\,i,j=1,\ldots,n}$. Ako je $n \geq m$, jedinstveno rješenje minimizacijskog problema (4.3) je $\hat{r} = \sum_{j=1}^{n} b_j s_j$, pri čemu je $b = (b_1, \ldots, b_n)^T$ jedinstveno rješenje sustava*

$$(X^T X + n\lambda\Omega)c = X^T Y, \tag{4.36}$$

po varijabli c i matrica Ω je definirana s

$$\Omega = \Big[\int_0^1 s_i^{(m)}(x) s_j^{(m)}(x) dx \Big]_{i,j=1,\ldots,n}. \tag{4.37}$$

Dokaz.
Uzmemo li proizvoljnu funkciju $f \in W_2^m[\,0,1]$, lema (4.0.4) pokazuje da ćemo ako zamijenimo f s prirodnim splajnom koji se podudara s f u točkama $x_i, i = 1, \ldots, n$, smanjiti $J_m(f)$, odnosno drugi dio kriterija (4.3), dok će LS izraz ostati nepromijenjen. Drugim riječima, za proizvoljnu funkciju $f \in W_2^m[\,0,1]$ postoji prirodni splajn koji strogo smanjuje kriterij (4.3) ako ona sama nije prirodni splajn. Iz toga slijedi da je izglađivač koji je rješenje optimizacijskog problema (4.3) upravo prirodni splajn. Dakle, možemo izglađivač tražiti na skupu funkcija oblika $s = \sum_{j=1}^{n} c_j s_j$, minimizacijom po c. Zapišimo kriterij (4.3) matrično:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - r(x_i))^2 + \lambda \int_0^1 (r^{(m)}(x))^2 dx \tag{4.38}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{n} c_j s_j(x_i))^2 + \lambda \int_0^1 (\sum_{j=1}^{n} c_j s_j^{(m)}(x))^2 dx \tag{4.39}$$

$$= \frac{1}{n}(Y - Xc)^T(Y - Xc) + \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \int_0^1 s_i^{(m)}(x) s_j^{(m)}(x) dx \tag{4.40}$$

$$= \frac{1}{n}(Y - Xc)^T(Y - Xc) + \lambda c^T \Omega c \tag{4.41}$$

$$= \frac{1}{n}(Y^T Y - Y^T X c - c^T X Y + c^T X^T X c) + \lambda c^T \Omega c \tag{4.42}$$

Deriviranjem po c dobije se

$$-2Y^T X - Y^T X + 2c^T X^T X + 2n\lambda c^T \Omega = 0 \tag{4.43}$$

$$Y^T X = c^T (X^T X + n\lambda\Omega) \tag{4.44}$$

$$\Longrightarrow (X^T X + n\lambda\Omega)c = X^T Y. \tag{4.45}$$

Još samo treba pokazati invertibilnost matrice $X^T X + n\lambda\Omega$. X je matrica punog ranga po lemi (4.0.4) pa je zbog toga matrica $X^T X$ simetrična i pozitivno definitna.

Nadalje, zbog $J_m(s) = c^T \Omega c \geq 0$ slijedi da je matrica Ω pozitivno semidefinitna. Iz toga slijedi da je matrica $X^T X + n\lambda\Omega$ pozitivno definitna pa posebno i invertibilna. Stoga, sustav (4.45) ima jedinstveno rješenje. ■

Iz teorema (4.0.5) vidimo da je izglađivač splajnovima oblika $\hat{r} = \sum_{j=1}^{n} b_j s_j$, pri čemu je $b = (X^T X + n\lambda\Omega)^{-1} X^T Y$. Ako uvedemo oznaku $S = (s_1, \ldots, s_n)$, slijedi da je

$$\hat{r}(x) = S(x)b = S(x)(X^T X + n\lambda\Omega)^{-1} X^T Y \tag{4.46}$$

pa vidimo da zadovoljava definiciju linearnog izglađivača. Posebno, $\begin{bmatrix} S(x_1) \\ \vdots \\ S(x_n) \end{bmatrix} = X$

pa je na točkama uzorka

$$\begin{bmatrix} \hat{r}(x_1) \\ \vdots \\ \hat{r}(x_n) \end{bmatrix} = \begin{bmatrix} S(x_1) \\ \vdots \\ S(x_n) \end{bmatrix} b = X(X^T X + n\lambda\Omega)^{-1} X^T Y, \tag{4.47}$$

što znači da je matrica izglađivanja $L = X(X^T X + n\lambda\Omega)^{-1} X^T$.

Iako lema (4.0.3) govori nešto o obliku prirodnih splajnova, još nismo naveli nijednu bazu za prostor $NS^{2m}(x_1, \ldots, x_n)$. Jedan primjer baze koji je više teoretskog karaktera, u smislu da je korisna za izračun greške, ali se u praksi koriste numerički optimalnije baze, je Demmler-Reinschova baza. Ona ima zatvoren oblik samo za slučaj $m = 1$, odnosno za linearne splajnove pa ovdje navodimo samo taj poseban slučaj.
Neka je zadan dizajn $x_i = \frac{2i-1}{2n}$ za $i = 1, \ldots, n$. Demmler-Reinschova baza $\{s_1, \ldots, s_n\}$ interpolira konstantu i vrijednosti funkcija $\sqrt{2}\cos(j\pi x)$ za $j = 1, \ldots, n-1$ respektivno u točkama $x_i, i = 1, \ldots, n$. Primjetimo da smo time za svaki element baze zadali n interpolacijskih uvjeta i time ga jednoznačno odredili. Eksplicitno baza glasi:

$$s_1(x) \equiv 1 \tag{4.48}$$

$$s_{j+1}(x) = \begin{cases} \sqrt{2}\cos(j\pi x_1), & 0 \leq x < x_1 \\ \sqrt{2}\cos(j\pi x_i) + \sqrt{2}\frac{x - x_i}{x_{i+1} - x_i}\left[\cos(j\pi x_{i+1}) - \cos(j\pi x_i)\right], & x_i \leq x < x_{i+1}, \\ & i = 1, \ldots, n-1 \\ \sqrt{2}\cos(j\pi x_n), & x_n \leq x \leq 1 \end{cases} \tag{4.49}$$

$$j = 1, \ldots, n-1. \tag{4.50}$$

Može se pokazati da ta baza dijagonalizira matrice X^TX i Ω te se u tom slučaju linearni izglađivač splajnovima može napisati kao

$$\hat{r}(x) = \frac{1}{n}\sum_{i=1}^{n}(1 + \sum_{j=1}^{n-1}\frac{\cos(j\pi x_i)s_{j+1}(x)}{1+\lambda\gamma_j})Y_i = \frac{1}{n}\sum_{i=1}^{n} l(x, x_i)Y_i, \tag{4.51}$$

pri čemu je definirano $l(x,y) = 1 + \sum_{j=1}^{n-1}\frac{\cos(j\pi y)s_{j+1}(x)}{1+\lambda\gamma_j})$, a

$$\gamma_j = (2n\sin\left(\frac{j\pi}{2n}\right))^2, j = 1, \ldots, n-1 \tag{4.52}$$

su tzv. Demmler-Reinschove svojstvene vrijednosti.

Sljedeći teorem (vidjeti [2], str.253-254) daje ocjenu asimptotskog rizika izglađivača splajnovima.

Teorem 4.0.6. *Neka je* $r \in C^2([\,0,1\,])$ *. Ako je* r' *omeđena, barem jedan od* $r'(0)$,$r'(1)$ *je različit od* 0 *i* $n\lambda^{\frac{3}{2}} \to \infty$*, vrijedi*

$$R(\hat{r}) \sim \frac{\lambda^{\frac{3}{2}}}{2}[\,r'(0)^2 + r'(1)^2\,] + \frac{\sigma^2}{4n\sqrt{\lambda}}. \tag{4.53}$$

Ako je dodatno r'' *Lipschitz neprekidna,* $r'(0) = r'(1) = 0$ *i* $n\lambda^2 \to \infty$*, tada je*

$$R(\hat{r}) \sim \lambda^2\int_0^1 r''(x)^2dx + \frac{\sigma^2}{4n\sqrt{\lambda}}. \tag{4.54}$$

.

U prvom slučaju, odnosno kada vrijedi izraz (4.53), pokazuje se da je optimalni parametar izglađivanja λ reda $O(n^{-\frac{1}{2}})$, zbog čega optimalni rizik opada brzinom $O(n^{-\frac{3}{4}})$. U drugom slučaju optimalni parametar izglađivanja reda $O(n^{-\frac{2}{5}})$, a cjelokupni rizik opada brzinom $O(n^{-\frac{4}{5}})$.

Poglavlje 5

Odabir parametra izglađivanja

Vidjeli smo da veličine poput MSE i rizika kod svih navedenih linearnih izglađivača ovise o širini pojasa h, odnosno λ. Kraće, oba parametra ćemo zvati parametrom izglađivanja. Također, u uvodnom poglavlju naveli smo unakrsno vrednovanje kao jednu metodu odabira hiperparametara modela. Ovdje ćemo navesti još neke metode i pokazati rezultat vezan uz lakše računanje *CV-score*-a u slučaju lokalnog polinomnog izglađivanja i izglađivača splajnovima.
Navedimo prvo rezultat vezan uz unakrsno vrednovanje.

Lema 5.0.1. *Neka je $k \in \{1, \dots, n\}$ i $\hat{r}^{-k}$ izglađivač koji minimizira težinsku sumu kvadrata (3.110) za neki fiksni broj x ili globalnu penaliziranu sumu kvadrata (4.3) na uzorku $(x_1, Y_1), \dots, (x_{k-1}, Y_{k-1}), (x_{k+1}, Y_{k+1}), \dots, (x_n, Y_n)$. Tada $\hat{r}^{-k}$ minimizira isti kriterij na uzorku $(x_1, Y_1), \dots, (x_{k-1}, Y_{k-1}), (x_k, \hat{r}^{-k}(x_k)), (x_{k+1}, Y_{k+1}), \dots, (x_n, Y_n)$.*

Dokaz.
Pokažimo prvo rezultat u slučaju lokalnog polinomnog izglađivanja.

$$\sum_{i \neq k} K(\frac{x - x_i}{h})(Y_i - \hat{r}^{-k}(x_i))^2 + K(\frac{x - x_k}{h})(\hat{r}^{-k}(x_k) - \hat{r}^{-k}(x_k))^2 \tag{5.1}$$

$$= \sum_{i \neq k} K(\frac{x - x_i}{h})(Y_i - \hat{r}^{-k}(x_i))^2 \leq \sum_{i \neq k} K(\frac{x - x_i}{h})(Y_i - \hat{r}(x_i))^2 \tag{5.2}$$

$$\leq \sum_{i \neq k} K(\frac{x - x_i}{h})(Y_i - \hat{r}(x_i))^2 + K(\frac{x - x_k}{h})(\hat{r}^{-k}(x_k) - \hat{r}(x_k))^2, \tag{5.3}$$

pri čemu je $\hat{r}$ proizvoljni lokalni polinomni izglađivač istog stupnja kao $\hat{r}^{-k}$. Jednak

slijed zaključivanja daje isti rezultat za izglađivače splajnovima:

$$\sum_{i\neq k}(Y_i - \hat{r}^{-k}(x_i))^2 + (\hat{r}^{-k}(x_k) - \hat{r}^{-k}(x_k))^2 + \lambda J_m(\hat{r}^{-k}) \tag{5.4}$$

$$= \sum_{i\neq k}(Y_i - \hat{r}^{-k}(x_i))^2 + \lambda J_m(\hat{r}^{-k}) \leq \sum_{i\neq k}(Y_i - \hat{r}(x_i))^2 + \lambda J_m(\hat{r}) \tag{5.5}$$

$$\leq \sum_{i\neq k}(Y_i - \hat{r}(x_i))^2 + (\hat{r}^{-k}(x_k) - \hat{r}(x_k))^2 + \lambda J_m(\hat{r}). \tag{5.6}$$

■

Pogledajmo kakvu nam praktičnu korist lema (5.0.1) daje u slučaju LOO unakrsnog vrednovanja. Ako za neki fiksni parametar izglađivanja h želimo izračunati procjenu prediktivnog rizika, trebamo n puta pronaći izglađivač na uzorcima od kojih je svaki bez jednog elementa početnog uzorka i broj kojeg želimo dobiti je

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{r}^{-i}(x_i))^2, \tag{5.7}$$

pri čemu ovdje implicitno podrazumijevamo ovisnost svakog izglađivača $\hat{r}$ o h. Lema (5.0.1) kaže da se svaki takav izglađivač $\hat{r}^{-k}$ može dobiti minimizacijom istog kriterija na uzorku koji se od početnog razlikuje po tome što mu je k-ta vrijednost varijable odaziva promijenjena u $\hat{r}^{-k}(x_k)$. To nam je bitno jer u slučaju lokalnih polinoma i splajnova matrica izglađivanja, odnosno težine koje se pridaju pojedinom Y_i ne ovise o njima samima već samo o dizajnu, odnosno o vrijednostima kovarijate. Stoga, izglađivač $\hat{r}^{-k}$ dobiven na uzorku $(x_1, Y_1), \ldots, (x_k, \hat{r}^{-k}(x_k)), \ldots, (x_n, Y_n)$ i izglađivač $\hat{r}$ dobiven na uzorku $(x_1, Y_1), \ldots, (x_k, Y_k), \ldots, (x_n, Y_n)$ imaju jednake matrice izglađivanja. Iz toga slijedi:

$$\hat{r}(x_k) = \sum_{j=1}^{n} L_{kj}Y_j \tag{5.8}$$

$$\hat{r}^{-k}(x_k) = \sum_{j\neq k} L_{kj}Y_j + L_{kk}\hat{r}^{-k}(x_k) \tag{5.9}$$

$$\implies \hat{r}^{-k}(x_k) - \hat{r}(x_k) = L_{kk}\hat{r}^{-k}(x_k) - L_{kk}Y_k \tag{5.10}$$

$$\hat{r}^{-k}(x_k) = \frac{\hat{r}(x_k) - L_{kk}Y_k}{1 - L_{kk}}. \tag{5.11}$$

Konačno, slijedi

$$Y_k - \hat{r}^{-k}(x_k) = \frac{Y_k - L_{kk}Y_k + L_{kk}Y_k - \hat{r}(x_k)}{1 - L_{kk}} \tag{5.12}$$

$$= \frac{Y_k - \hat{r}(x_k)}{1 - L_{kk}}. \tag{5.13}$$

Uvrštavanjem u izraz za unakrsno vrednovanje dobijemo

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_k - \hat{r}(x_k)}{1 - L_{kk}}\right)^2. \tag{5.14}$$

Iz ovoga vidimo da u slučaju LOO unakrsnog vrednovanja za fiksni h nije potrebno pronalaziti n izglađivača na n poduzoraka, već je dovoljno pronaći jedan izglađivač na cijelom uzorku da bismo dobili vrijednost $CV(h)$. To znatno ubrzava račun u slučaju velikih vrijednosti n. Kriterij (5.14) možemo koristiti općenito kao kriterij odabira parametra izglađivanja. Samo u slučaju nekih izglađivača, poput Gasser-Müllerovog, nemamo ovakvu interpretaciju.
Izraz (5.14) dobar je uvod u još jednu često korištenu metodu. To je generalizirano unakrsno vrednovanje definirano izrazom

$$GCV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_k - \hat{r}(x_k)}{1 - \frac{\nu}{n}}\right)^2, \tag{5.15}$$

pri čemu su $\nu = tr(L)$ generalizirani stupnjevi slobode. Drugim riječima, u izrazu (5.14) smo elemente dijagonale matrice izglađivanja zamijenili s njihovim prosjekom. Generalizirano unakrsno vrednovanje ima svojstvo invarijantnosti na ortogonalne transformacije od Y. Nadalje, vrijedi i teorem (vidjeti [2], str. 44, teorem 2.1):

Teorem 5.0.2. *Neka je $\nu = tr(L), \tilde{\nu} = tr(L^T L)$ i pretpostavimo da je $\nu < 1$. Tada je*

$$\frac{|\mathbb{E}(GCV(h)) - \mathrm{PR}(h)|}{\mathrm{R}(h)} \leq g(h), \tag{5.16}$$

pri čemu je

$$g(h) = \frac{2\nu + \frac{\nu^2}{\tilde{\nu}}}{(1-\nu)^2}. \tag{5.17}$$

Odnosno, kada je vrijednost $g(h)$ mala tada je i očekivana vrijednost generaliziranog unakrsnog vrednovanja približno jednaka prediktivnom riziku relativno prema veličini rizika.

Postavlja se pitanje, koliko su obično i generalizirano unakrsno vrednovanje efikasni u smislu brzine konvergencije. Neka je $\hat{h}$ vrijednost parametra dobivenu optimizacijom $CV(h)$ ili $GCV(h)$ kriterija. Označimo s $\hat{h}_0$ vrijednost parametra koja minimizira ARSS. Pokazuje se da

$$n^{\frac{1}{10}} \frac{\hat{h} - \hat{h}_0}{\hat{h}_0} \xrightarrow{d} \upsilon Z, \tag{5.18}$$

pri čemu je Z varijabla iz standardne normalne razdiobe, a υ^2 određuje varijancu i ovisi o σ^2,r i K. Ovaj rezultat pokazuje da $\hat{h}$ kao procjenitelj $\hat{h}_0$ konvergira brzinom $O(n^{\frac{-1}{10}})$.

Izbor jezgre

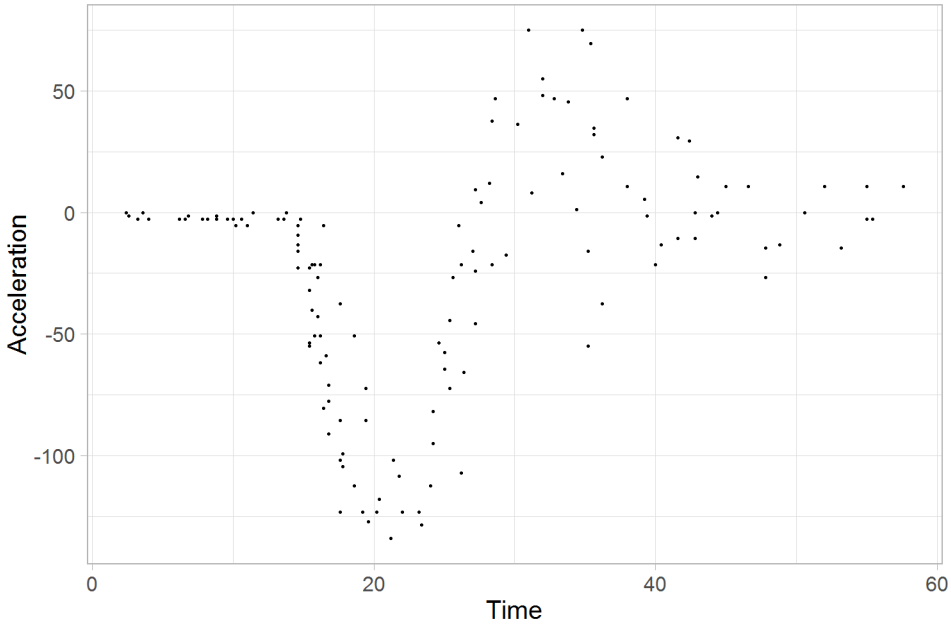
Već smo kod Nadaraya-Watson procjenitelja naveli da izbor jezgre u lokalnom izglađivanju nije toliko bitan jer su dobiveni izglađivači numerički dosta slični. Kao dodatni argument, recimo da želimo izabrati jezgru na način da minimiziramo asimptotski optimalan rizik kao u teoremu (3.1.10). Tada vidimo da nam je to ekvivalentno pronalasku jezgre K koja minimizira izraz $M_2(K)^{\frac{2}{5}} V(K)^{\frac{4}{5}}$. Jezgra koja je rješenje tog problema je Epanechnikova, a vrijednost tog izraza za nju iznosi $M_2^{\frac{2}{5}} V^{\frac{4}{5}} = \frac{3}{5\sqrt{5}}$. Zbog toga za proizvoljnu jezgru K možemo promatrati omjer $\frac{M_2(K)^{\frac{2}{5}} V(K)^{\frac{4}{5}}}{\frac{3}{5\sqrt{5}}}$ kao mjeru relativne neefikasnosti jezgre prema kriteriju optimizacije asimptotskog rizika. U donjoj tablici je navedena mjera neefikasnosti za neke jezgre.

Jezgra	Neefikasnost
Epanechnikova	1
Uniformna	1.0758
Biweight	1.0061
Triweight	1.0135
Gaussova	1.0513

Iz tablice možemo zaključiti da je asimptotski rizik neosjetljiv na izbor jezgre i da jezgru možemo birati na temelju stvari poput lakše izračunljivosti ili svojstava koja želimo za $\hat{r}$.

5.1 Primjer: *mcycle*

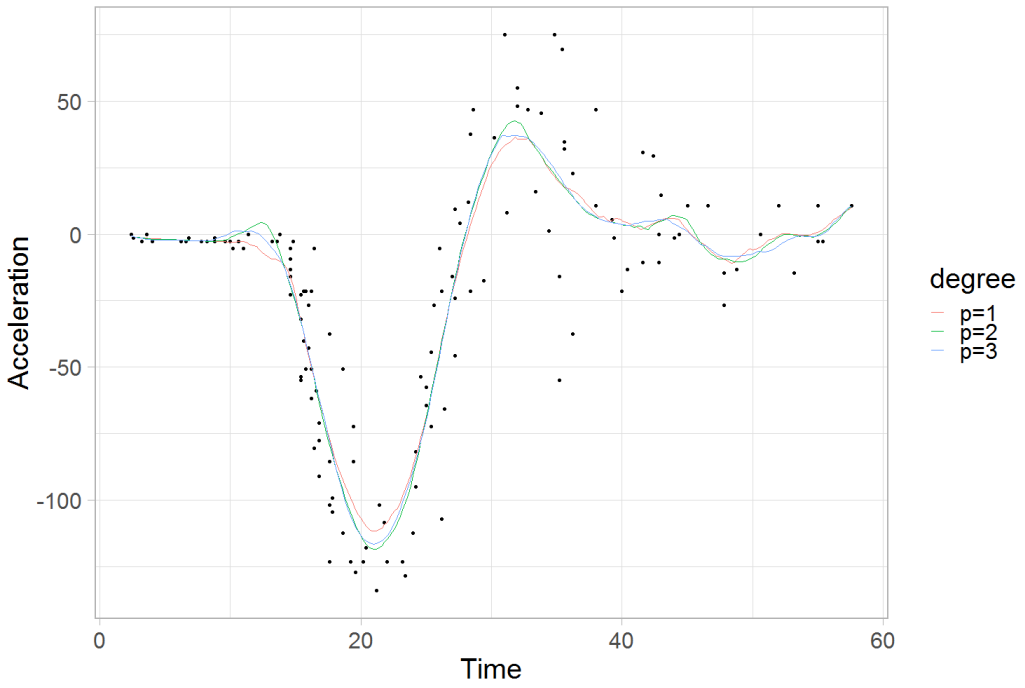
Preostaje nam još na primjeru pokazati primjenu obrađenih metoda. Odabrani podaci nalaze se pod nazivom "mcycle" u R-ovom paketu "MASS" Podaci su mjerenja ubrzanja glave u simuliranim motorističkim nesrećama koja su korištena za testiranja kaciga. Sastoji se od dvije varijable: vremena nakon udara u milisekundama i akceleracije u g. Podaci su prikazani na Slici 5.1.



Slika 5.1: mcycle podaci

Od metoda lokalne regresije odabrano je lokalno polinomno izglađivanje stupnja 1,2 i 3, a od metoda globalne regresije splajnovi. U svim slučajevima kao metoda odabira parametra izglađivanja korišteno je LOO unakrsno vrednovanje. Sve je rađeno u R-u. Lokalno polinomno izglađivanje implementirano je u paketu "locpol", a izglađivanje splajnovima u paketu "stats".

Slika 5.2 prikazuje lokalne polinomne izglađivače za koje je korištena Epanechnikova jezgra. Za lokalni kubični izglađivač optimalna širina pojasa je $h = 7.114456$, dok je procjena prediktivnog rizika $CV(h) = 542.1867$ i to je najbolji izglađivač po kriteriju unakrsnog vrednovanja. Za lokalni kvadratni izglađivač širina pojasa je $h = 5.717588$, a procjena rizika je $CV(h) = 546.9584$, dok je za lokalni linearni $h = 3.430307$ i $CV(h) = 575.0025$.



Slika 5.2: Usporedba lokalnih polinomnih izglađivača stupnja 1,2 i 3 na mcycle podacima

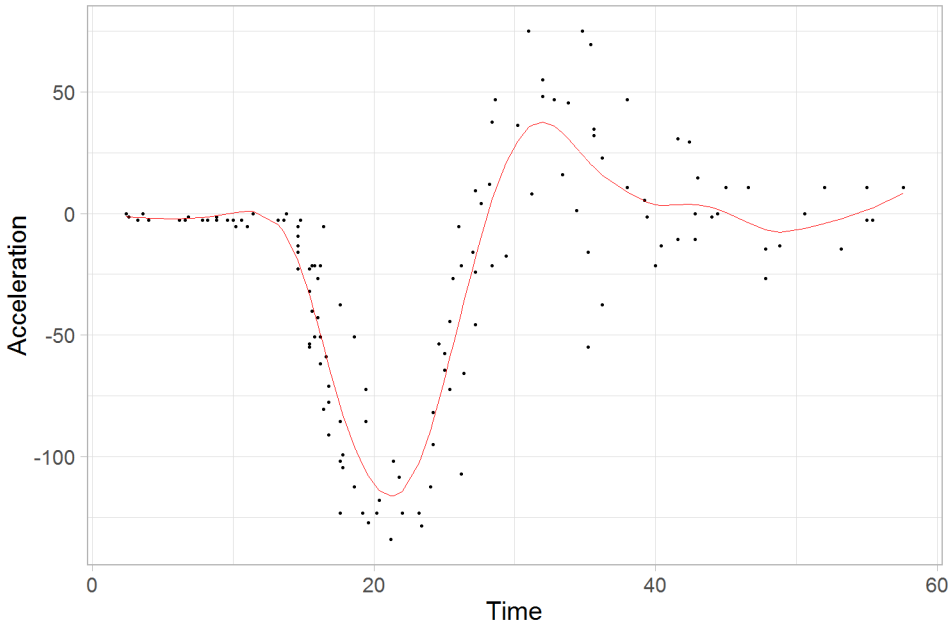
Slika 5.3 prikazuje izglađivač kubičnim splajnom s $\lambda = 9.120377e-05$ i $CV(\lambda) = 543.1041$.

Vidimo da nema neke velike razlike u prediktivnom riziku između kubičnog splajna i lokalnog kubičnog izglađivača. Možda je malo iznenađujuće što smo morali uzeti treći stupanj polinoma da bismo dobili bolji rezultat od splajna.
Na slici 5.4 vidi se primjer ovisnosti CV-score-a o parametru izglađivanja za lokalni kubični izglađivač. Već navedena optimalna vrijednost označena je crvenom bojom.

5.2 Primjer: *airquality*

Podaci *airquality* nalaze se pod istim nazivom u R-ovom paketu "datasets". Podaci su dnevna mjerenja kvalitete zraka u New Yorku u razdoblju od svibnja do rujna 1973. Mi ćemo uzeti samo dvije varijable: redni broj dana u odnosu na 1.5.1973. (dakle, mjerna jedinica je dan) i maksimalna dnevna temperatura na taj dan u stupnjevima Fahrenhajta mjerena u zračnoj luci La Guardia. Podaci su prikazani na Slici 5.5.

Slika 5.6 prikazuje lokalni linearni izglađivač sa širinom prozora $h = 2.125255$

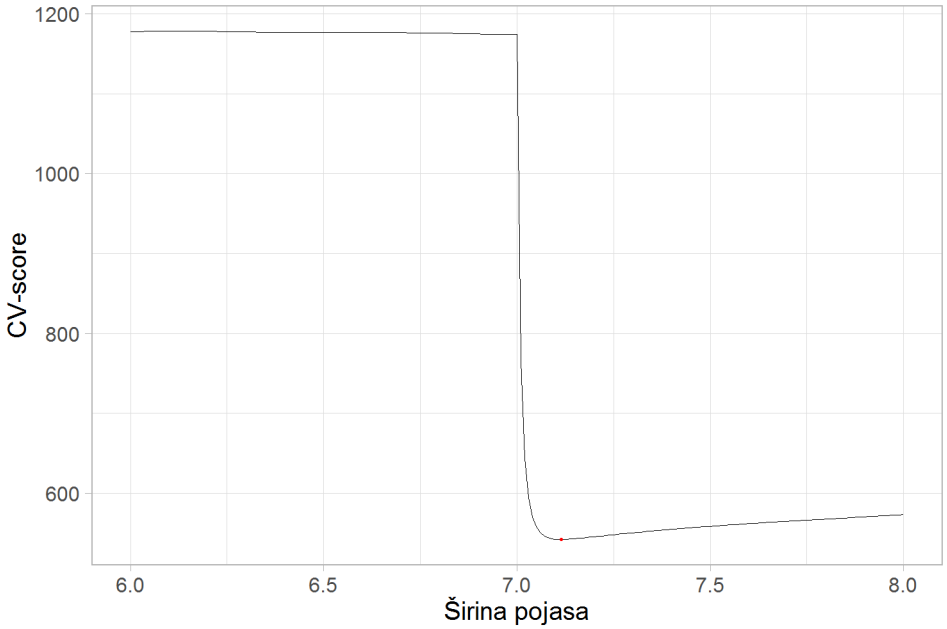


Slika 5.3: Kubični splajn na mcycle podacima

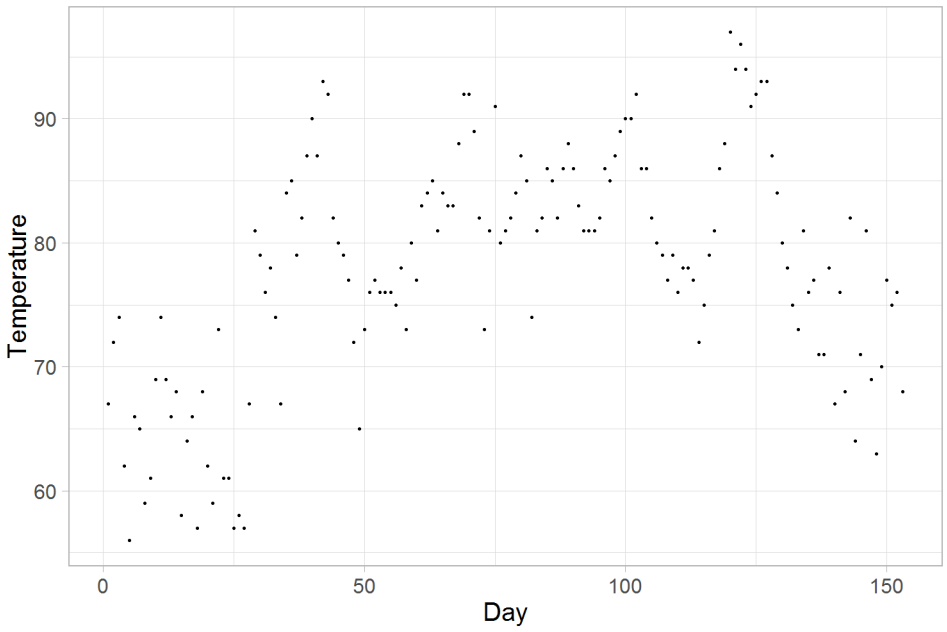
odabranom unakrsnim vrednovanjem i procijenjenim prediktivnom rizikom $CV(h) = 20.67577$.

Slika 5.7 prikazuje izglađivač kubičnim splajnom sa parametrom izglađivanja $\lambda = 6.294312e-07$, također odabranim unakrsnim vrednovanjem, i procijenjenim prediktivnom rizikom $CV(h) = 21.97132$.

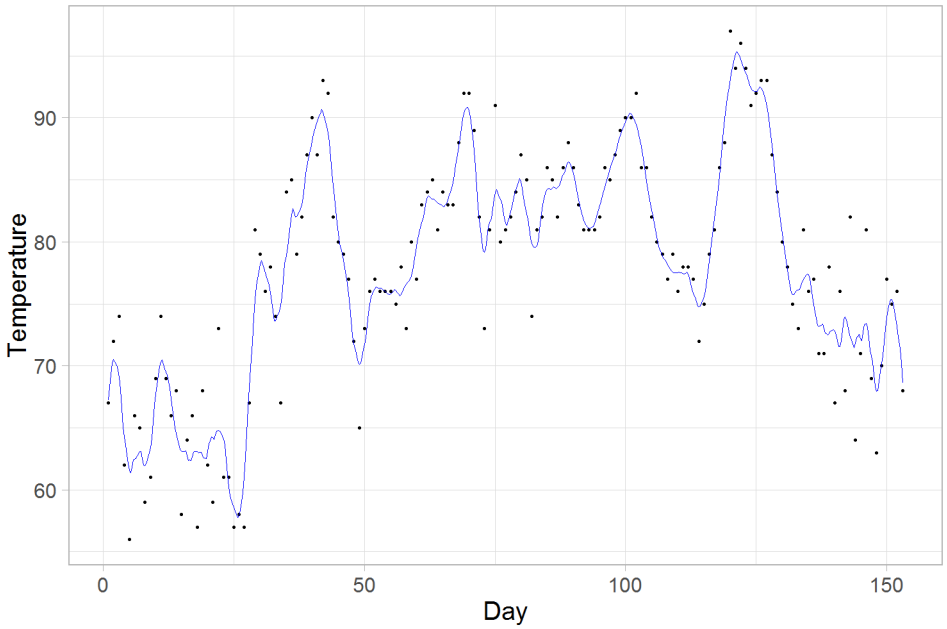
Slika 5.8 prikazuje ta dva izglađivača na jednoj slici radi usporedbe.



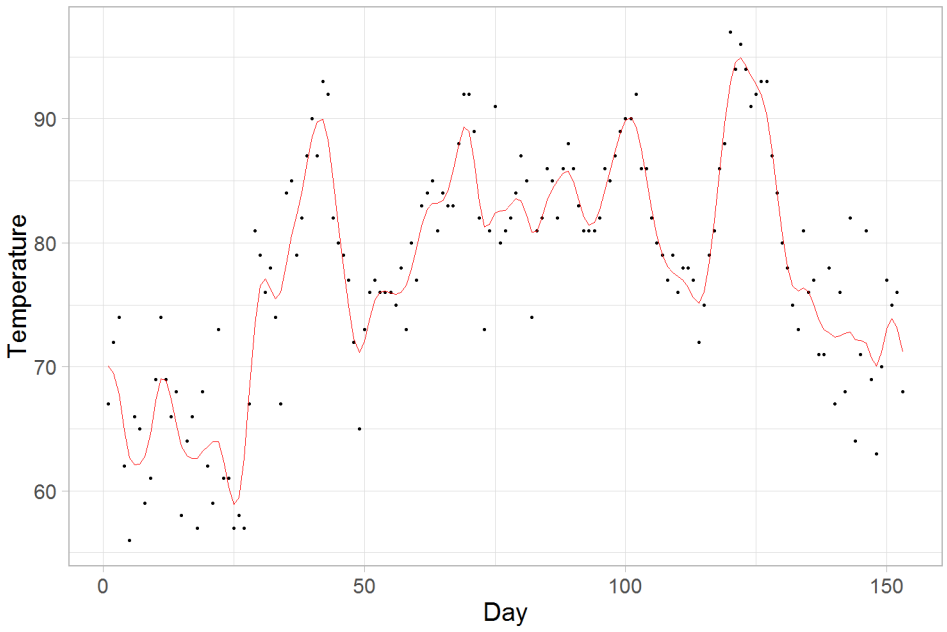
Slika 5.4: Ovisnost CV-score-a o parametru izglađivanja



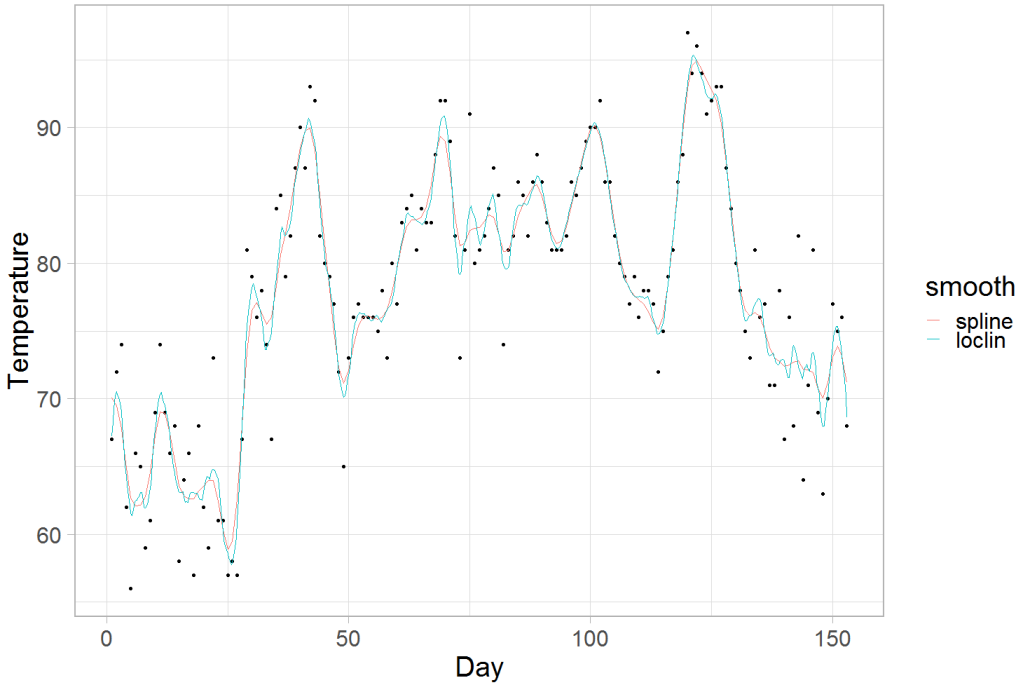
Slika 5.5: airquality podaci



Slika 5.6: Lokalni linearni izglađivač na airquality podacima



Slika 5.7: Kubični splajn na airquality podacima



Slika 5.8: Usporedba izglađivača na airquality podacima

Bibliografija

[1] L. Wasserman, *All of Nonparametric Statistics*, Springer Science+Business Media, New York, 2006.

[2] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York, 1999.

[3] M. P. Wand, M. C. Jones, *Kernel Smoothing*, Chapman & Hall, New York, 1995

[4] Y. Wang, *Smoothing Splines*, Taylor & Francis Group, Boca Raton, 2011.

[5] J. S. Simonoff, *Smoothing Methods in Statistics*, Springer-Verlag, New York, 1996.

[6] J. Fan, I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman & Hall, New York, 1996.

[7] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

[8] A. Buja, T. Hastie, R. Tibshirani, *Linear Smoothers and Additive Models*, The Annals of Statistics, Vol. 17, No. 2. (Jun., 1989), pp. 453-510.

Sažetak

Glavni cilj ovog rada je predstaviti najčešće korištene linearne izglađivače neparametarske regresije.
Započinjemo uvođenjem terminologije teorije aproksimacije i nekih poznatih koncepata primjenjene statistike potrebne za razumijevanje daljnjeg materijala. Nakon toga definiramo linearne izglađivače i dajemo nekoliko njihovih jednostavnih primjera. Nadalje, linearne izglađivače svrstavamo u dvije glavne skupine i dajemo osnovne rezultate vezane uz procjenu pogreške procjene. Na kraju predlažemo metode odabira hiperparametara izglađivača i provodimo linearno izglađivanje na primjeru odabrana dva skupa podataka.

Summary

The main goal of this thesis is to present most commonly used linear smoothers in nonparametric regression.
We begin first by introducing terminology of approximation theory and some well known concepts from applied statistics necessary for understanding the material. We continue by defining linear smoothers and giving several simple examples. Next, two main categories of linear smoothers are discussed and basic results considering approximation of estimation error are given. Furthermore, we suggest methods for smoothing models hyperparameter tuning and show examples of applying linear smoothing on two chosen datasets.

Životopis

Rođena sam 11.12.1994. u Osijeku. Osnovnu školu pohađala sam u Đakovačkim Selcima, a gimnaziju Antuna Gustava Matoša u Đakovu. Preddiplomski studij matematike upisala sam 2013. godine na Prirodoslovno-matematičkom fakultetu u Zagrebu na kojem sam 2016. godine stekla titulu univ. bacc. math. i upisala diplomski studij Matematička statistika.