

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni diplomski studij Financijska matematika i statistika

Karla Višić

Neparametarska regresija

Diplomski rad

Osijek, 2022.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku
Sveučilišni diplomski studij Financijska matematika i statistika

Karla Višić

Neparametarska regresija

Diplomski rad

Mentor: prof. dr. sc. Mirta Benšić

Osijek, 2022.

Sadržaj

Uvod	1
1 Osnovni pojmovi	2
1.1 Regresijska analiza	2
1.2 Jednostavna linearna regresija	4
1.3 Jezgre	5
2 Izgladivači jezgrama	8
2.1 Procjenitelj prosjeka stupaca	8
2.2 Lokalni procjenitelji	8
2.3 Asimptotska svojstva	11
2.4 Referentna širina pojasa	18
2.5 Procjena graničnih vrijednosti	19
2.6 Reziduali i greške predikcije	20
2.7 Unakrsna validacija	21
2.8 Asimptotska distribucija	24
2.9 Uvjetna varijanca i pouzdani intervali	27
3 Multivarijatni slučaj	29
3.1 Prokletstvo dimenzionalnosti	30
3.2 Parcijalna linearna regresija	31
Bibliografija	33
Sažetak	34
Životopis	36

Uvod

U ovom diplomskom radu se bavimo problemom neparametarske regresije u smislu procjene regresijske funkcije nekim lokalnim procjeniteljem. Za početak ćemo dati generalni pregled regresijske analize i pobliže objasniti parametarski i neparametarski pristup. Nakon toga ćemo definirati nekoliko neparametarskih procjenitelja od kojih izdvajamo Nadaraya-Watson i lokalni linearni procjenitelj. Za njih ćemo analizirati asimptotska svojstva, pristranost, varijancu i srednje kvadratnu grešku.

Tehnike lokalnog modeliranja počivaju na ideji lokalnog okruženja pa je od važnosti imati razvijene metode za određivanje optimalne veličine okruženja. Eng. bandwidth, tj. parametar izgladivanja ili širina pojasa govori koliko je veliko okruženje pa ćemo predstaviti dvije tehnike za određivanje istoga: *rule-of-thumb* metodu i metodu unakrsne validacije. Pokazat će se da parametar izgladivanja ima utjecaj na pristranost i varijancu pa je prilikom odabira najboljeg procjenitelja potrebno uzeti u obzir sve parametre. Dotaknut ćemo se graničnog ponašanja procjenitelja, odnosno proučavat ćemo ponašanje i učinkovitost procjenitelja u rubnim točkama s obzirom da se u praksi često pokaže da su granični slučajevi od interesa i važni za razumijevanje. Za kraj ćemo ponuditi preporuku procjenitelja za kojeg smatramo da je prema većini svojstava optimalan te ćemo nizom primjera ilustrirati metode.

Poglavlje 1

Osnovni pojmovi

1.1 Regresijska analiza

Neka su X i Y dvije slučajne varijable te $(X_1, Y_1), \dots, (X_n, Y_n)$ nezavisni, jednako distribuirani vektori kao i (X, Y) gdje je X neovisna varijabla, a Y ovisna varijabla. X možemo zvati još i prediktor ili regresor. Željeli bismo pronaći vezu između varijabli X i Y , kvantificirati ovisnost i efekt koji neovisna varijabla ima na ovisnu te na temelju podataka reći više o distribuciji varijable Y . Drugim riječima, ideja je pronaći funkciju $m(X)$ koja je na neki način dobra aproksimacija varijable Y . Prethodno opisana tehnika se naziva regresijska analiza i jedna je od najčešće korištenih tehnika u statistici. Postoji parametarski i neparametarski pristup procjenjivanja $m(X)$. Ukoliko koristimo parametarski pristup, pretpostavljamo unaprijed zadanu strukturu regresijske funkcije koja ovisi o konačno mnogo parametara. Linearna regresija je jedan primjer parametarske regresije gdje pretpostavljamo da je $m(X)$ linearna funkcija. Zbog svoje jednostavnosti, ujedno je i najviše korištena.

Međutim, u primjenama pretpostavka linearnosti regresijske funkcije često nije zadovoljena i $m(X)$ može imati razne forme. Ako podaci sugeriraju nelinearnu formu, umjesto linearne regresije možemo koristiti polinomijalnu regresiju povećanjem broja parametara. Jedan od problema koji se javlja u polinomijalnoj regresiji je veliki utjecaj rubnih točaka na izgled $m(X)$. Postoji nekoliko metoda kojima se zaobilaze nedostaci polinomijsne regresije od kojih ćemo opisati lokalni (linerani) pristup. Umjesto povećanja broja parametara i korištenja polinomijalne regresije, možemo primijeniti linearnu regresiju lokalno, tj. za $X = x$, lokalno primijenimo linearnu regresiju na skup točaka koje su u blizini od x . Veličina koja definira susjedstvo od x može biti određena objektivno na temelju podataka ili subjektivno od strane analitičara. Na taj način parametarski problem i određivanje unaprijed zadanog broja parametara svodimo na rješavanje mnogo linearnih regresija pri čemu u svakoj određujemo samo dva parametra.

Za dani x , neka je h veličina koja definira susjedstvo oko x . Tada je lokalni linearni model:

$$Y_i = a(x) + b(x)X_i + e \tag{1.1.1}$$

Procjena uređenog para (a, b) se svodi na minimizaciju sume kvadrata odstupanja eksperimentalnih od teorijskih vrijednosti za one realizacije od X_i za koje je $x_i \in [x - h, x + h]$. Kažemo da smo linearnu regresiju primijenili lokano, tj. na dio podataka oko x .

S prethodno opisanom metodom lokalnog modeliranja dolazimo do neparametarske regresije.

Neparametarska regresija ne pretpostavlja unaprijed zadanu distribuciju već za pretpostavku uzima da $m(X)$ može imati bilo koji nelinearni oblik. Na taj način možemo obuhvatiti širu klasu funkcija.

Definirajmo precizno model neparametarske regresije.
Želimo procijeniti funkciju uvjetnog očekivanja:

$$\mathbb{E}[Y|X = x] = m(x) \quad (1.1.2)$$

gdje $m(x)$ zovemo regresijska funkcija.

Za početak promatramo slučaj kada je prediktor X slučajna varijabla. Pretpostavimo da imamo slučajni uzorak $(X_1, Y_1), \dots, (X_n, Y_n)$ gdje su slučajni vektori (X_i, Y_i) nezavisni i jednako distribuirani. U narednim poglavljima se fokusiramo na procjenu regresijske funkcije u $x \in \mathbb{R}$. Neka je x fiksna i ne nalazi se na granicama nosača od X . Neparametarski regresijski model je:

$$\begin{aligned} Y &= m(X) + \epsilon \\ \mathbb{E}[\epsilon|X] &= 0 \\ \mathbb{E}[\epsilon^2|X] &= \sigma^2(X) \end{aligned}$$

Neka je $\hat{m}(x)$ procjenitelj regresijske funkcije (1.1.2) u x . Učinkovitost procjene temeljimo na srednjoj kvadratnoj greški (MSE) ili riziku. MSE definiramo na sljedeći način

$$\text{MSE} = R(m(x), \hat{m}(x)) = \mathbb{E}[(m(x) - \hat{m}(x))^2]$$

Cilj je minimizirati MSE po svim mogućim procjeniteljima \hat{m} . Može se pokazati da MSE ima dekompoziciju na varijancu i pristranost što će u nekim slučajevima dovesti do pojednostavljenja izraza.

Teorem 1. Za procjenitelja $\hat{m}(x)$ od $m(x)$ vrijedi:

$$R(m(x), \hat{m}(x)) = (m(x) - \mathbb{E}[\hat{m}(x)])^2 + \text{Var}(\hat{m}(x))$$

pri čemu prvi izraz zovemo pristranost procjenitelja .

Dokaz:

$$\begin{aligned} \mathbb{E}[(m(x) - \hat{m}(x))^2] &= \mathbb{E}[(m(x) - \mathbb{E}[\hat{m}(x)] + \mathbb{E}[\hat{m}(x)] - \hat{m}(x))^2] \\ &= \mathbb{E}[(m(x) - \mathbb{E}[\hat{m}(x)])^2] + \mathbb{E}[(\mathbb{E}[\hat{m}(x)] - \hat{m}(x))^2] \\ &\quad + 2\mathbb{E}[(m(x) - \mathbb{E}[\hat{m}(x)])(\mathbb{E}[\hat{m}(x)] - \hat{m}(x))] \\ &= (m(x) - \mathbb{E}[\hat{m}(x)])^2 + \text{Var}(\hat{m}(x)) + 2(m(x) - \mathbb{E}[\hat{m}(x)])\mathbb{E}[\mathbb{E}[\hat{m}(x)] - \hat{m}(x)] \\ &= (m(x) - \mathbb{E}[\hat{m}(x)])^2 + \text{Var}(\hat{m}(x)). \end{aligned}$$

gdje zadnja jednakost slijedi iz $\mathbb{E}[\mathbb{E}[\hat{m}(x)] - \hat{m}(x)] = \mathbb{E}[\hat{m}(x)] - \mathbb{E}[\hat{m}(x)] = 0$

□

Definicija 1. Kažemo da je procjenitelj $\hat{m}(x)$ od $m(x)$ nepristran ako vrijedi:
 $\mathbb{E}[\hat{m}(x)] = m(x)$.

Iako se na prvu može činiti da prilikom dekompozicije imamo minimizaciju dva izraza, ukoliko se postigne nepristranost procjenitelja, minimizacija MSE se svodi na minimizaciju varijance.

Teorem 2. *Neka su Y i X dvije slučajne varijable tako da $\mathbb{E}Y^2 < \infty$. Tada za \forall prediktor $g(X)$ od Y vrijedi:*

$$\mathbb{E}[Y - g(X)]^2 \geq \mathbb{E}[Y - \mathbb{E}[Y|X]]^2$$

Dokaz:

$$\begin{aligned} \mathbb{E}[Y - g(X)]^2 &= \mathbb{E}[Y - \mathbb{E}[Y|X] + \mathbb{E}[Y|X] - g(X)]^2 \\ &= \mathbb{E}[Y - \mathbb{E}[Y|X]]^2 + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))] + \mathbb{E}[\mathbb{E}[Y|X] \\ &\quad - g(X)]^2 = (*) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))] &= \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - g(X))|X]] \\ &= \mathbb{E}[(\mathbb{E}[Y|X] - g(X))\mathbb{E}[Y - \mathbb{E}[Y|X]|X]] = 0 \end{aligned}$$

zbog $\mathbb{E}[Y - \mathbb{E}[Y|X]|X] = \mathbb{E}[Y|X] - \mathbb{E}[\mathbb{E}[Y|X]|X] = 0$.

$$\mathbb{E}[Y - g(X)]^2 = (*) = \mathbb{E}[Y - \mathbb{E}[Y|X]]^2 + \mathbb{E}[\mathbb{E}[Y|X] - g(X)]^2 \geq \mathbb{E}[Y - \mathbb{E}[Y|X]]^2$$

jer je $\mathbb{E}[\mathbb{E}[Y|X] - g(X)]^2 \geq 0$. □

Prethodni teorem govori da je najbolji prediktor od Y pomoću X u srednjem kvadratnom smislu uvjetno očekivanje. Posebno je zanimljiv slučaj kada je uvjetno očekivanje linearno:

$$\mathbb{E}[Y|X = x] = m(x) = \alpha + \beta x$$

1.2 Jednostavna linearna regresija

S obzirom da je linearna regresija najčešće korištena regresijska metoda, ima ju smisla pobliže objasniti. Razmatramo slučaj kada je regresor jednodimenzionalna slučajna varijabla pa takav model zovemo jednostavna linearna regresija.

Pretpostavimo da imamo n slučajnih vektora $(X_1, Y_1), \dots, (X_n, Y_n)$ koje želimo opisati modelom

$$Y_i = aX_i + b + \epsilon_i, \quad i = 1, \dots, n \tag{1.2.1}$$

pri čemu je ϵ_i slučajna varijabla koja opisuje grešku modela te $a, b \in \mathbb{R}$.

Ako su $(x_1, y_1), \dots, (x_n, y_n)$ realizacije slučajnih vektora $(X_1, Y_1), \dots, (X_n, Y_n)$, uvodimo sljedeće oznake.

$$\beta = [a, b]^T \quad \mathbf{y} = [y_1, \dots, y_n]^T$$

$$\mathbf{x} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \tag{1.2.2}$$

y_1, \dots, y_n zovemo eksperimentalne ili izmjerene vrijednosti, a $y(x_i) = ax_i + b$ teorijske vrijednosti. Za dani model pretpostavljamo sljedeće:

- $EY_i = ax_i + b$ za svaki $i = 1, \dots, n$, ako je $\epsilon = Y_i - ax_i - b$ onda je $E\epsilon_i = 0$
- Y_1, \dots, Y_n su međusobno nezavisne i imaju istu varijancu σ^2 . Za $\epsilon = [\epsilon_1, \dots, \epsilon_n]^T$, matrica kovarijanci je oblika $\sigma^2 I$
- Y_1, \dots, Y_n su normalno distribuirane slučajne varijable, $Y_i \sim N(ax_i + b, \sigma^2)$

Procjena uređenog para (a, b) se svodi na minimizaciju sume kvadrata odstupanja eksperimentalnih od teorijskih vrijednosti. Primijetimo kako procjenitelja nepoznatih parametara a i b možemo dobiti rješavanjem sljedećeg minimizacijskog problema

$$\hat{\beta} = \operatorname{argmin}_{a,b} \sum_{i=1}^n (y(x_i) - y_i)^2 \quad (1.2.3)$$

gdje je $(a, b) \in \mathbb{R} \times \mathbb{R}$.

Rješavanjem prethodnog minimizacijskog problema dobivamo procjenu za vektorski parametar (a, b) te je

$$\hat{\beta} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}$$

Uvrštavanjem slučajnog vektora $\mathbf{Y} = [Y_1, \dots, Y_n]^T$ u prethodni izraz umjesto \mathbf{y} , dobivamo procjenitelja za nepoznati parametar β .

1.3 Jezgre

Da bismo mogli konstruirati izgladivače jezgrama (procjenitelje) prije svega moramo definirati jezgru koja je važna u brojnim metodama procjene. Najlakše je možemo okarakterizirati kao težinsku funkciju, a formalnu definiciju dajemo u nastavku.

Definicija 2. Jezgra (drugog reda) je funkcija $K : \mathbb{R} \rightarrow \mathbb{R}$ koja zadovoljava sljedeća svojstva:

1. $0 \leq K(u) \leq \overline{K} \leq \infty$
2. $K(u) = K(-u)$
3. $\int_{-\infty}^{\infty} K(u) du = 1$
4. $\int_{-\infty}^{\infty} |u|^r K(u) < \infty$ za sve pozitivne cijele brojeve r

Dakle, iz uvjeta (1) - (3) vidimo da je jezgra omeđena funkcija gustoće koja je simetrična oko 0. Uvjet (4) navodimo zbog pojednostavljenja pri čemu uvjet ne isključuje jezgre od važnosti u praksi.

Veliki broj funkcija zadovoljava svojstva (1) - (4) a neke od najčešće korištenih jezgri su rektangularna, Gaussova, Epanechnikova i triangularna (vidi [4], str. 332).

Jezgre se uobičajno koriste za procjenu funkcije gustoće $f(x)$ u modelu jednostavnog slučajnog uzorka (X_1, \dots, X_n) na sljedeći način:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \quad (1.3.1)$$

Procjenitelj (1.3.1) se naziva i Rosenblatt-Parzen¹ procjenitelj funkcije gustoće po matematičarima koji su ga prvi predložili. Može se pokazati da je procjena koju dobijemo dobro definirana funkcija gustoće. Nenegativnost slijedi iz definicije jezgre, a integriranjem po cijelom \mathbb{R} dobijemo 1:

$$\int_{-\infty}^{\infty} \hat{f}(x) dx = \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{\infty} K\left(\frac{X_i - x}{h}\right) dx = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} K(u) du = 1 \quad (1.3.2)$$

U drugoj jednakosti smo iskoristili zamjenu varijabli $u = (X_i - x)/h$, a treća jednakost slijedi iz definicije jezgre.

Definicija 3. *Realni broj $h > 0$ nazivamo parametar izgladivanja ili širina pojasa.*

Definicija 4. *Normalizirana jezgra je funkcija koja zadovoljava uvjet $\int_{-\infty}^{\infty} u^2 K(u) = 1$.*

Zbog jednostavnosti, u nastavku pretpostavljamo da je jezgra K normalizirana.

Napomenimo da su procjenitelji jezgrama invarijantni na reskaliranje funkcije jezgre i širine pojasa. Drugim riječima, ukoliko koristimo jezgru $K_b(u) = K(u/b)/b$ sa širinom pojasa u/b dobit ćemo istog procjenitelja kao i da smo koristili $K(u)$. Za potrebe neparametarske regresije preporuka je da se koriste Gaussova i Epenchnikova jezgra koje daju slične rezultate. Spomenimo kako izbor jezgre nije od velikog značaja jer procjene koje dobijemo korištenjem različitih jezgri su numerički slične. Međutim, izbor širine pojasa h nije toliko trivijalan i utječe na izgled izgladivanja. Male veličine h će smanjiti grešku aproksimacije i pristranost modeliranja. Za jako male h , posljedično imamo i jako mali broj podataka koji se nalazi u okolini od x što rezultira velikom varijancom. S druge strane, veliki h će rezultirati procjenama koje su zaglađenije, ali pristranost će biti veća.

Neka je lokalni linearni model dan s (1.1.1) te $M_h : h \in (0, \infty)$ familija svih lokalnih linearnih modela indeksiranih parametrom h . Tada za jako male h , procijenjena regresijska funkcija interpolira točke, a za jako velike h se svodi na parametarski linearni model i globalno modeliranje. Stoga za h možemo reći da utječe na kompleksnost modela. Tako izbor h možemo promatrati kao izbor modela M_{h_0} iz familije modela $M_h : h \in (0, \infty)$ koji najbolje opisuje podatke. Postoje razne metode za odabir optimalne širine pojasa, a neke od njih ćemo opisati u kasnijim poglavljima. Svakako, izbor h zbog suprotnog učinka na pristranost i varijancu, u literaturi poznato pod nazivom eng. *bias-variance tradeoff*, nije trivijalan te se odabir na temelju podataka čini boljim u odnosu na subjektivan odabir.

U tablici 1.1 navodimo prethodno spomenute jezgre u normaliziranoj formi.

¹Murray Rosenblatt (1926. - 2019.), američki matematičar
Emanuel Parzen (1929. - 2016.), američki matematičar

Jezgra	Formula	R_k
Rektangularna	$K(u) = \begin{cases} \frac{1}{2\sqrt{3}}, & \text{ako je } u < \sqrt{3} \\ 0, & \text{inače} \end{cases}$	$\frac{1}{2\sqrt{3}}$
Gaussova	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$	$\frac{1}{2\sqrt{\pi}}$
Epanechnikova	$K(u) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{u^2}{5}), & \text{ako je } u < \sqrt{5} \\ 0, & \text{inače} \end{cases}$	$\frac{3\sqrt{5}}{25}$
Triangularna	$K(u) = \begin{cases} \frac{1}{\sqrt{6}}(1 - \frac{ u }{\sqrt{6}}), & \text{ako je } u < \sqrt{6} \\ 0, & \text{inače} \end{cases}$	$\frac{\sqrt{6}}{9}$

Tablica 1.1: Najčešće korištene jezgre drugog reda u normaliziranoj formi

Poglavlje 2

Izgladivači jezgrama

2.1 Procjenitelj prosjeka stupaca

Pretpostavimo da želimo procijeniti uvjetno očekivanje od Y za $X = x$. Razlikujemo dva slučaja ovisno o distribuciji od X . Kada je X diskretna slučajna varijabla, za procjenitelja od $m(x)$ možemo uzeti prosjek Y_i za koje je pripadni $X_i = x$. Ukoliko slučajna varijabla X ima neprekidnu distribuciju ne možemo postupati na analogan način. Naime, vjerojatnost da je $X_i = x$ jednaka je 0, za svaki i . Međutim, ukoliko gledamo okolinu od x i procjenitelja tražimo u odnosu na tu okolinu možemo dobiti dobru aproksimaciju. Drugim riječima, procjenitelj od $m(x)$ je prosjek svih Y_i za koje je realizacija X_i udaljena od x za manje od h , tj. $|X_i - x| \leq h$, za neki $h > 0$. Tako za x definiramo interval $[x - h, x + h]$ kao okolinu od x ili *prozor*. Na taj način dobijemo procjenitelja prosjeka stupaca $\hat{m}(x)$:

$$\hat{m}(x) = \frac{\sum_{i=1}^n 1\{|X_i - x| \leq h\} Y_i}{\sum_{i=1}^n 1\{|X_i - x| \leq h\}} \quad (2.1.1)$$

Procjenitelj kojeg smo dobili je stepenasta funkcija sa skokovima u rubovima prozora. Takva procjena regresijske funkcije je dosta gruba ako $m(x)$ procjenjujemo na ograničenom skupu vrijednosti. Da bismo dobili zaglađeniju procjenu, možemo $m(x)$ procijeniti na bilo kojem skupu vrijednosti, ne nužno ograničenom. Na taj način dobivamo finiju procjenu koja predstavlja generalizaciju procjenitelja prosjeka stupaca kojeg zovemo „Rolling Binned Mean-s“. Pokazat će se da je ovo poseban slučaj Nadaraya - Watson procjenitelja.

Prethodno opisana metoda je lokalno procjenjivala $m(x)$, odnosno ograničila se na prozor unutar kojeg se x nalazi. Takve metode nazivamo metode lokalne regresije.

2.2 Lokalni procjenitelji

U prethodnom dijelu je definiran procjenitelj prosjeka stupaca kojeg možemo generalizirati ako umjesto indikator funkcije koristimo funkciju jezgre. Procjenitelj kojeg dobijemo naziva se Nadaraya - Watson¹ procjenitelj (u nastavku NW procjenitelj) po matematičarima

¹Élizbar Nadaraya (1936. -), gruzijski matematičar
Geoffrey Stuart Watson (1921. - 1998.), australski statističar

koji su ga predložili, neovisno jedno o drugome.

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad (2.2.1)$$

Odabir jezgre nije od značaja i može se koristiti bilo koja spomenuta iz tablice 1.1. Uglavnom se koristi Gaussova jezgra jer takav procjenitelj $\hat{m}(x)$ ima derivacije svakog reda.

Spomenuli smo da izbor širine pojasa utječe na izgled procijenjene regresijske funkcije. Kada $h \rightarrow \infty$, modeliranje postaje globalno i procjena regresijske funkcije se svodi na prosjek po podacima, $\hat{m}(x) \rightarrow \bar{y}$. S druge strane smanjivanjem okoline od x , smanjuje se broj i -ova za koje je $|x_i - x| \leq h$ pa samim time i broj y_i kojima se dodijeli značajna težina. Stoga, $h \rightarrow 0$ povlači $\hat{m}(x_i) \rightarrow y_i$.

Lako se može vidjeti da NW procjenitelj zadovoljava minimizacijski problem

$$\hat{m}_{NW}(x) = \operatorname{argmin}_m \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (Y_i - m)^2 \quad (2.2.2)$$

Naime, vrijedi sljedeće:

$$\begin{aligned} S(m) &= \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) (y_i - m)^2 = \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) (y_i^2 - 2my_i + m^2) \\ &= \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) y_i - 2m \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) + m^2 \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) \end{aligned}$$

$$\hat{m} = \operatorname{argmin}_m S(m)$$

$$\frac{\partial}{\partial m} S(m) = 0$$

$$-2 \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) Y_i + 2m \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) = 0$$

Uvrštavanjem slučajnog vektora X_i u prethodni izraz umjesto x_i , dobivamo NW procjenitelja

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

U tom smislu možemo reći da NW procjenitelj lokalno aproksimira $m(x)$ kao konstantnu funkciju. Prethodna interpretacija nas navodi da je moguće lokalno aproksimirati $m(x)$ nekom drugom funkcijom koja ovisi o više parametara.

$$Y = m(X) + e \simeq m(x) + m'(x)(X - x) + e \quad (2.2.3)$$

Na taj način dobivamo lokalnog linearnog procjenitelja (u nastavku LL procjenitelj). Vidjeli smo da NW procjenitelj zadovoljava minimizacijski problem (2.2.2) te na sličan način LL procjenitelja možemo potražiti kao rješenje minimizacijskog problema:

$$\{\hat{m}_{LL}(x), \hat{m}'_{LL}(x)\} = \underset{\alpha, \beta}{\operatorname{argmin}} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (Y_i - \alpha - \beta(X_i - x))^2 \quad (2.2.4)$$

Sada je aproksimacijski model

$$Y \simeq Z(x)^\tau \beta(x) + e \quad (2.2.5)$$

gdje je

$$\begin{aligned} \beta(x) &= (m(x), m'(x))^\tau \\ Z(x) &= \begin{pmatrix} 1 \\ X - x \end{pmatrix} \end{aligned}$$

(2.2.5) je linearna regresija Y na $Z(x)$ pa primjenom težinske metode najmanjih kvadrata, pri čemu su $K\left(\frac{X_i - x}{h}\right)$ težine (vidi [3], str. 105), eksplicitno dobivamo rješenje za procjenitelja vektora koeficijenata.

$$\begin{aligned} \hat{\beta}_{LL}(x) &= \left(\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Z_i(x) Z_i(x)^\tau \right)^{-1} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Z_i(x) Y_i \\ &= (\mathbf{Z}^\tau \mathbf{K} \mathbf{Z})^{-1} \mathbf{Z}^\tau \mathbf{K} \mathbf{Y} \end{aligned}$$

gdje je

$$Z_i = Z(X_i), \quad \mathbf{Z} = \begin{pmatrix} Z_1^\tau \\ \vdots \\ Z_n^\tau \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

te $\mathbf{K} = \operatorname{diag}K((X_1 - x)/h), \dots, K((X_n - x)/h)$.

Primijetimo da osim $\hat{m}_{LL}(x)$ kao rješenje od (2.2.4) dobivamo i procjenu za derivaciju od $m(x)$. Kada $h \rightarrow \infty$, možemo reći da su svi podaci blizu x što za posljedice ima da svi x_i imaju jednake težine. Tada za dobivenu procjenu vrijedi $\hat{m}_{LL}(x) \rightarrow \hat{\alpha} + \hat{\beta}(x)$ te je LL procjenitelj fleksibilna generalizacija procjenitelja nepoznatih parametara u linearnoj regresiji korištenjem metode najmanjih kvadrata.

Prethodna dva spomenuta procjenitelja možemo generalizirati i promatramo ih kao posebne slučajeve lokalnog polinomijalnog procjenitelja. Naime, ideja je aproksimirati regresijsku funkciju nekim polinomom stupnja p te zatim lokalno procijeniti. Pri tome koristimo Taylorov razvoj u red za $(X, Y) \in \mathbb{R}^2$:

$$\begin{aligned} Y &= m(X) + e \\ Y &\simeq m(x) + m'(x)(X - x) + \dots + m^{(p)}(x) \frac{(X - x)^p}{p!} + e \\ &= Z(x)^\tau \beta(x) + e \end{aligned}$$

pri čemu je

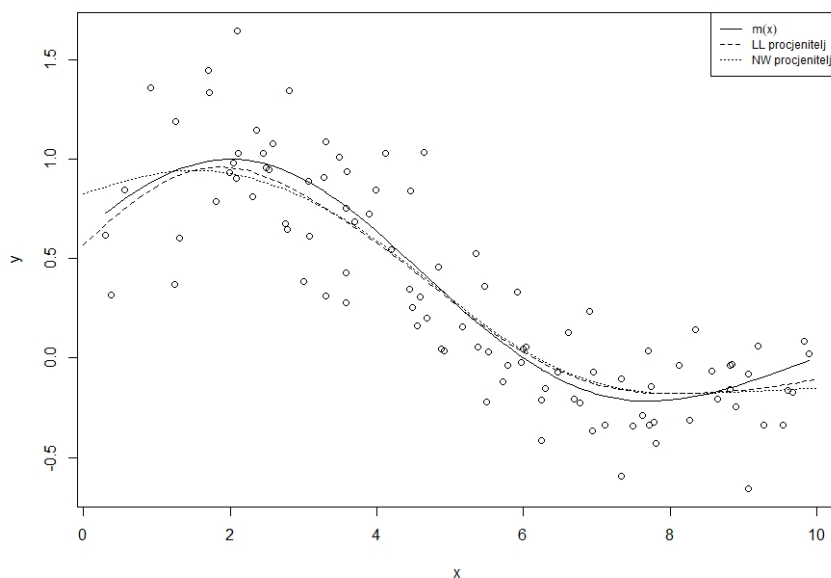
$$Z(x) = \begin{pmatrix} 1 \\ X - x \\ \vdots \\ \frac{(X - x)^p}{p!} \end{pmatrix} \quad \beta(x) = \begin{pmatrix} m(x) \\ m'(x) \\ \vdots \\ m^{(p)}(x) \end{pmatrix}$$

Dobivamo lokalnog polinomijalnog procjenitelja:

$$\begin{aligned} \hat{\beta}_{LP}(x) &= \left(\sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) Z_i(x) Z_i(x)^\tau \right)^{-1} \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) Z_i(x) Y_i \\ &= (\mathbf{Z}^\tau \mathbf{K} \mathbf{Z})^{-1} \mathbf{Z}^\tau \mathbf{K} \mathbf{Y} \end{aligned}$$

Možemo primijetiti da je za NW procjenitelja $p = 0$, a za LL procjenitelja $p = 1$.

Primjer 1. Neka je $m(x) = \frac{\sin((x-2)\pi/4)}{(x-2)\pi/4}$ te generiramo 100 slučajnih parova točaka s greškom $\mathcal{N}(0, 1)$.



Slika 2.1: Usporedba NW i LL procjenitelja

Na slici 2.1 su osim prave funkcije $m(x)$ prikazane procjene dobivene korištenjem NW i LL procjenitelja, kada je $h = 1$. Možemo primijetiti da obje procjene dobro aproksimiraju $m(x)$, međutim u rubnim područjima LL procjenitelj ima veću preciznost.

2.3 Asimptotska svojstva

Sljedeće što ćemo razmotriti su asimptotska svojstva NW i LL procjenitelja. Neka je $f(x)$ funkcija gustoće od X , a $\sigma^2(x) = \mathbb{E}[e^2 | X = x]$ uvjetna varijanca od $e = Y - m(X)$. Asimptotska svojstva razmatramo kada $h \rightarrow 0$, a $nh \rightarrow \infty$. Dodatno pretpostavimo:

1. $m(x)$, $f(x)$ i σ^2 su neprekidne funkcije u okolini N od x .
2. $f(x) > 0$, za $\forall x$ iz nosača od X

Primijetimo da zahtijevamo da širina pojasa postaje sve manja, a broj realizacija sve veći i divergira u beskonačnost. Također, zahtijeva se da je funkcija gustoće nenegativna na nosaču te na taj način osiguravamo dovoljno podataka blizu x .

$m(x)$ označava uvjetno očekivanje od Y uvjetno na $X = x$. Tada je uvjetno očekivanje NW procjenitelja jednako:

$$\mathbb{E}[\hat{m}_{NW}(x)|\mathbf{X}] = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \mathbb{E}[Y_i|X_i]}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) m(X_i)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad (2.3.1)$$

Sljedeće dajemo rezultat o asimptotskoj pristranosti procjenitelja.

Pristranost

Neka vrijede sljedeće oznake:

$$x_n = o(a_n), \quad \text{ako } \frac{x_n}{a_n} \rightarrow 0 \quad \text{za } n \rightarrow \infty$$

$$x_n = O(a_n), \quad \text{ako } \left| \frac{x_n}{a_n} \right| \text{ je ograničen za veliki } n$$

$$X_n = o_P(a_n), \quad \text{ako } \frac{X_n}{a_n} \xrightarrow{p} 0 \quad \text{za } n \rightarrow \infty$$

$$X_n = O_P(a_n), \quad \text{ako } \left| \frac{X_n}{a_n} \right| \text{ u vjerojatnosnom smislu je ograničen za veliki } n$$

Teorem 3. *Pod pretpostavkama ovog poglavlja uz dodatne uvjete o neprekidnosti $m''(x)$ i $f'(x)$ u okolini N od x , za $h \rightarrow 0$ i $nh \rightarrow \infty$ vrijedi:*

$$1. \mathbb{E}[\hat{m}_{NW}(x)|\mathbf{X}] = m(x) + h^2 B_{NW}(x) + o_p(h^2) + O_p\left(\sqrt{\frac{h}{n}}\right)$$

gdje je

$$B_{NW}(x) = \frac{1}{2}m''(x) + f(x)^{-1}f'(x)m'(x)$$

$$2. \mathbb{E}[\hat{m}_{LL}(x)|\mathbf{X}] = m(x) + h^2 B_{LL}(x) + o_p(h^2) + O_p\left(\sqrt{\frac{h}{n}}\right)$$

gdje je

$$B_{LL}(x) = \frac{1}{2}m''(x)$$

Dokaz:

Dokazujemo prvu tvrdnju teorema, za drugu tvrdnju pogledati [1], str.62 .

$$\begin{aligned} \mathbb{E}[\hat{m}_{NW}(x)|\mathbf{X}] &= \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) m(X_i)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} = \frac{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (m(X_i) - m(x) + m(x))}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \\ &= m(x) + \frac{\hat{b}(x)}{\hat{f}(x)} \end{aligned} \quad (2.3.2)$$

pri čemu je $\hat{f}(x)$ procjenitelj funkcije gustoće definiran s (1.3.1), a $\hat{b}(x)$ s

$$\hat{b}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (m(X_i) - m(x)). \quad (2.3.3)$$

Može se pokazati da vrijedi $\hat{f}(x) \xrightarrow{p} f(x)$, (vidi [3], str. 345) pa je dokaz završen ako pokažemo

da vrijedi $\hat{b}(x) = h^2 f(x) B_{NW} + o(h^2) + O_p\left(\sqrt{\frac{h}{n}}\right)$.

Kako je $\hat{b}(x)$ prosjek uzorka nezavisnih komponenti, ima definirano očekivanje.

$$\begin{aligned} \mathbb{E}[\hat{b}(x)] &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{X - x}{h}\right) (m(X) - m(x))\right] \\ &= \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{v - x}{h}\right) (m(v) - m(x)) f(v) dv \\ &= \int_{-\infty}^{\infty} K(u) (m(x + hu) - m(x)) f(x + hu) du \end{aligned} \quad (2.3.4)$$

Treća jednakost slijedi iz transformacije $v = x + hu$.

Koristimo Taylorov razvoj u red:

$$\begin{aligned} m(x + hu) - m(x) &= m'(x)hu + \frac{1}{2}m''(x)h^2u^2 + o(h^2) \\ f(x + hu) &= f(x) + f'(x)hu + o(h) \end{aligned}$$

Uvrštavanjem u (2.3.4) dobivamo:

$$\begin{aligned} &= \int_{-\infty}^{\infty} K(u) \left(m'(x)hu + \frac{1}{2}m''(x)h^2u^2 + o(h^2)\right) (f(x) + f'(x)hu + o(h)) du \\ &= h \left(\int_{-\infty}^{\infty} uK(u)du\right) m'(x) (f(x) + o(h)) \\ &+ h^2 \left(\int_{-\infty}^{\infty} u^2K(u)du\right) \left(\frac{1}{2}m''(x)f(x) + m'(x)f'(x)\right) \\ &+ h^3 \left(\int_{-\infty}^{\infty} u^3K(u)du\right) \left(\frac{1}{2}m''(x)f'(x) + o(h^2)\right) \\ &= h^2 \left(\frac{1}{2}m''(x)f(x) + m'(x)f'(x)\right) + o(h^2) \\ &= h^2 B_{NW} f(x) + o(h^2) \end{aligned}$$

iz čega slijedi da je $\mathbb{E}[\hat{b}(x)] = h^2 f(x) B_{NW} + o(h^2)$.
 Varijanca je manja od drugog momenta pa vrijedi:

$$\begin{aligned} \text{var}[\hat{b}(x)] &= \frac{1}{nh^2} \text{var} \left[K \left(\frac{X-x}{h} \right) (m(X) - m(x)) \right] \\ &\leq \frac{1}{nh^2} \mathbb{E} \left[K \left(\frac{X-x}{h} \right)^2 (m(X) - m(x))^2 \right] \\ &= \frac{1}{nh} \int_{-\infty}^{\infty} K(u)^2 (m(x+hu) - m(x))^2 f(x+hu) du \\ &= \frac{1}{nh} \int_{-\infty}^{\infty} u^2 K(u)^2 du (m'(x))^2 f(x) (h^2 + o(1)) \\ &\leq \frac{h}{n} \overline{K} (m'(x))^2 f(x) + o\left(\frac{h}{n}\right) \end{aligned}$$

gdje zadnja nejednakost slijedi iz definicije jezgre.

Pokazali smo da vrijedi $\text{var}[\hat{b}(x)] \leq O\left(\frac{h}{n}\right)$.

Zaključujemo

$$\hat{b}(x) = h^2 f(x) B_{NW} + o(h^2) + O_p \left(\sqrt{\frac{h}{n}} \right)$$

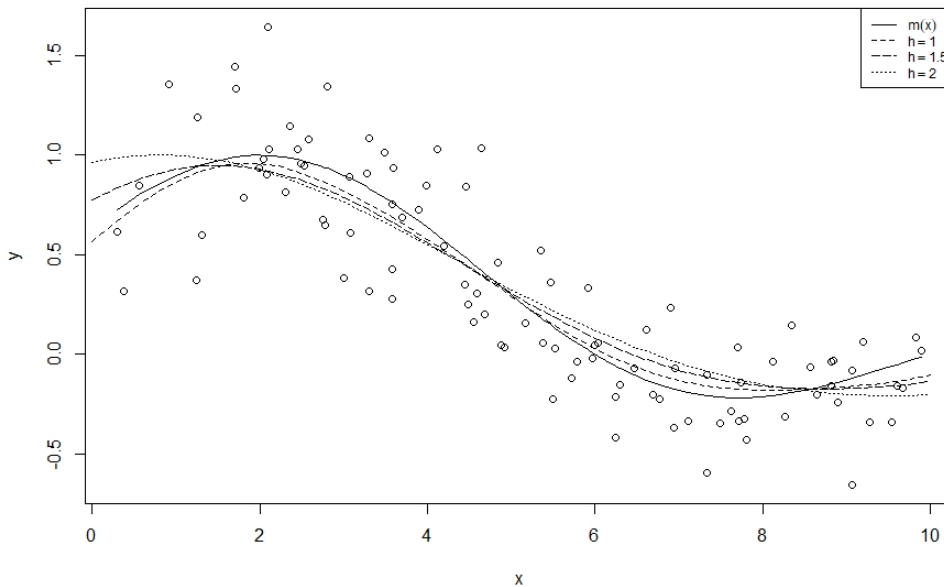
i

$$\frac{\hat{b}(x)}{\hat{f}(x)} = h^2 f(x) B_{NW} + o(h^2) + O_p \left(\sqrt{\frac{h}{n}} \right)$$

te skupa s (2.3.2) dokazuje teorem. □

Izraze $h^2 B_{NW}(x)$ i $h^2 B_{LL}(x)$ zovemo asimptotska pristranost te je proporcionalna kvadratu širine pojasa i funkciji $B_{NW}(x)$, odnosno $B_{LL}(x)$. Što je manja širina pojasa, manja je asimptotska pristranost. Međutim, posljedično će manja širina pojasa dovesti do veće varijance procjenitelja što ćemo vidjeti u sljedećem poglavlju. Funkcija $B_{LL}(x)$ ovisi o drugoj derivaciji $m(x)$ koja nam govori o zakrivljenosti funkcije $m(x)$ što nazivamo pristranost zaglađivanja. U odnosu na $B_{LL}(x)$, $B_{NW}(x)$ ima dodatni izraz što nas upućuje da je LL procjenitelj smanjio asimptotsku pristranost u odnosu na NW procjenitelja.

Slika 2.2 prikazuje ponašanje LL procjenitelja ovisno o izboru h . Prikazane su procjene za $h = 1, 1.5, 2$. Možemo primijetiti kako se porastom širine pojasa pristranost povećava te procjena postaje manje precizna. Također, povećanjem h , procijenjena funkcija postaje sve manje zakrivljena što odgovara prethodnim razmatranjima kada $h \rightarrow \infty$, LL procjenitelj se svodi na procjenitelja u linearnoj regresiji, a dobivena procjena je regresijski pravac.



Slika 2.2: Pristranost LL procjenitelja u ovisnosti o h

Varijanca

Po definiciji, varijanca je srednje kvadratno odstupanje od očekivanja, odnosno $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$. Iz (2.3.1) dobivamo

$$\hat{m}_{NW}(x) - \mathbb{E}[\hat{m}_{NW}(x)|\mathbf{X}] = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) e_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

Slijedi,

$$\text{var}[\hat{m}_{NW}(x)|\mathbf{X}] = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)^2 \sigma^2(X_i)}{\left(\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)\right)^2} \quad (2.3.5)$$

Imamo sve potrebno za dati rezultate o asimptotskoj varijanci procjenitelja.

Teorem 4. *Neka je $R_K = \int_{-\infty}^{\infty} K(u)^2 du$ grubost jezgre $K(u)$ te neka vrijede prethodno definirane pretpostavke. Vrijedi sljedeće:*

1. $\text{var}[\hat{m}_{NW}(x)|\mathbf{X}] = \frac{R_K \sigma^2(x)}{f(x)nh} + o_p\left(\frac{1}{nh}\right)$
2. $\text{var}[\hat{m}_{LL}(x)|\mathbf{X}] = \frac{R_K \sigma^2(x)}{f(x)nh} + o_p\left(\frac{1}{nh}\right)$

Dokaz:

Dokazat ćemo prvu tvrdnju teorema, odnosno varijancu NW procjenitelja (za dokaz druge

vidi [1]).

(2.3.2) možemo zapisati na sljedeći način

$$nh \operatorname{var}[\hat{m}_{NW}(x)|\mathbf{X}] = \frac{\hat{v}(x)}{\hat{f}(x)^2}$$

gdje je $\hat{f}(x)$ procjenitelj funkcije gustoće od $f(x)$, a

$$\hat{v}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)^2 \sigma^2(X_i)$$

Može se pokazati da vrijedi $\hat{f}(x) \xrightarrow{p} f(x)$, (vidi [3], str. 345), pa se dokaz teorema svodi na dokazivanje tvrdnje $\hat{v}(x) \xrightarrow{p} R_K \omega^2(x) f(x)$.

Vrijedi

$$\begin{aligned} \mathbb{E}[\hat{v}(x)] &= \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{v-x}{h}\right)^2 \sigma^2(v) f(v) dv \\ &= \int_{-\infty}^{\infty} K(u)^2 \sigma^2(x+hu) f(x+hu) du \\ &= \int_{-\infty}^{\infty} K(u)^2 du (\sigma^2(x) + o(1)) (f(x) + o(1)) \\ &= \int_{-\infty}^{\infty} K(u)^2 du (\sigma^2(x) f(x) + o(1)) \\ &= R_K \sigma^2(x) f(x) \end{aligned}$$

U drugoj jednakosti smo iskoristili transformaciju $v = x + hu$.

$$\begin{aligned} nh \operatorname{var}[\hat{v}(x)] &= \frac{1}{h} \operatorname{var} \left[K\left(\frac{X-x}{h}\right)^2 \omega^2 \right] \\ &\leq \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{v-x}{h}\right)^4 \sigma^4(v) f(v) dv \\ &= \int_{-\infty}^{\infty} K(u)^4 \sigma^4(x+hu) f(x+hu) du \\ &\leq \bar{K}^2 R_K \sigma^4(x) f(x) + o(1) \end{aligned}$$

Druga nejednakost slijedi iz činjenice da je $\hat{v}(x)$ prosjek nezavisnih slučajnih varijabli i varijanca je manja od drugog momenta. Sada za $n \rightarrow \infty$ slijedi $\operatorname{var}[\hat{v}(x)] \rightarrow 0$. Prema Markovljevoj nejednakosti (vidi [3], str. 151) slijedi $\hat{v}(x) \xrightarrow{p} R_K \omega^2(x) f(x)$ čim je dokaz završen. □

Izraz $\frac{R_K \sigma^2(x)}{f(x)nh}$ zovemo asimptotska varijanca.

Primijetimo da oba procjenitelja imaju jednaku asimptotsku varijancu. Iz teorema 4 i izraza za uvjetnu varijancu procjenitelja možemo zaključiti kako varijanca raste porastom uvjetne varijance greške e , za one x za koje je $f(x)$ mala, te smanjenjem širine pojasa h . Vidjeli smo da su NW i LL procjenitelji zapravo lokalni procjenitelji pa u slučaju da se u okolini od x ne nalazi relativno dovoljno podataka (drugim riječima, $f(x)$ je mala) procjena koju dobijemo nije precizna.

AIMSE

Za početak definiramo asimptotsku srednje kvadratnu grešku $AMSE$ kao

$$AMSE[x] \stackrel{def}{=} h^4 B(x)^2 + \frac{R_K \sigma^2(x)}{nhf(x)} \quad (2.3.6)$$

odnosno

$$AMSE = \text{asimptotska pristranost}^2 + \text{asimptotska varijanca}$$

Integriranjem možemo dobiti globalnu mjeru optimalnosti $AIMSE$ pri čemu je uobičajno integrirati po $f(x)\omega(x)$

$$\begin{aligned} AIMSE &= \int_S \left(h^4 B(x)^2 \frac{R_K \sigma^2(x)}{nhf(x)} \right) f(x)\omega(x) dx \\ &= h^4 \bar{B} + \frac{R_K \bar{\sigma}^2}{nh} \end{aligned}$$

gdje je S nosač od X , $\omega(x)$ neka integrabilna težinska funkcija te

$$\begin{aligned} \bar{B} &= \int_S B(x)^2 f(x)\omega(x) dx \\ \bar{\sigma}^2 &= \int_S \sigma^2(x)\omega(x) dx \end{aligned}$$

Najčešće korištena težinska funkcija je indikator funkcija tako da $\omega(x) = 1\{\xi_1 \leq x \leq \xi_2\}$ kojom osiguravamo da vrijedi $\bar{\sigma}^2 < \infty$ u slučaju da je nosač od X neograničen. Tada za $[\xi_1, \xi_2]$ možemo postaviti neko područje od interesa ili kao rubove intervala uzeti kvantile distribucije od X_i . U slučaju da je nosač S od X ograničen, težinsku funkciju možemo izostaviti. Možemo primijetiti da izbor od h utječe na $AIMSE$ pa je cilj pronaći takav h koji je optimalan.

Teorem 5. Širina pojasa h koji minimizira $AIMSE$ je

$$h_o = \left(\frac{R_K \sigma^2}{4B} \right)^{\frac{1}{5}} n^{-\frac{1}{5}} \quad (2.3.7)$$

Za $h \sim n^{-\frac{1}{5}}$ vrijedi $AIMSE[\hat{m}(x)] = O(n^{-\frac{4}{5}})$.

Prethodni teorem daje izraz za optimalni h u smislu minimizacije $AIMSE$ (vidi [4], Teorem 19.3). Uvrštavanjem u izraz za $AIMSE$, može se izračunati da korištenjem optimalnog h_o vrijedi:

$$AIMSE_0 \simeq 1.65 (R_K^4 \bar{B} \bar{\sigma}^8)^{1/5} n^{-4/5}$$

Postavlja se pitanje može li se izborom funkcije jezgre utjecati na minimizaciju $AIMSE_0$. Lako se može vidjeti da $AIMSE_0$ ovisi o funkciji jezgre samo kroz grubost jezgre. Ukoliko pogledamo u tablici 1 uočavamo da Epanechnikova jezgra ima najmanju vrijednost R_K što potvrđuje i sljedeći teorem (vidi [4], Teorem 19.4).

Teorem 6. *Minimizacija AIMSE za NW i LL procjenitelja se postiže korištenjem Epanechnikove jezgre.*

Sljedeće računamo relativni gubitak učinkovitosti mjeren korijenom od $AIMSE$, ukoliko koristimo Gaussovu jezgru u odnosu na Epanechnikovu jezgru.

$$\left(\frac{AIMSE_0(Gaussian)}{AIMSE_0(Epanechnikov)} \right)^{1/2} = \left(\frac{R_K(Gaussian)^{4/5}}{R_K(Epanechnikov)^{4/5}} \right)^{1/2} = \left(\frac{1/2\sqrt{\pi}}{3\sqrt{5}/25} \right)^{2/5} \simeq 1.02$$

Iz prethodnog proizlazi da je relativni gubitak učinkovitosti jednak 2% ako koristimo Gaussovu jezgru, a 1% i 3% za triangularnu i rektangularnu jezgru što ne predstavlja značajno odstupanje. Naša preporuka je ipak u smjeru korištenja Gaussove jezgre iz razloga što daje zaglađenije procjene. Procjene dobivene korištenjem Gaussove jezgre imaju derivacije svih redova za razliku od Epanechnikove jezgre čija prva derivacija nije neprekidna na granicama nosača.

2.4 Referentna širina pojasa

Lokalni procjenitelji definirani su pomoću širine pojasa te je stoga potrebno razviti metodu za odabir iste. U ovom dijelu ćemo izvesti izraz za referentnu širinu pojasa koja oponaša optimalnu širinu pojasa (2.3.7) te predstavlja temelj za daljnje analize. Metodu su razvili Fan i Gijbels (1996.) i to za lokalnog linearnog procjenitelja.

1. Neka je težinska funkcija indikator funkcija tako da $\omega(x) = 1\{\xi_1 \leq x \leq \xi_2\}$. Za granice intervala možemo uzeti nosač od X ako je ograničen. U suprotnom za $[\xi_1, \xi_2]$ možemo postaviti neko područje od interesa.
2. Regresijsku funkciju procijenimo nekim polinomom reda q za neki $q \geq 2$, $m(x) = \beta_0 + \beta_1 x + \dots + \beta_q x^q$, te metodom najmanjih kvadrata dobijemo procjene koeficijenta $\hat{\beta}_0, \dots, \hat{\beta}_q$. Osim toga, dobivamo izraz za drugu derivaciju procijenjene regresijske funkcije $\hat{m}''(x) = 2\hat{\beta}_2 + \dots + q(q-1)\hat{\beta}_q x^{q-2}$
3. Ako \bar{B} zapišemo u obliku očekivanja

$$\bar{B} = \mathbb{E}[B(X)^2 \omega(x)] = \mathbb{E} \left[\left(\frac{1}{2} m''(X_i) \right)^2 1\{\xi_1 \leq x \leq \xi_2\} \right]$$

, jedan procjenitelj za \bar{B} je

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{2} \hat{m}''(X_i) \right)^2 1\{\xi_1 \leq x \leq \xi_2\}$$

Pretpostavimo da varijanca $\sigma^2 = \mathbb{E}[e^2|X]$ ne ovisi o X te $\bar{\sigma}^2 = \sigma^2(\xi_2 - \xi_1)$. Varijancu procijenimo s $\hat{\sigma}^2$ iz drugog koraka. Optimalni h_0 (2.3.7) možemo zapisati kao

$$h_o = \left(\frac{R_K}{4} \right)^{\frac{1}{5}} \left(\frac{\bar{\sigma}^2}{n\bar{B}} \right)^{\frac{1}{5}} \simeq 0.58 \left(\frac{\bar{\sigma}^2}{n\bar{B}} \right)^{\frac{1}{5}}$$

Uvrštavanjem procjenitelja u gornji izraz dobivamo referentnu širinu pojasa koju nazivamo *rule-of-thumb* (ROT) širina pojasa.

$$h_{ROT} = 0.58 \left(\frac{\hat{\sigma}^2(\xi_2 - \xi_1)}{n\hat{B}} \right)^{1/5} \quad (2.4.1)$$

Iako Fan i Gijbels nisu razvili ROT za NW procjenitelja, koristeći uniformnu distribuciju za $f(x)$ može se pokazati da se optimalna širina pojasa podudara za NW i LL procjenitelja.

2.5 Procjena graničnih vrijednosti

Analizirajući asimptotska svojstva procjenitelja, uočavamo da LL procjenitelj ima smanjenu pristranost u odnosu na NW procjenitelja. Sljedeća dva teorema daju rezultate o graničnom ponašanju procjenitelja u smislu pristranosti.

Teorem 7. *Neka je $\mu_K = 2 \int_0^\infty K(u)du$, a nosač od X neka je $S = [\underline{x}, \bar{x}]$. Ako $m''(\underline{x}+)$, $\sigma^2(\underline{x}+)$ i $f'(\underline{x}+)$ postoje i $f(\underline{x}+) > 0$ tada vrijedi :*

$$\mathbb{E}[\hat{m}_{NW}(\underline{x})|\mathbf{X}] = m(\underline{x}) + hm'(\underline{x})\mu_K + o_p(h^2) + O_p\left(\left(\sqrt{\frac{h}{n}}\right)\right)$$

Ako $m''(\bar{x}-)$, $\sigma^2(\bar{x}-)$ i $f'(\bar{x}-)$ postoje i $f(\bar{x}-) > 0$ tada vrijedi :

$$\mathbb{E}[\hat{m}_{NW}(\bar{x})|\mathbf{X}] = m(\bar{x}) + hm'(\bar{x})\mu_K + o_p(h^2) + O_p\left(\left(\sqrt{\frac{h}{n}}\right)\right)$$

Iz prethodnog teorema vidimo kako je asimptotska pristranost NW procjenitelja za granične točke reda $O(h)$ te ovisi o nagibu uvjetnog očekivanja, $m'(x)$. Zaključujemo kako NW procjenitelj ima veliku asimptotsku pristranost za granične točke.

Prije nego što iskažemo analogan rezultat za LL procjenitelja potrebno je definirati projekcijsku jezgru. Označimo momente $v_j = \int_0^\infty u^j K(u)du$ te neka je $\pi_j = \int_0^\infty u^j K(u)^2 du$. Projekcijska jezgra definirana je sa:

$$K^*(u) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} v_0 & v_1 \\ v_1 & v_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ u \end{bmatrix} K(u) = \frac{v_2 - v_1 u}{v_0 v_2 - v_1^2} K(u)$$

Definiramo drugi moment i grubost jezgre:

$$\sigma_{K^*}^2 = \int_0^\infty u^2 K^*(u) du = \frac{v_2^2 - v_1 v_3}{v_0 v_2 - v_1^2}$$

$$R_K^* = \int_0^\infty K^*(u)^2 du = \frac{v_2^2 \pi_0 - 2v_1 v_2 \pi_1 + v_1^2 \pi_2}{(v_0 v_2 - v_1^2)^2}$$

Sada imamo sve potrebno za rezultat o graničnom ponašanju LL procjenitelja.

Teorem 8. *Uz pretpostavke iz prethodnog teorema za graničnu točku \underline{x} vrijedi:*

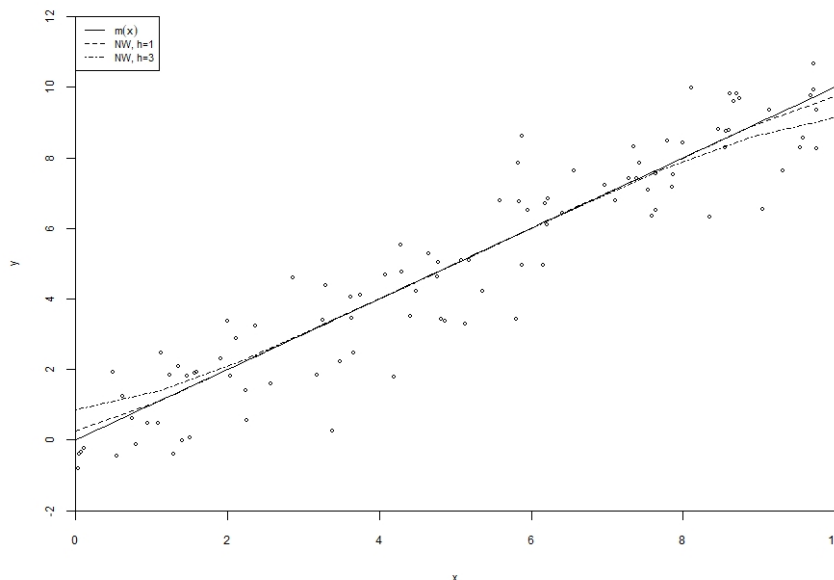
1. $\mathbb{E}[\hat{m}_{LL}(\underline{x})|\mathbf{X}] = m(\underline{x}) + \frac{h^2 m''(\underline{x}) \sigma_{K^*}^2}{2} + o_p(h^2) + O_p\left(\sqrt{\frac{h}{n}}\right)$

$$2. \text{var}[\hat{m}_{LL}(\underline{x})|\mathbf{X}] = \frac{R_K^* \sigma^2(\underline{x})}{f(\underline{x})nh} + o_p\left(\frac{1}{nh}\right)$$

Teoreme 7 i 8 navodimo bez dokaza, a isti se mogu pronaći u Cheng, Fan i Marron (1997) i Imbens i Kalyahnaman (2012).

Možemo primijetiti kako je asimptotska pristranost LL procjenitelja za granične točke reda $O(h^2)$, ista kao za unutarnje točke, invarijantna je na nagib $m(x)$ te je varijanca istog reda kao za unutarnje točke. Usporedbom rezultata o asimptotskoj pristranosti procjenitelja za granične točke vidimo da LL procjenitelj ima asimptotsku pristranost manjeg reda nego NW procjenitelj. Ako analiziramo asimptotsku pristranost procjenitelja za granične i unutarnje točke, LL procjenitelj ima bolja svojstva od NW procjenitelja. U primjenama su granične vrijednosti često od velike važnosti pa je preporuka da se koristi LL procjenitelj.

Primjer 2. Neka je $m(x) = x$ te generiramo 100 parova točaka s greškom $\mathcal{N}(0,1)$.



Slika 2.3: Ponašanje NW procjenitelja u rubnim točkama

Slika 2.3 potvrđuje tvrdnje teorema 7 koji govori o graničnom ponašanju NW procjenitelja. Naime, sa slike uočavamo da je preciznost procjene smanjena u graničnim točkama u odnosu na unutarnje točke. Također, procjena je bolja kada je h manji, tj za $h = 1$.

2.6 Reziduali i greške predikcije

Do sada smo definirali nekoliko različitih procjenitelja i analizirali asimptotska svojstva. Zanima nas koliko je neki procjenitelj prikladan te bismo htjeli definirati mjeru kvalitete procjenitelja. Neka je $x = x_i$ pa rezidual definiramo kao

$$\hat{e}_i = y_i - \hat{m}(x_i)$$

S obzirom da su greške modela nemjerljive, rezidual može poslužiti kao procjenitelj grešaka modela. Može se pokazati da rezidual kao mjera kvalitete nije najbolji odabir. Naime, za jako malu širinu pojasa, za $h \rightarrow 0$, $\hat{m}(x_i) \rightarrow y_i$. Ako pogledamo definiciju reziduala, slijedi da za $h \rightarrow 0$ povlači $\hat{e}_i \rightarrow 0$, što nikako nije točno. Razlog tome leži u činjenici da su NW i LL lokalni izgladivači pa je u procjenu od y_i uključen i y_i . Posljedično će teorijska vrijednost biti uvijek blizu y_i . Ukoliko isključimo y_i iz procjenjivanja, možemo se riješiti prethodno opisanog problema.

Neka je $\tilde{m}_{-i}(x)$ procjena dobivena izbacivanjem i -tog podatka gdje notacija $-i$ označava da je i -ti podatak izostavljen. Tako u slučaju NW procjenitelja imamo sljedeće:

$$\tilde{y}_i = \tilde{m}_{-i}(x) = \frac{\sum_{j \neq i} K\left(\frac{x_j - x}{h}\right) y_j}{\sum_{j \neq i} K\left(\frac{x_j - x}{h}\right)}$$

Sada za $x = x_i$, predikcija od y_i je

$$\tilde{y}_i = \tilde{m}_{-i}(x_i)$$

te greška predikcije iznosi

$$\tilde{e}_i = y_i - \tilde{y}_i$$

koju možemo uzimati kao mjeru kvalitete procjenitelja regresijske funkcije.

2.7 Unakrsna validacija

Metoda unakrsne validacije je popularan odabir za izbor širine pojasa većine neparametarskih procjenitelja. Ideja je procijeniti $m(x)$ na temelju eng. *leave-one-out* validacije koja je ujedno najjednostavniji oblik unakrsne validacije. Kako bismo naglasili ovisnost procjenitelja uvjetnog očekivanja $\hat{m}(x)$ o h , uvodimo oznaku $\hat{m}(x, h)$. Cilj je pronaći h tako da minimizira $IMSE$ od $\hat{m}(x, h)$:

$$IMSE_n(h) = \int_S E[(\hat{m}(x, h) - m(x))^2] f(x) \omega(x) dx$$

Izraz unutar očekivanja, $\hat{m}(x, h) - m(x)$, za $x = x_i$ možemo procijeniti greškom predikcije \tilde{e}_i .

Procjenitelj za $IMSE_n(h)$ je težinski prosjek kvadrata grešaka predikcije \tilde{e}_i :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \tilde{e}_i(h)^2 \omega(X_i) \quad (2.7.1)$$

Funkciju $CV(h)$ zovemo unakrsno validacijski kriterij. Težinsku funkciju $\omega(X_i)$ možemo zanemariti ukoliko X_i ima ograničen nosač. Sljedeći teorem kaže da je $CV(h)$ nepristrani procjenitelj od $IMSE_{n-1}(h)$ do na neku konstante.

Teorem 9. *Vrijedi*

$$\mathbb{E}[CV(h)] = \bar{\sigma}^2 + IMSE_{n-1}(h)$$

pri čemu je $\bar{\sigma}^2 = E[e^2 \omega(X)]$.

Dokaz:

Primijetimo kako je $m(X_i) - \tilde{m}_{i-1}(X_i, h)$ nekorelirano s e_i . Naime, $m(X_i) - \tilde{m}_i(X_i, h)$ možemo zapisati kao funkciju samo od $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ i $(e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n)$.

Vrijedi:

$$\tilde{e}_i(h) = Y_i - \tilde{m}_{-i}(X_i, h) = m(X_i) + e_i - \tilde{m}_{-i}(X_i, h)$$

Imamo sljedeće

$$\begin{aligned} \mathbb{E}[CV(h)] &= \mathbb{E}(\tilde{e}_i(h)^2 \omega(X_i)) \\ &= \mathbb{E}[e_i^2(h)^2 \omega(X_i)] + \mathbb{E}[(\tilde{m}_{-i}(X_i, h) - m(X_i))^2 \omega(X_i)] \\ &\quad + 2\mathbb{E}[(\tilde{m}_{-i}(X_i, h) - m(X_i)) \omega(X_i) e_i] \\ &= \bar{\sigma}^2 + \mathbb{E}[(\tilde{m}_{-i}(X_i, h) - m(X_i))^2 \omega(X_i)] \end{aligned} \quad (2.7.2)$$

Drugi izraz je očekivanje slučajnih varijabli X_i i $\tilde{m}_{-i}(X_i, h)$ koje su nezavisne jer $\tilde{m}_{-i}(X_i, h)$ nije funkcija od X_i . Ako uzmemo uvjetno očekivanje uvjetno na slučajni uzorak bez i -te slučajne varijable, imamo očekivanje jedino od X_i te vrijedi

$$\mathbb{E}_{-i}[(\tilde{m}_{-i}(X_i, h) - m(X_i))^2 \omega(X_i)] = \int (\tilde{m}_{-i}(x, h) - m(x))^2 f(x) \omega(x) dx.$$

Očekivanje gornjeg izraza je jednako

$$\begin{aligned} \mathbb{E}[(\tilde{m}_{-i}(X_i, h) - m(X_i))^2 \omega(X_i)] &= \mathbb{E} \left(\int (\tilde{m}_{-i}(x, h) - m(x))^2 f(x) \omega(x) dx \right) \\ &= IMSE_{n-1} h \end{aligned}$$

Skupa s (2.7.2) dokazuje tvrdnju teorema. □

Primijetimo kako je $\bar{\sigma}^2$ konstantna funkcija nezavisna od h pa $E[CV(h)]$ ovisi o širini pojasa samo kroz drugi član, odnosno $E[CV(h)]$ je jednak $IMSE_{n-1}(h)$ do na konstantu. Stoga je h koji minimizira $IMSE_{n-1}(h)$ jednak h koji minimizira $E[CV(h)]$. Osim toga, za veliki h je $IMSE_{n-1}(h)$ približno jednako $IMSE_n(h)$ što navodi da je $CV(h)$ nepristrani procjenitelj od $IMSE_n(h) + \bar{\sigma}^2$. Posljedično se izbor h koji minimizira $IMSE_n(h)$ svodi na izbor h koji minimizira $CV(h)$.

Vrijedi:

$$h_{CV} = \operatorname{argmin}_{h \geq h_l} CV(h) \quad (2.7.3)$$

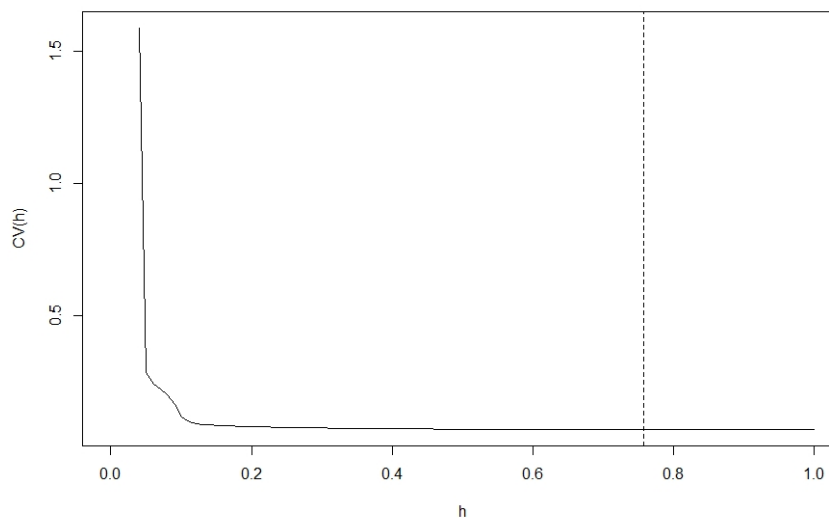
za neki $h_l > 0$, a uvjet $h \geq h_l$ osigurava da ne uzmemo jako mali h . Rješavanje minimizacijskog problema se svodi na numeričko rješavanje s obzirom da ne postoji eksplicitno rješenje. Neke od metoda koje se mogu koristiti su metoda zlatnog reza (vidi [5], str.25) koja je računalno učinkovitija te metoda mrežasto pretraživanje koje je dovoljna za većinu primjena. Odaberemo j vrijednosti za h , $[h_1, \dots, h_j]$ te procijenimo $CV(h_j)$ za svaki h_j . Tada optimalan h zadovoljava:

$$h_{cv} = \operatorname{argmin}_{h \in [h_1, \dots, h_j]} CV(h)$$

Na slici 2.4 je prikazana funkcija

$$CV^*(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{m}_{-i}(x_i))^2 \quad (2.7.4)$$

te identificirana optimalna širina pojasa iscrtanom okomitom linijom. Širina pojasa koju dobijemo je aproksimacija minimuma funkcije (2.7.4) korištenjem mrežastog pretraživanja pri čemu smo unaprijed zadali interval unutar kojeg se traži h (za više detalja vidi [7]) te je za procjenitelja uzet LL procjenitelj. Dobiveni optimalni h iznosi 0.75.



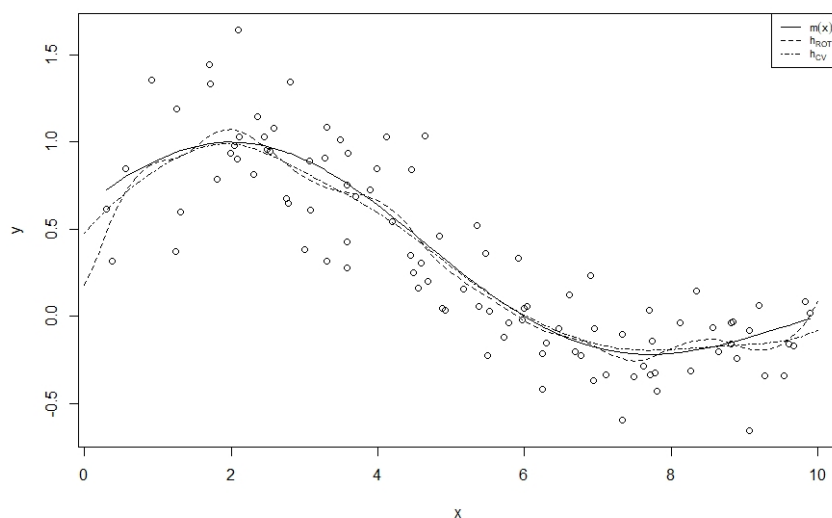
Slika 2.4: Vrijednost CV ovisno o h

Spomenimo i ekstremne slučajeve koji su mogući pa kada $h_{CV} = \infty$, $CV(h)$ se smanjuje, imamo slučaj da je $\hat{m}_{NW}(x) = \bar{Y}$ i $\hat{m}_{LL}(x) = \hat{\alpha} + \hat{\beta}x$. Drugim riječima, NW procjenitelj se svodi na prosjek slučajnog uzorka, a LL procjenitelj na linearnu regresiju.

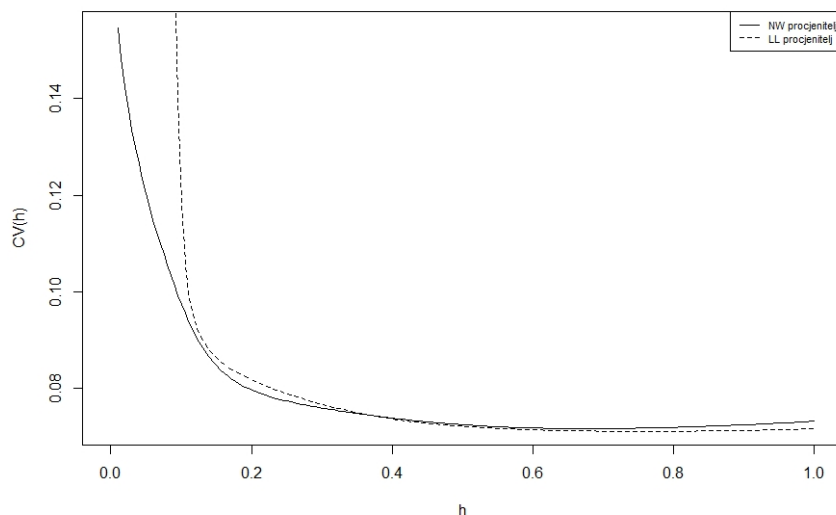
Usporedbu LL procjenitelja korištenjem širine pojasa dobivenih unakrsno validacijskim kriterijem i ROT vidimo na slici 2.5. Pritom je $h_{CV} = 0.75$, a $h_{ROT} = 0.35$. Sa slike uočavamo kako je procjena za h_{CV} zaglađenija te bolje aproksimira $m(x)$.

Osim za određivanje optimalne širine pojasa, unakrsna validacija se koristi za usporedbu različitih neparametarskih procjenitelja. Odabirom procjenitelja s najmanjom vrijednosti unakrsno validacijskog kriterija (2.7.1) dobiva se CV - odabrani procjenitelj.

Na slici 2.6 vidimo usporedbu $CV(h)$ za NW i LL procjenitelja. Možemo uočiti da se minimum unakrsno validacijskog kriterija postiže za LL procjenitelja pa je prema ovom kriteriju, LL procjenitelj CV - odabrani procjenitelj.



Slika 2.5: Procjene $m(x)$ dobivene za različite h -ove



Slika 2.6: Usporedba unakrsno validacijskog kriterija za NW i LL procjenitelja

2.8 Asimptotska distribucija

NW i LL procjenitelji su konzistentni procjenitelji uz vrlo slabe pretpostavke o čemu govori sljedeći teorem.

Teorem 10. *Uz pretpostavke koje su vrijedile do sada , $\hat{m}_{NW}(x) \xrightarrow{p} m(x)$ i $\hat{m}_{LL}(x) \xrightarrow{p} m(x)$*

Dokaz:

Teorem dokazujemo za slučaj NW procjenitelja, dok se dokaz za LL procjenitelja može pronaći u Fan i Gijbels (1996), (vidi [1]) .

NW procjenitelj možemo zapisati na sljedeći način:

$$\hat{m}(x) = m(x) + \frac{\hat{b}(x)}{\hat{f}(x)} + \frac{\hat{g}(x)}{\hat{f}(x)}$$

gdje je $\hat{f}(x)$ procjenitelj funkcije gustoće jezgrom definiran s (1.3.1), $\hat{b}(x)$ je definiran s (2.3.3), a $\hat{g}(x)$ s

$$\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) e_i \quad (2.8.1)$$

Može se pokazati da $\hat{f}(x) \xrightarrow{p} f(x) > 0$, (vidi [4], str. 345), pa se dokaz teorema svodi na dokazivanje dvije tvrdnje, $\hat{b}(x) \xrightarrow{p} 0$ i $\hat{g}(x) \xrightarrow{p} 0$.

Dokažimo prvo $\hat{g}(x) \xrightarrow{p} 0$.

Iz (2.8.1) vidimo da je $\hat{g}(x)$ linearan u e_i , $\mathbb{E}[e|X] = 0$ pa slijedi da je $\mathbb{E}[\hat{g}(x)] = 0$. $\hat{g}(x)$ je prosjek nezavisnih slučajnih varijabli pa je varijanca manja od drugog momenta, $\sigma^2(X) = \mathbb{E}[e^2|X]$ te vrijedi sljedeće:

$$\begin{aligned} nh \text{ var}[\hat{g}(x)] &= \frac{1}{h} \text{ var} \left[K\left(\frac{X-x}{h}\right) e \right] \\ &\leq \frac{1}{h} \mathbb{E} \left[K\left(\frac{X-x}{h}\right)^2 e^2 \right] \\ &= \frac{1}{h} \mathbb{E} \left[K\left(\frac{X-x}{h}\right)^2 \sigma^2(X) \right] \\ &= \int_{-\infty}^{\infty} K(u)^2 \sigma^2(x+hu) f(x+hu) du \\ &= R_K \sigma^2(x) f(x) + o(1) \end{aligned}$$

Zamjenom $u = (X-x)/h$ dobivamo treću jednakost. Za $n \rightarrow \infty$ slijedi $\text{var}[\hat{g}(x)] \rightarrow 0$. Iz Markovljeve nejednakosti (vidi [3], str. 151), slijedi $\hat{g}(x) \xrightarrow{p} 0$.

Dokažimo $\hat{b}(x) \xrightarrow{p} 0$. Prvo računamo očekivanje i varijancu od $\hat{b}(x)$.

Zbog neprekidnosti $m(x)$ i $f(x)$ za $h \rightarrow \infty$ i iz () slijedi:

$$\begin{aligned} \mathbb{E}[\hat{b}(x)] &= o(1) \\ nh \text{ var}[\hat{b}(x)] &\leq o(1) \end{aligned}$$

odnosno,

$$\text{var}[\hat{b}(x)] \rightarrow 0$$

Iz Markovljeve nejednakosti (vidi [3], str. 151), slijedi $\hat{b}(x) \xrightarrow{p} 0$. Skupa s dokazom $\hat{g}(x) \xrightarrow{p} 0$ završavamo dokaz. □

Uz jače pretpostavke o neprekidnosti, sljedeći teorem govori o asimptotskoj normalnosti neparametarskih procjenitelja.

Teorem 11. *Neka su $m''(x)$ i $f'(x)$ neprekidne funkcije u N te neka vrijede pretpostavke kao i sada. Pretpostavimo da za neke $r > 2$ i $x \in N$ vrijedi*

$$E[|e|^r | X = x] \leq \bar{\sigma} < \infty$$

i

$$nh^5 = O(1)$$

Slijedi:

$$\begin{aligned} \sqrt{nh} (\hat{m}_{NW}(x) - m(x) - h^2 B_{NW}(x)) &\xrightarrow{d} N\left(0, \frac{R_K \sigma^2(x)}{f(x)}\right) \\ \sqrt{nh} (\hat{m}_{LL}(x) - m(x) - h^2 B_{LL}(x)) &\xrightarrow{d} N\left(0, \frac{R_K \sigma^2(x)}{f(x)}\right) \end{aligned}$$

Dokaz se može pronaći u Fan i Gijbels (1996) (vidi [[1]]). Dodatni uvjete koje spominjemo su tehničke prirode. Drugi uvjet o ograničenosti širine pojasa osigurava da širina pojasa mora konvergirati u 0 najmanje pri stopi $n^{-1/5}$.

Postoji nekoliko bitnih razlika u distribuciji neparametarskih i parametarskih procjenitelja u regresiji. Stopa konvergencije procjenitelja je \sqrt{nh} a ne \sqrt{n} kao kod parametarskih procjenitelja što za posljedicu ima da neparametarski procjenitelji sporije konvergiraju. Prethodno možemo objasniti činjenicom da je neparametarska procjena zahtjevnija od parametarske. Stoga možemo reći kako je veličina uzorka potrebna za konstrukciju $\hat{m}(x)$ proporcionalna nh a ne n . Drugo, asimptotska distribucija sadrži izraz koji označava pristranost. Proporcionalan je kvadratnoj širini pojasa i funkciji B_{NW} (B_{LL}) za koju smo prethodno pokazali da ovisi o nagibu i zakrivljenosti uvjetnog očekivanja $m(x)$. Zanimljivo je pogledati slučaj kada je $m(x)$ konstanta. Kada je $m(x)$ konstanta povlači da je $B_{NW} = B_{LL} = 0$ što znači da ne postoji asimptotska pristranost. Možemo zaključiti da pristranost raste sa zakrivljenošću regresijske funkcije. Treće, asimptotska varijanca je obrnuto proporcionalana funkciji gustoće $f(x)$ što ide u prilog tezi da procjene graničnih vrijednosti imaju manju preciznost.

Prethodni teorem 11 možemo pojednostavniti ukoliko izostavimo izraz o pristranosti. Naime, u slučajevima kada odaberemo h koji konvergira u 0 brže nego pri optimalnoj stopi $n^{-1/5}$, dolazi do smanjenja pristranosti, ali i rasta varijance. Takav h se naziva eng. *under – smoothing* širina pojasa. Sljedeće navodimo tehnički iskaz (vidi [3], Teorem 19.10).

Teorem 12. *Pod pretpostavkama teorema 11 i $nh^5 = o(1)$ vrijedi:*

$$\begin{aligned} \sqrt{nh} (\hat{m}_{NW}(x) - m(x)) &\xrightarrow{d} N\left(0, \frac{R_K \sigma^2(x)}{f(x)}\right) \\ \sqrt{nh} (\hat{m}_{LL}(x) - m(x)) &\xrightarrow{d} N\left(0, \frac{R_K \sigma^2(x)}{f(x)}\right) \end{aligned}$$

Iako su ovaj pristup koristili mnogi autori, postoji nekoliko negativnih posljedica koje je potrebno napomenuti. Unatoč izostavljanju izraza za pristranost odabirom male širine pojasa, pristranost nije eliminirana. Posljedično dolazi do povećanja varijance što dovodi u pitanje učinkovitost procjenitelja. Osim toga, izbor *under – smoothing* širine pojasa nije najjasniji kako bi se postiglo da konvergira u 0 brže nego pri optimalnoj stopi.

2.9 Uvjetna varijanca i pouzdani intervali

U primjenama je nerijetko od značaja poznavanje uvjetne varijance

$$\sigma^2(x) = \text{var}[Y|X = x] = \mathbb{E}[e^2|X = x]$$

Iz druge jednakosti možemo primijetiti da je $\sigma^2(x)$ uvjetno očekivanje od e^2 uvjetno na $X = x$. Kao u prethodnim razmatranjima za $m(x) = \mathbb{E}[Y|X = x]$, $\sigma^2(x)$ možemo neparametarski procijeniti koristeći jezgre. Imamo:

$$\bar{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) e_i^2}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

Kako greške e_i nisu mjerljive, moramo ih procijeniti. Prirodno se javljaju reziduali $\hat{e} = Y_i - \hat{m}(X_i)$ kao procjenitelji grešaka. Osim reziduala, greške regresije možemo procijeniti grešakama predikcije $\tilde{e}_i = Y_i - \hat{m}_{-i}(X_i)$ pri čemu $\hat{m}_{-i}(X_i)$ predstavlja procjenu za Y_i dobivenu izbacivanjem i -te slučajne varijable. Može se pokazati da su greške predikcije manje sklone *overfitanju* od reziduala te ćemo ih stoga koristiti kao procjenitelje za greške. NW procjenitelj za uvjetnu varijancu je jednak

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \tilde{e}_i^2}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad (2.9.1)$$

Primijetimo da procjenitelj uvjetne varijance ovisi o širini pojasa koju možemo odrediti koristeći ROT ili unakrsnu validaciju neovisno o izboru h korištenog za procjenu uvjetnog očekivanja. S obzirom da je $\sigma^2(x) \geq 0$, procjenitelj kojeg koristimo bi trebao zadovoljavati to svojstvo. Iz (2.9.1) vidimo da je NW procjenitelj težinski prosjek kvadrata grešaka predikcije te je kao takav pozitivan za svaki x , što ne možemo jamčiti za LL procjenitelja. Stoga je preporuka da se kao procjenitelj uvjetne varijance koristi NW procjenitelj.

Sada ćemo izračunati uvjetnu varijancu lokalnih procjenitelja egzaktno. Neka je

$$\begin{aligned} \hat{\beta}(x) &= (\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{K}\mathbf{Z}) \\ &= (\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{K}\mathbf{m}) + (\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{K}\mathbf{e}) \end{aligned}$$

izraz za nekog lokalnog procjenitelja, pri čemu je \mathbf{m} $n + 1$ vektor uvjetnih očekivanja $m(X_i)$. Prvi izraz je funkcija samo od \mathbf{X} , drugi izraz je linearan u \mathbf{e} pa je uvjetna varijanca od $\hat{\beta}$

$$\mathbf{V}_{\hat{\beta}(x)} = \text{var}[\hat{\beta}|\mathbf{X}] = (\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{K}\mathbf{D}\mathbf{K}\mathbf{Z})(\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1}$$

gdje je $\mathbf{D} = \text{diag}(\sigma^2(X_1), \dots, \sigma^2(X_n))$. Zamjenom $\hat{\sigma}^2(X_i)$ sa \hat{e}_i^2 ili \tilde{e}_i^2 dobivamo procjenitelja asimptotske varijance.

$$\hat{\mathbf{V}}_{\hat{\beta}(x)} = (\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1} \left(\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Z_i(x) Z_i(x)' \tilde{e}_i^2 \right) (\mathbf{Z}'\mathbf{K}\mathbf{Z})^{-1} \quad (2.9.2)$$

Osim toga, umjesto $\sigma^2(X_i)$ možemo uvrstiti procjenitelja $\hat{\sigma}^2(X_i)$ te dobivamo

$$\hat{V}_{\hat{m}(x)} = \frac{R_K \hat{\sigma}^2(x)}{nh \hat{f}(x)}$$

pri čemu je $\hat{f}(x)$ procjenitelj funkcije gustoće. Primijetimo da je $\hat{V}_{\hat{m}(x)}$ prvi dijagonalni element matrice $\hat{V}_{\hat{\beta}(x)}$. (2.9.2) je najbliže matrici kovarijanci u slučaju konačnog uzorka pa je preporuka korištenje $\hat{V}_{\hat{\beta}(x)}$ kao procjenitelja uvjetne varijance.

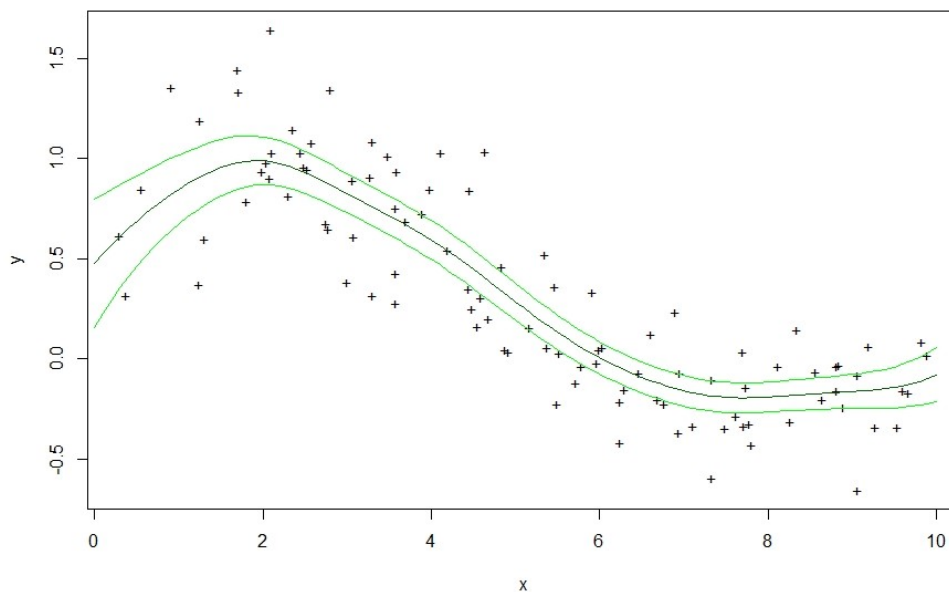
Ostaje konstruirati asimptotske pouzdane intervale za $\hat{m}(x)$. Prethodno smo pokazali da je $\hat{V}_{\hat{m}(x)}$ uvjetna varijanca procjenitelja, a u 2.8 asimptotsku normalnost procjenitelja. 95%-tni pouzdani interval za $m(x)$ možemo konstruirati na sljedeći način:

$$\hat{C} = \left[\hat{m}(x) - 1.96 \sqrt{\hat{V}_{\hat{m}(x)}}, \hat{m}(x) + 1.96 \sqrt{\hat{V}_{\hat{m}(x)}} \right] \quad (2.9.3)$$

gdje je 1.96 dobiven kao rješenje

$$P(|N(0, 1)| \geq c) = 0.95 \quad (2.9.4)$$

Pouzdana intervali se konstruiraju za svaki $m(x)$ posebno što je u literaturi poznato kao „pointwise“ pouzdani intervali. Također, pouzdani intervali ne uzimaju u obzir pristranost procjenitelja pa je ispravnije kazati da je (2.9.3) pouzdani interval za $m(x) + h^2 B(x)$, a ne za $m(x)$.



Slika 2.7: Procijenjena regresijska funkcija i pouzdani intervali

Poglavlje 3

Multivarijatan slučaj

Do sada je pretpostavka bila da imamo n nezavisnih, jednako distribuiranih parova slučajnih vektora (X_i, Y_i) gdje je $X_i \in \mathbb{R}$. U ovom poglavlju razmatramo slučaj kada postoji više neovisnih varijabli, tj. neka je X_i

$$X_i = \begin{pmatrix} X_{1i} \\ \vdots \\ X_{di} \end{pmatrix} \in \mathbb{R}^d.$$

Želimo procijeniti uvjetno očekivanje $\mathbb{E}[Y_i | X_i = x] = m(x)$, za fiksni x , pri čemu je sada $x = (x_1, \dots, x_d) \in \mathbb{R}^d$. Jezgru definiramo kao:

$$K_i(x) = K\left(\frac{X_{1i} - x_1}{h_1}\right) K\left(\frac{X_{2i} - x_2}{h_2}\right) \dots K\left(\frac{X_{di} - x_d}{h_d}\right) \quad (3.0.1)$$

Kao i prije, jezgra na neki način predstavlja težinsku funkciju pri čemu sada udaljenost X_i do x promatramo u Euklidskom prostoru \mathbb{R}^d . Svakom regresoru je pridružena širina pojasa h_i te h predstavlja vektor širina pojasa

$$h = \begin{pmatrix} h_1 \\ \vdots \\ h_d \end{pmatrix}.$$

Za odabir optimalne širine pojasa, koristimo unakrsnu validaciju koja je definirana na analogan način kao u slučaju jednog regresora s tim da je sada $CV(h)$ funkcija d -dimenzionalne širine pojasa.

Nadaraya - Watson procjenitelj ima oblik

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_i(x) Y_i}{\sum_{i=1}^n K_i(x)}$$

gdje je $K_i(x)$ definirana s (3.0.1).

Neka je

$$Z_i(x) = \begin{pmatrix} 1 \\ X_i - x \end{pmatrix},$$

a lokalni linearni procjenitelj, $\hat{\alpha}(x)$, je definiran na sljedeći način, gdje je

$$\begin{pmatrix} \hat{\alpha}(x) \\ \hat{\beta}(x) \end{pmatrix} = \left(\sum_{i=1}^n K_i(x) Z_i(x) Z_i(x)' \right)^{-1} \sum_{i=1}^n K_i(x) Z_i(x) Z_i(x) Y_i$$

Nadalje, neka je $f(x)$ funkcija gustoće od X , $\sigma^2(x) = \mathbb{E}[e^2|X = x]$ uvjetna varijanca od $e = Y - m(X)$ te $|h| = h_1 h_2 \cdots h_d$.

Propozicija 1. *Neka je $\hat{m}(x)$ Nadaraya - Watson ili lokalni linearni procjenitelj regresijske funkcije $m(x)$. Kada $n \rightarrow \infty$ i $h_j \rightarrow 0$ tako da $n|h| \rightarrow \infty$, asimptotska distribucija od $\hat{m}(x)$ je definirana na sljedeći način:*

$$\sqrt{n|h|} \left(\hat{m}(x) - m(x) - \sum_{j=1}^d h_j^2 B_j(x) \right) \xrightarrow{d} N \left(0, \frac{R_K^d \sigma^2(x)}{f(x)} \right)$$

pri čemu je za Nadaraya - Watson procjenitelja

$$B_j(x) = \frac{1}{2} \frac{\partial^2}{\partial x_j^2} m(x) + f(x)^{-1} \frac{\partial}{\partial x_j} f(x) \frac{\partial}{\partial x_j} m(x)$$

a za lokalnog linearnog procjenitelja

$$B_j(x) = \frac{1}{2} \frac{\partial^2}{\partial x_j^2} m(x).$$

Dokaz se može pronaći u Fan i Gijbels (1996) (vidi [1]).

3.1 Prokletstvo dimenzionalnosti

Može se pokazati da konvergencija procjenitelja usporava povećanjem dimenzije od X što je u literaturi poznato pod nazivom „prokletstvo dimenzionalnosti“. Tada procjena regresijske funkcije postaje sve zahtjevnija i nepreciznija. U slučaju jedne neovisne varijable, procjena $m(x)$ se temeljila na lokalnoj aproksimaciji $m(x)$. Slično, i sada se u okolini od x mora nalaziti dovoljno realizacija kako bismo mogli aproksimirati $m(x)$. To u višedimenzionalnom prostoru zbog veće raspršenosti podataka postaje manje vjerojatno, broj realizacija od X_i blizu x se smanjuje što za posljedicu ima smanjenje stope konvergencije procjenitelja.

Optimalnost procjenitelja se diskutirala u kontekstu AIMSE te se u višedimenzionalnom slučaju, AIMSE definira na analogan način s tim da ćemo, zbog jednostavnosti, za širinu pojasa uzeti h koji ima sve jednake koordinate h_j (vidi [3], str. 707):

$$AIMSE = h^4 \int_s \left(\sum_{i=1}^d B_j(x) \right)^2 f(x) \omega(x) dx + \frac{R_K^d}{nh^d} \int_s \sigma^2(x) \omega(x) dx \quad (3.1.1)$$

Primijetimo da je varijanca reda $(nh^d)^{-1}$ dok je u jednodimenzionalnom slučaju varijanca reda $(nh)^{-1}$. Sada je varijanca većeg reda što je u skladu s očekivanjima. Pristranost je istog reda kao i prije, reda h^4 . Kao i u jednodimenzionalnom slučaju, cilj je pronaći h tako da minimizira AIMSE o čemu govori sljedeći teorem (vidi [3], Teorem 19.11).

Teorem 13. *Minimizaciju AIMSE dobivamo za $h \sim n^{-1/(4+d)}$. Tada je*

$$AIMSE = O(n^{-4/(4+d)})$$

Prvo, uočavamo da je optimalna širina pojasa $h = cn^{-1/(4+d)}$ za neku konstantu c . Drugo, stopa optimalnog AIMSE-a ovisi o veličini d pa se povećavanjem dimenzije od d stopa smanjuje. Ovim smo pokazali da se povećanjem dimenzije od X preciznost procjene pogoršava. Napomenimo da ovo nije isključivo za procjenitelje jezgrama već je fenomen prisutan kod većine neparametarskih procjenitelja. Stoga se u praksi neparametarska regresija najčešće koristi kad je samo jedan regresor. Za više od tri regresora nije uobičajno koristiti neparametarsku regresiju.

3.2 Parcijalna linearna regresija

Vidjeli smo da se u s slučajevima velike dimenzionalnosti smanjuje preciznost procjene stoga je potrebno razviti drugačiji pristup rješavanja problema. Jedna od metoda je parcijalna linearna regresija. Regresijsku funkciju možemo podijeliti na dva dijela pri čemu prvom dijelu pristupamo parametarski, a drugom na neparametarski način. Neka je (X, Z) particija regresora gdje je X d - dimenzionalan, a Z k - dimenzionalan.

Tada je model parcijalne linearne regresije oblika:

$$\begin{aligned} Y &= m(X) + Z'\beta + \epsilon \\ \mathbb{E}[e|X, Z] &= 0 \end{aligned} \tag{3.2.1}$$

Dvije su bitne pretpostavke koje treba naglasiti. Prvo, ne postoji neparametarska interakcija između X i Z te je uvjetno očekivanje razdvojivo između X i Z . Drugo, uvjetno očekivanje je linearno u Z . S obzirom da je teško uvijek zadovoljiti prethodne pretpostavke, uzimamo da su one aproksimacija. Particiju (X, Z) ćemo kreirati s obzirom na distribuciju neovisnih varijabli. Regresori koji su diskretne slučajne varijable pripadaju Z , a neprekidne slučajne varijable X . Ukoliko je diskretna slučajna varijabla kategorijalna s više od dvije kategorije, uvjetno očekivanje od Y uvjetno na tu slučajnu varijablu ne možemo zapisati kao linearnu funkciju. Naime, uvjetno očekivanje za svaku kategoriju ima jedinstvenu vrijednost. Linearnost uvjetnog očekivanja se može postići konstruirajući $p - 1$ indikator varijabli, ako je ukupno p kategorija.

Neka je W_1 kategorijalna slučajna varijabla koja može poprimiti 3 vrijednosti.

$$W_1 = \begin{cases} 1, & \text{ako je } W_1 = a \\ 2, & \text{ako je } W_1 = b \\ 3, & \text{ako je } W_1 = c \end{cases}, \quad \mathbb{E}[Y|W_1] = \begin{cases} \mu_1, & \text{ako je } W_1 = 1 \\ \mu_2, & \text{ako je } W_1 = 2 \\ \mu_3, & \text{ako je } W_1 = 3 \end{cases}$$

Indikator varijable su sljedeće:

$$W_2 = \begin{cases} 1, & \text{ako je } W_1 = b \\ 0, & \text{ako je } W_1 \text{ različito od } b \end{cases}, \quad W_3 = \begin{cases} 1, & \text{ako je } W_1 = c \\ 0, & \text{ako je } W_1 \text{ različito od } c \end{cases}$$

EksPLICITNA veza između W_1 i (W_2, W_3) je dana s

$$W_1 = \begin{cases} 1, & \text{ako je } W_2 = 0 \text{ i } W_3 = 0 \\ 2, & \text{ako je } W_2 = 1 \text{ i } W_3 = 0, \\ 3, & \text{ako je } W_2 = 0 \text{ i } W_3 = 1 \end{cases}$$

pa je

$$\mathbb{E}[Y|W_1] = \begin{cases} \mu_{00}, & \text{ako je } W_2 = 0 \text{ i } W_3 = 0 \\ \mu_{10}, & \text{ako je } W_2 = 1 \text{ i } W_3 = 0 \\ \mu_{01}, & \text{ako je } W_2 = 0 \text{ i } W_3 = 1 \end{cases}$$

Sada uvjetno očekivanje od Y uvjetno na W_1 možemo zapisati kao linearnu funkciju od W_2 i W_3 . Imamo:

$$\mathbb{E}[Y|W_1] = \beta_0 + \beta_1 W_2 + \beta_2 W_3$$

gdje koeficijente β_0, β_1 i β_2 možemo dobiti kao rješenje sustava:

$$\begin{aligned} \beta_0 &= \mu_{00} \\ \beta_1 &= \mu_{10} - \mu_{00} \\ \beta_2 &= \mu_{01} - \mu_{00}. \end{aligned}$$

X je najčešće jednodimenzionalan ili dvodimenzionalan. Najveći doprinos parcijalnoj linearnoj regresiji je dao Robins (1988) koji je predložio eliminaciju neparametarske komponente transformacijom. Uzmimo očekivanje od (3.2.1) uvjetno na X . Imamo:

$$\mathbb{E}[Y|X] = m(X) + \mathbb{E}[Z|X]'\beta \quad (3.2.2)$$

Oduzimanjem od (3.2.1) dobijemo

$$Y - \mathbb{E}[Y|X] = (Z - \mathbb{E}[Z|X])'\beta + \epsilon \quad (3.2.3)$$

Prethodno možemo promatrati kao model linearne regresije. Imamo regresiju neparametarske regresijske greške $Y - \mathbb{E}[Y|X]$ na vektor neparametarskih regresijskih grešaka $Z - \mathbb{E}[Z|X]$. Procjenitelj kojeg Robinson predlaže ćemo definirati u 3 koraka.

1. Koristeći neki od spomenutih neparametarskih procjenitelja, Nadaraya - Watson ili lokalni linearni procjenitelj, napraviti k regresija, Y_i na X_i , Z_{1i} na X_i , Z_{2i} na X_i , ..., i Z_{ki} na X_i . Kao rezultat dobivamo teorijske vrijednosti \hat{g}_{0i} , \hat{g}_{1i} , ..., i \hat{g}_{ki} .
2. Napraviti regresiju $Y_i - \hat{g}_{0i}$ na $Z_{1i} - \hat{g}_{1i}, \dots, Z_{ki} - \hat{g}_{ki}$. Kao rezultat dobivamo procjenu od $\hat{\beta}$ i standardne greške.
3. Napraviti neparametarsku regresiju $Y_i - Z_{ki}'\hat{\beta}$ na X_i . Kao rezultat dobivamo neparametarski procjenitelj $\hat{m}(x)$ i pripadajuće intervale pouzdanosti.

Bibliografija

- [1] J. Fan i I. Gijbels *Local Polynomial Modeling and Its Applications*, Chapman Hall, New York, 1996.
- [2] L.Gyorfi, M.Kohler, A.Krzyzak, H.Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York, 2002.
- [3] Bruce E. Hansen, *Econometrics*, Princeton University Press, 2022.
- [4] Bruce E. Hansen, *Introduction to Econometrics*, University of Wisconsin, Department of Economics, 2020.
- [5] Rudolf Scitovski, Ninoslav Truhar, Zoran Tomljanovic, *Metode optimizacije*, Sveučilište Josipa Jurja Strossmayera u Osijeku - Odjel za matematiku, 2014.
- [6] Larry Wasserman, *All of Nonparametric Statistics*, Springer Science+Business Media, 2006.
- [7] timwaite/nprtw: Nonparametric regression package. Dostupno na:
<https://github.com/timwaite/nprtw>)

Sažetak

U ovom diplomskom radu su predstavljene i opisane metode neparametarskog procjenjivanja funkcije uvjetnog očekivanja (regresijska funkcija) pri čemu smo se fokusirali na lokalne izgladivače jezgrama. Za Nadaraya - Watson i lokalni linearni procjenitelj su analizirana asimptotska svojstva, pristranost, varijanca i srednja kvadratna greška. Širina pojasa utječe na izgled regresijske funkcije pa je prilikom procjenjivanja ključno odabrati optimalnu širinu. Stoga su predstavljene dvije metode za odabir, *rule-of-thumb* i unakrsno validacijski kriterij. U zadnjem poglavlju smo se osvrnuli na multivarijantni slučaj i problematiku velike dimenzionalnosti u procjenjivanju. Kao odgovor na smanjenu preciznost opisana je metoda parcijalne regresije.

Ključne riječi: regresijska funkcija, neparametarska regresija, linearna regresija, srednja kvadratna greška, Nadaraya - Watson procjenitelj, lokalni linearni procjenitelj, asimptotska pristranost, asimptotska varijanca, širina pojasa, unakrsno -validacijski kriterij, prokletstvo dimenzionalnosti

Nonparametric regression

Summary

In this thesis we present and describe nonparametric estimation of the conditional expectation function (regression function) with focus on local kernel smoothing estimators. For Nadaraya - Watson estimator and local linear estimator we analyzed their asymptotic properties, bias, variance and mean squared error. Bandwidth affects the level of curvature of the regression function so it is crucial to use optimal bandwidth when estimating. We present two methods for choosing the optimal bandwidth, Rule-of-Thumb and cross validation criterion. In the last chapter we consider the multivariate case and problems of large dimensionality. As a response to reduced precision we describe partial linear regression method.

Keywords: regression function, nonparametric regression, linear regression, mean squared error, Nadaraya - Watson estimator, local linear estimator, asymptotic bias, asymptotic variance, bandwidth, cross validation criterion, curse of dimensionality

Životopis

Rođena sam 27.6.1995. u Splitu. Osnovnu školu završavam u Osnovnoj školi kralja Zvonimira u Segetu Donjem, a srednjoškolsko obrazovanje u III.gimnaziji u Splitu. Nakon završene srednje škole, 2014. godine upisujem prediplomski studij na Prirodoslovno matematičkom fakultet u Zagreb koji završavam 2019. godine i dobivam titulu bacc.mag.edu.math. Iste godine upisujem Diplomski studij Financijska matematika i statistika u Osijeku. Tijekom diplomskog studija sam odradila stručnu praksu u digitalnoj agenciji Escape na poziciji digital marketing asistenta. Osim toga, odradila sam praksu u Hrvatskoj agenciji za poljoprivredu i hranu u Osijeku. Od 10.2021. sam zaposlena u A1 Hrvatska na poziciji Junior Data Analysta gdje sam došla u sklopu A1 Start programa namijenjenog apsolventima ili mladima do 1 godine radnog iskustva.