UDK: 551.49/517.55
Originalni naučni rad

NEPARAMETARSKE FUNKCIJE RASPODELE U HIDROLOGIJI

Mr Đurica MARKOVIĆ
Fakultet tehničkih nauka Univerziteta u Prištini sa sedištem u Kosovskoj Mitrovici
Dr Jasna PLAVŠIĆ, Dr Miloš STANIĆ
Građevinski fakultet Univerziteta u Beogradu
Dr Goran SEKULIĆ
Građevinski fakultet Univerziteta u Podgorici

REZIME

Kao alternativni postupak za parametarske metode u statističkoj analizi se nameću neparametarske metode, pre svega zbog svoje flaksibilnosti i mogućnosti potpune automatizacije određivanja funkcije raspodele. Kao jedna od glavnih osobina koja nameće ovu metodu, u odnosu na parametarske metode, jeste odsustvo potrebe za izbor raspodele na osnovu osmotrenih podataka ili kombinovanje dve teorijske funkcije raspodele za pokrivanje krajeva raspodele.

U radu se prikazuju prednosti i nedostaci primene neparametarskih raspodela baziranih na jezgrima na primeru ocene raspodela nedeljnih proticaja na reci Veliki Rzav u okviru generisanja dugačkih hidroloških nizova za primenu u optimizaciji rada akumulacija. Posebna prednost neparametarskih metoda u takvoj primeni je mogućnost potpune automatizacije postupka generisanja nizova nedeljnih proticaja iz njihovih empirijskih raspodela.

Ključne reči: neparametarske metode raspodele, metoda funkcije jezgra, širina intervala funkcije jezgra

1 UVODNA RAZMATRANJA

Uobičajeni su problemi u hidrologiji da se hidrološke veličine posmatraju kao slučajne promenljive i da se njihove nepoznate raspodele verovatnoće ocenjuju na osnovu osmotrenih nizova podataka određivanjem empirijske raspodele verovatnoće. Ovaj postupak nazivamo statistička analiza. Rezultati statističke analize u hidrologiji se koriste za: projektovanje objekata i sistema za zaštitu od poplava (analiza maksimalnih proticaja, kiša – analiza velikih voda), analizu raspoloživih količina vode za potrebe svih vidova korišćenja vode (vodosnabdevanje, hidroenergetika, poljoprivreda), analizu dugotrajnih sušnih perioda za potrebe vodosnabdevanja i poljoprivrede, analize kvaliteta voda i garantovanih ekoloških proticaja (analiza minimalnih proticaja – analiza malih voda).

Raspodele verovatnoće hidroloških promenljivih su se u doskorašnjoj praksi uobičajeno predstavljale parametarskim metodama. Postupak određivanja raspodele se svodio na izbor najbolje teorijske raspodele u odnosu na empirijsku raspodelu osmotrenog niza. Mada se ova metoda primenjuje veoma uspešno, ona ima i svojih nedostataka. Najčešći problem je da se osmotrenom nizu podataka dovoljno dobro prilagođava više teorijskih raspodela, pri čemu su najveće razlike upravo u delu najznačajnijem za statističku analizu, u krajevima raspodela koji se ekstrapoluju izvan osmotrenih vrednosti. Ne postoji statistički metod obrade podataka koji će prevazići nedostatak nepoznavanja razmatrane teorijske raspodele (Dooge, 1986). Istinska raspodela je uvek nepoznata u praksi, dok proizvoljan izbor ili pretpostavka o obliku stvarne raspodele povećava neizvesnost (Faucher, i drugi, 2001).

Sa praktičnog stanovišta, smatra se da metode za statističku analizu velikih voda trebaju da budu (Lall, i drugi, 1993):

1) jednostavnije u smislu izbora modela i uticaja izuzetaka u osmotrenim podacima,
2) primenljivi na različitim mernim mestima,
3) budu dovoljno fleksibilni za primenu na različite osmotrene podatke,
4) pruže maksimum informacija iz datog uzorka,
5) na odgovarajući način opišu krajeve raspodele, koji su od velikog interesa, i
6) vode doslednim rezultatima i primeni bez obzira na objektivnost istraživača.

Parametarske metode ne zadovoljavaju potpuno navedene kriterijume (Lall, i drugi, 1993).

Svesni ovih nedostataka S. Yakowitz i K. Adamowski su nezavisno počeli da u hidrologiji koriste neparametarske funkcije raspodele. Tako su dalje u svojim radovima (Adamowski, 1985), (Schuster, i drugi, 1985), (Adamowski, 1987), (Adamowski, 1989) primenili ove metode na analizu pojave velikih voda kao i analizu pojave malih voda.

2 NEPARAMETARSKE RASPODELE

Postoji više neparametarskih metoda za određivanje gustine raspodele osmotrenog niza. Najčešće korišćene metode su: metoda ocene gustine raspodele pomoću tzv. funkcije-jezgra (*kernel density estimator*), metoda K-najbližih susednih vrednosti (*K-nearest neighbor*), ocena pomoću splajnova (*spline estimator*) i sl.

Od neparametarskih raspodela se ne može očekivati da zadovolje sve kriterijume koje metode za statističku analizu trebaju da poseduju, a navedeni su u uvodnim razmatranjima. Za neparametarske metode se može reći da (Lall, i drugi, 1993):

1) ne zavise od izbora modela,
2) su primenljive na različitim mernim mestima i pružaju doslednost rezultata bez obzira na objektivnost istrazivača,
3) mogu da jasnije odrede uticaj izuzetaka u podacima na raspodelu,
4) dosta su fleksibilnije, i
5) kako su zasnovane na lokalnim srednjim vrednostima daju bolje težine krajevima raspodele u odnosu na parametarske metode (čija je težina centralnog dela raspodele mnogo veća).

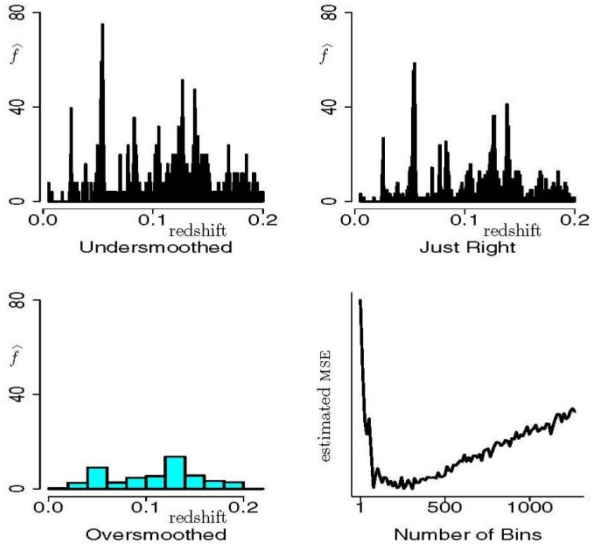
Nedostaci neparametarskih metoda su:

- da dobijena raspodela ne predstavlja teorijsku raspodelu razmatrane veličine,
- s obzirom da se opisuje preko osmotrenih podataka, postoji mala mogućnost ekstrapolacije izvan osmotrenih podataka,
- da postoji subjektivnost istraživača pri optimalnom izboru širine intervala ili faktora izglađenosti funkcije raspodele *h*.

2.1 Metod histograma

Najjednostavniji način prikaza empirijske gustine raspodele odnosno grafičke predstave podataka predstavlja – metod histograma. Na osnovu ukupnog broja podataka iz uzorka se određuju klasni intervali. Zatim se po odgovarajućim klasama raspoređuju vrednosti iz uzorka. Kao srednja vrednost gornjih i donjih granica klasa određuje se reprezentativna vrednost klasnog intervala a spajanjem reprezentativnih vrednosti klasnih intervala mogli bi smo da dobijemo najjednostavniji grafički prikaz emprijske gustine raspodele.

Kod ovog metoda je bitno da se odredi takav broj klasnih intervala koji će omogućiti najbolji prikaz funkcije gustine raspodele. Ukoliko bi se odabrao veliki broj klasnih intervala, dobio bi se dijagram sa velikim brojem štapova što bi liniju funkcije gustine raspodele činilo veoma grubom - neizglađenom. Međutim, ukoliko se izabere manji broj klasnih intervala, linija funkcije gustine raspodele će izgledati izglađeno, ali može davati sasvim druge informacije u odnosu na raspodelu koju imaju ili kojoj teže empirijski podaci. To se može videti sa slike 1 na kojoj su prikazani podaci sa različitim brojem klasnih intervala. Prava mera broja klasnih intervala se može odrediti srednjom kvadratnom greškom (mean square error - MSE) u odnosu na broj klasa. Ovo se može videti na slici 1.



Slika 1. Različiti broj klasnih intervala za iste podatke: gore levo je veći broj klasnih intervala od optimalnog a dole levo manji. Gore desno je optimalni broj klasa i dole desno njegovo određivanje na osnovu srednje kvadratne greške u odnosu na broj klasa

Znači, može se reći da je metoda ocene gustine raspodele uz pomoć histograma najjednostavnija neparametarska metoda za određivanje gustine raspodele na osnovu osmotrenog uzorka, gde se jedino određuje širina klasnog intervala i od njega zavisi kakva će biti ocena gustine.

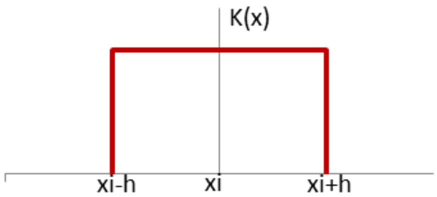
Međutim, kako histogrami nemaju glatku liniju, nastala je ideja da se istim principom, razviju neparametarske metode pomoću određenih funkcija, kao npr. funkcije jezgra, koja izgleda izglađeno i koja brže konvergira ka istinskoj funkciji gustine raspodele (Silverman, i drugi, 1989).

2.2 Metod ocene funkcije gustine raspodele pomoću funkcije jezgra

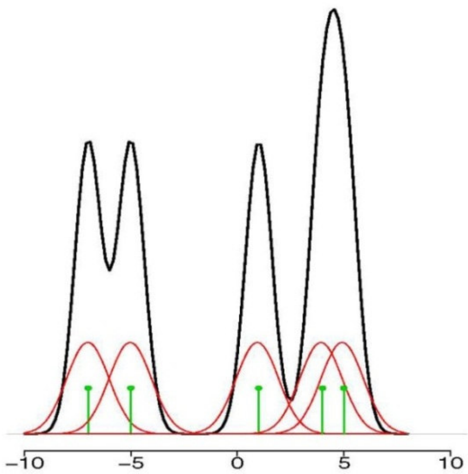
Ocena gustine raspodele pomoću funkcije-jezgra ili neparametarska funkcija gustine raspodele $f_n(x)$ iz uzorka $\{x_1, ..., x_i, ..., x_n\}$ veličine n data je izrazom

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

gde je $K(\ )$ funkcija jezgra centrirana u odnosu na podatak x_i, dok je h širina intervala koji određuje prostiranje individualnih funkcija jezgra oko osmotrene tačke, što je prikazano na slici 2 na primeru pravougaonog jezgra.



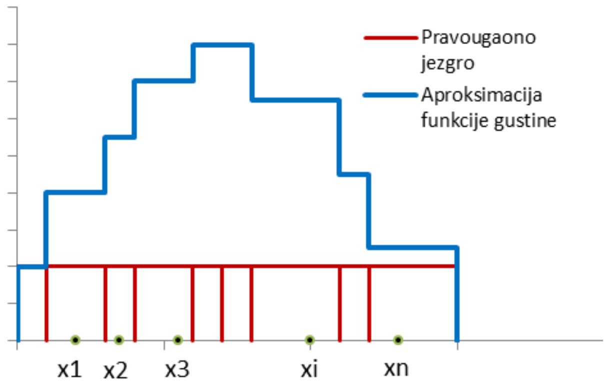
Slika 2. Prikaz pravougaonog jezgra u odnosu na podatak x_i i širina opsega h



Slika 3. Aproksimacija neparametarske funkcije gustine na osnovu jezgara individualnih podataka za Gausovo jezgro

Na slici 3 i slici 4 je prikazana aproksimacija gustine raspodele na osnovu Gausovih i pravougaonih jezgara za individualne podatke. Neparametarska funkcija gustine u stvari predstavlja konvoluciju individualnih funkcija jezgara. Neparametarska funkcija raspodele je onda data izrazom:

$$F_n(x) = \int_{-\infty}^{x} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t - x_i}{h}\right) dt = \frac{1}{n}\sum_{i=1}^{n} K^*\left(\frac{x - x_i}{h}\right)$$



Slika 4. Aproksimacija neparametarske funkcije gustine na osnovu jezgara individualnih podataka za pravougaono jezgro

Neke od osobina koje funkcija jezgra treba da ima su: jezgro je kontinualna funkcija, simetrična oko 0 i sa integralom koji je jednak 1:

$$K(t) = K(-t)$$

$$\int K(t)dt = 1$$

Postoji više različitih tipova funkcija jezgra, među kojima je Gausovo jezgro čest i popularan izbor (Sharma, i drugi, 1997). U tabeli 1 su date funkcije najčešće korišćenih jezgara u metodi ocene funkcije gustine raspodele pomoću funkcije jezgra.

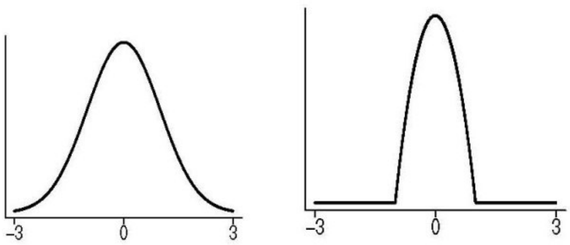
Treba napomenuti da neparametarska funkcija raspodele ne zavisi mnogo od izbora jezgra i da su razlike, u zavisnosti od jezgra, neznatne jer glavni uticaj na oblik funkcije daje vrednost širine intervala h (Adamowski, 1989). Pored toga, treba reći da se Epanečnikovo jezgro smatra optimalnim, a zatim slede dvotežinsko, trougaono, Gausovo i pravougaono. Hardle and Linton (Hardle, i drugi, 1994) su dali koeficijente kojima se optimalna vrednost h za jednu vrstu jezgra može pretvoriti u optimalnu vrednost h za drugu vrstu jezgra. Ovi koeficijenti su dati u tabeli 2.

Tabela 1. Prikaz najčešće korišćenih jezgara u metodi ocene funkcije gustine raspodele, gde je $t = \left(\frac{x - x_i}{h}\right)$

Naziv jezgra	Funkcija
Epanečnikovo	$K(t) = \frac{3}{4}\left(1 - t^2\right)$ za $\lvert t \rvert \le 1$ $K(t) = 0$ u ostalim slučajevima
Gausovo	$K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$
Dvotežinsko (biweight)	$K(t) = \frac{15}{16}\left(1 - t^2\right)^2$ za $\lvert t \rvert \le 1$ $K(t) = 0$ u ostalim slučajevima
Trougaono	$K(t) = 1 - \lvert t \rvert$ za $\lvert t \rvert \le 1$ $K(t) = 0$ u ostalim slučajevima
Pravougaono	$K(t) = \frac{1}{2}$ za $\lvert t \rvert \le 1$ $K(t) = 0$ u ostalim slučajevima

Tabela 2. Koeficijenti različitih jezgara u odnosu na optimalno jezgro

Naziv jezgra	*Kopt*, *K*
Epanečnikovo	1
Dvotežinsko	1.005
Trougaono	1.011
Gausovo	1.041
Pravougaono	1.060



Slika 5. Izgled Gausovog i Epanečnikovog jezgra

Imajući u vidu da gustina raspodele zavisi od izbora širine intervala h a da je srednje kvadratno odstupanje uvek manje kada je h niže može se zaključiti da se najbolja funkcija raspodele dobija za najmanju širinu intervala h. Međutim, ovo nije tačno jer se prekomernim smanjivanjem širine intervala h povećava varijansa. Ovaj postupak rezultira time da je potrebno odrediti neku optimalnu širinu intervala koja će imati određeno srednje kvadratno odstupanje a istovremeno neće uvećati varijansu. Optimalno h se često određuje probanjem za više vrednosti širine intervala dok se ne odredi optimalna vrednost.

Navedeno važi za neparametarske funkcije raspodele sa fiksnim jezgrom. Značajno je napomenuti da se fiksno jezgro najčešće koristi kada se raspolaže sa podacima koji nemaju izraženu asimetriju. Koeficijent asimetrije se na neki način može ublažiti logaritamskom transformacijom podataka. Ukoliko se ni na taj način ne izbegne izrazita asimetričnost podataka primenjuje se funkcija sa promenljivim jezgrom. Promenljivo jezgro ima sledeći smisao: na mestima gde se nalazi veći broj podataka – bliže centralnom delu raspodele, koristi se manja širina intervala, dok se na krajevima raspodele – gde ima mali broj podataka, najčešće po jedan, koristi veća širina intervala i na taj način izbegava osetljivost raspodele na različitu koncentraciju podataka.

3 ODREĐIVANJE ŠIRINE INTERVALA FUNKCIJE – JEZGRA

3.1 Metode za određivanje širine intervala

Kao što se može videti iz prethodno izloženog, pri korišćenju neparametarskih funkcija raspodela nije presudna odluka o korišćenju vrste jezgra, već je najbitnije odrediti optimalnu vrednost širine opsega h. Kako od optimalnog izbora širine opsega zavisi i izgled raspodele na krajevima, što je od najvećeg interesa za statističku analizu, u literaturi se može pronaći veliki broj radova sa različitim metodama za izbor optimalnog h. Postoji više grupa metoda za određivanje optimalne širine intervala. Jedna grupa metoda se bazira na razmatranju grešaka kao što su: ISE – integralna kvadratna greška (*Integrated Square Error*), MISE – integralna srednja kvadratna greška (Mean Integrated Square Error) i AMISE – asimptotska integralna srednja kvadratna greška (*Asymptotic Mean Integrated Square Error*), i ova grupa metoda ima više teoretski značaj. U drugoj grupi su tzv. brzi i približni („*quick and dirty*“) metodi, gde spadaju i razna nepisana pravila, a zasnovani su na AMISE metodi. U treću grupu spadaju metode kros-validacije (*cross-validation*). U četvrtu grupu spadaju metode zamene (*plug-in methods*). Petu grupu čine „optimalni“ metodi. U šestoj grupi su ostale metode i metode koje predstavljaju neke varijacije ostalih metoda. U ovom radu će biti spomenuti samo neki od njih.

3.2 *Integralna srednja kvadratna greška (Integrated Mean Square Error IMSE)*

Na osnovu kriterijuma za računanje optimalne vrednosti na osnovu minimizacije IMSE koje je dato kao:

$$IMSE = \int_{-\infty}^{\infty} E\left[\hat{f}(x) - f(x)\right]^2 dx$$

gde je $\hat{f}(x)$ procena nepoznate funkcije gustine $f(x)$.

Korišćenjem ovog kriterijuma u radu (Adamowski, 1989) je dat obrazac za računanje optimalne vrednosti za *h*:

$$h = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-2}\left(X_i - x_j\right)}{5^{\frac{1}{2}} N\left(N - \frac{10}{3}\right)}$$

Na osnovu numeričkih proučavanja došlo se do zaključka da je ovako sračunato *h* optimalno i za male uzorke (*N=25* je najmanji razmatrani uzorak).

3.3 *Obrazac Adamowskog za trougaono jezgro*

U radu (Adamowski, 1987) je dat obrazac za računanje širine opsega *h*, za trougaono jezgro:

$$h = \left(\frac{5}{192}\pi N\right)^{\frac{1}{5}}$$

3.4 *Nepisano pravilo Silvermana*

Ovaj metod je dat u knjizi (Silverman, 1986) kao tzv. „*rule of thumb*“:

$$h = 1.06\hat{\sigma}N^{-\frac{1}{5}}$$

gde je $\hat{\sigma}$ standardna devijacija uzorka a *N* obim uzorka tj. broj osmotrenih vrednosti. Treba reći da se ovim obrascem računa optimalno *h* samo ako je populacija iz normalne raspodele. Za ostale slučajeve se ne dobija vrednost optimalnog *h*.

4 OCENA FUNKCIJE RASPODELE POMOĆU JEZGARA

Ako je gustina raspodele ocenjena pomoću jezgara, tada je ocena funkcije raspodele:

$$F(x) = \int_{-\infty}^{x} f(u)du = \int_{-\infty}^{x} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{u - x_i}{h}\right)du =$$

$$= \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{u - x_i}{h}\right)du$$

Veličina

$$K^*\left(\frac{u - x_i}{h}\right) = \int_{-\infty}^{x} K\left(\frac{u - x_i}{h}\right)du$$

naziva se integralnim jezgrom, pa se neparametarska funkcija raspodele može prikazati kao:

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} K^*\left(\frac{u - x_i}{h}\right)$$

Drugim rečima, neparametarska funkcija raspodele se može shvatiti kao konvolucija integralnih jezgara nad članovima niza.

Optimalna širina intervala *h* za integralno jezgro funkcije raspodele razlikuje se od optimalne širine za gustinu raspodele. U opštem slučaju se može reći da je širina intervala za funkciju raspodele jednaka

$$h = b\left(\frac{1}{N}\right)^{\frac{2m}{2m+1}}$$

gde je *b>0*, što određuje širinu intervala *h* samo u zavisnosti od broja članova uzorka *N*. Može se pokazati (Lall, i drugi, 1993) da je odnos optimalnih *h* za gustinu raspodele i funkciju raspodele jednak $N^{-1/3}$ / $N^{-1/5}$.

5 NUMERIČKI PRIMER

Za potrebe istraživanja u okviru ovog rada korišćene su neparametarske funkcije raspodele – kao metod ocene funkcije raspodele pomoću jezgra. Kao jezgro funkcije je korišćeno Gausovo jezgro i primenjeno na podacima za vodomerne stanice na reci Veliki Rzav. Osmotreni podaci imaju visok koeficijent asimetrije što je je na neki način uslovilo logaritamsku transformaciju podataka o proticajima. Korišćeni podaci osmotrenih proticaja su prethodno analizirani i eventualno dopunjeni regresijom sa ostalih vodomernih stanica u slivu Velikog Rzava, čime je napravljen niz od 1951-2009. godine. Podaci su svedeni na srednje nedeljne proticaje.

Kao što je već rečeno, izgled funkcije raspodele najviše zavisi od izbora širine intervala *h*, i to je ilustrovano slikama 6 - 8 gde su prikazane neparametarske gustine raspodele sračunate za proticaje 1. nedelje korišćenjem Gausovog jezgra sa različitim vrednostima *h*. Može se videti da na slici 6 imamo malu vrednost širine intervala *h* = 0,038 pa je funkcija gustine malo „nazubljena“. U ovom slučaju širina intervala je određena za Epanečnikovo jezgro a zatim transformisana koeficijentima iz tabele 3 na vrednost širine intervala za Gausovo jezgro. Na slici 7 širina intervala je povećana na *h* = 0,088, što vidimo po glađoj liniji funkcije gustine raspodele i ovo je ocenjeno kao prava mera za širinu intervala za date podatke. Širina intervala je određena korigovanim obrascem datim kao nepisano pravilo, pri čemu je broj podataka N stepenovan na $-\frac{1}{3}$ a ne $-\frac{1}{5}$

Na slici 8 širina intervala je još povećana na *h=0,152*, i dobijeno je obrascem za nepisano pravilo. Na slici se vidi da se dobija veoma glatka funkcija gustine raspodele.

Ovde treba naglasiti da sa previše malom širinom intervala oblik gustine raspodele postaje veoma osetljiv na individualne podatke. Sa druge strane, previše velika vrednost širine intervala daje previše izglađenu gustinu raspodele koja ne može da opiše stvarnu disperziju i asimetričnost podataka (Faucher, i drugi, 2001).

Međutim, pored funkcije gustine raspodele, potrebno je posmatrati i ponašanje funkcije raspodele. Na razmatranom primeru i za iste vrednosti h daju se dijagrami funkcije raspodele na slikama 9-11.

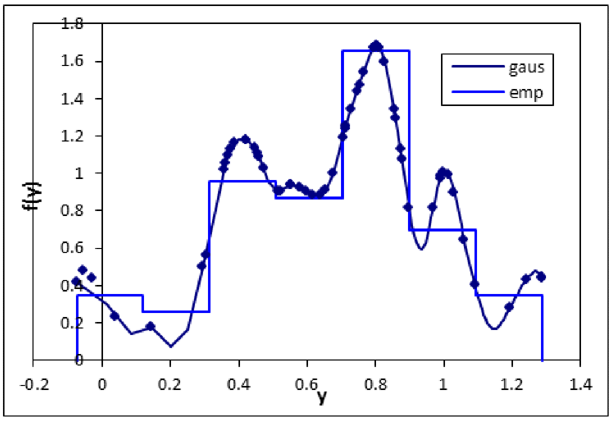
Na spomenutim slikama je lako uočljivo da neparametarska funkcija raspodele najviše „prati“ osmotrene podatke za najmanju vrednost h, dok se sa povećanjem h prvo udaljava od empirijskih tačaka na krajevima raspodele, a zatim, za veliko h, i na ostalim delovima funkcije raspodele. Takođe, na ekstrapolovanim krajevima raspodele za male vrednosti h dolazi do naglog rasta vrednosti izvan opsega opaženih proticaja, dok se taj efekat smanjuje za veće vrednosti h. Za najveće h neparametarska funkcija raspodele je sasvim drugačija od empirijske, tako da uopšte ne opisuje osmotrenu asimetriju i disperziju.

Može se dokazati zavisnost između vrednosti za širinu intervala h jednog jezgra sa drugim jezgrom (Hardle, i drugi, 1994). U tabeli 3 se daju koeficijenti za promenu širine intervala određenih jezgara. Npr. za širinu opsega *h=0.2* korišćenjem dvotežinskog jezgra, prevodimo u širinu opsega *h=0.133* za pravougaono jezgro i *h=0.076* za Gausovo jezgro.

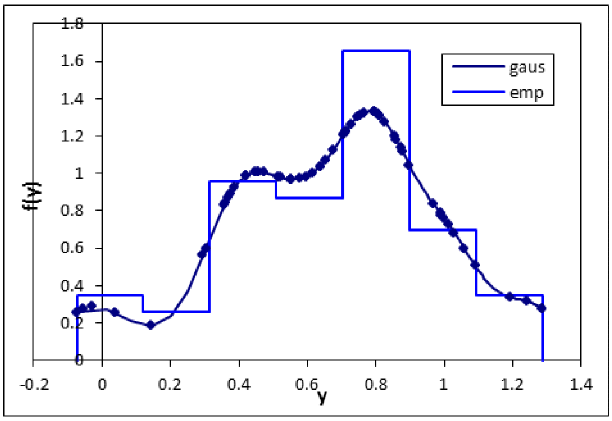
Prikazani postupak primenjen je za određivanje raspodela nedeljnih proticaja na reci Veliki Rzav na stanicama Roge i Radobuđa (za 52 nedelje tokom godine). Tako dobijene raspodele korišćene su kao prvi korak u postupku generisanja dugačkih nizova nedeljnih protoka. Ostali koraci u tom postupku obuhvatili su metode za postizanje korelacije između generisanih podataka koja odgovara osmotrenim podacima, što nije predmet ovog rada.

Tabela 3. Koeficijenti za promenu širine opsega određnih jezgara (Hardle, et al., 1994)

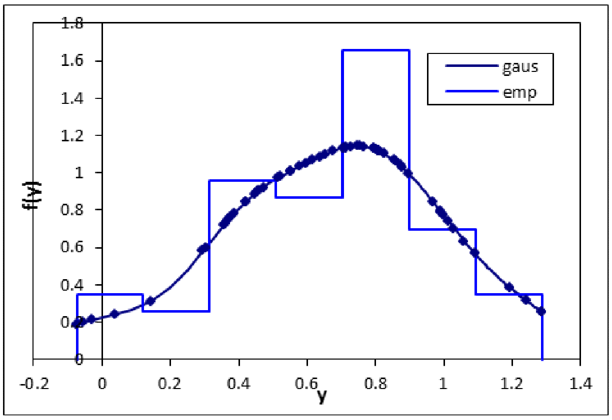
Jezgro (s_j/s_i)	Pravougaono	Trougaono	Epanečnikovo	Dvotežinsko	Gausovo
Pravougaono	1.000	0.715	0.786	0.663	1.740
Trougaono	1.398	1.000	1.099	0.927	2.432
Epanečnikovo	1.272	0.910	1.000	0.844	2.214
Biweight	1.507	1.078	1.185	1.000	2.623
Gausovo	0.575	0.411	0.452	0.381	1.000



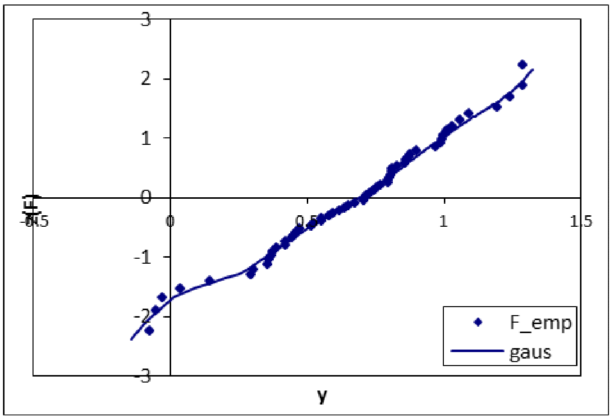
Slika 6. Neparametarska funkcija gustine za osmotrene nedeljne podatke na reci Velikom Rzavu stanica Roge za h=0,038



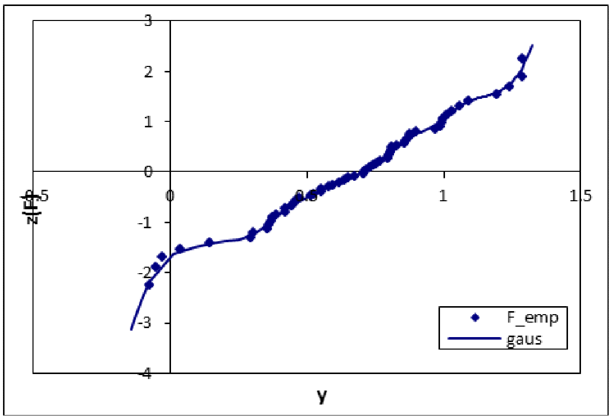
Slika 7. Neparametarska funkcija gustine za osmotrene nedeljne podatke na reci Velikom Rzavu stanica Roge za h=0,088



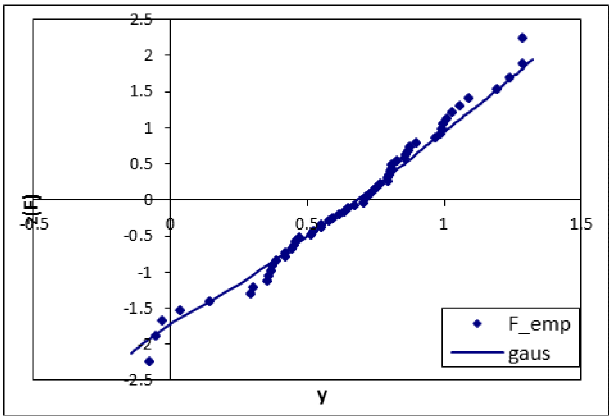
Slika 8. Neparametarska funkcija gustine za osmotrene nedeljne podatke na reci Velikom Rzavu stanica Roge za h=0,152



Slika 10. Neparametarska funkcija raspodele za osmotrene nedeljne podatke na reci Velikom Rzavu stanica Roge za h=0,088



Slika 9. Neparametarska funkcija raspodele za osmotrene nedeljne podatke na reci Velikom Rzavu stanica Roge za h=0,038



Slika 11. Neparametarska funkcija raspodele za osmotrene nedeljne podatke na reci Velikom Rzavu stanica Roge za h=0,152

6 ZAKLJUČAK

Određivanje ocene funkcije raspodele neparametarskim metodama je pre svega omogućeno razvojem računara, s obzirom na veliki broj računskih operacija. U tom smislu postoji potreba da se ceo postupak određivanja ocene obavlja automatski, što je i jedna od preporuka korišćenja neparametarskih metoda gde na osnovu osmotrenog uzorka dobijamo ocenu raspodele.

Radom su predstavljene sve prednosti i nedostatci neparametarskih metoda za ocenu gustine raspodele i funkcije raspodele. Na osnovu toga se može reći da neparametarske metode imaju prednost prilikom izbora metoda za statističku analizu, čak i u slučajevima analize malih i velikih voda gde se zahteva ekstrapolacija funkcije raspodele.

Dalja upotreba neparametarskih funkcija raspodela u hidrologiji se kreće u pravcu korišćenja poboljšanih neparametarskih metoda kao što je npr. metod sa promenljivim jezgrom. Razmatrane metode su korišćene sa fiksnim jezgrom i širina jezgra zavisi od izbora širine intervala *h*. Kako je *h* konstantno za ceo niz osmotrenih podataka, na delovima gde imamo veliki broj podataka (npr. oko medijane), *h* može da bude veliko pa se dobija funkcija koja je previše izglađena (analogno primeru prikazanom na slikama 6-13), dok na krajevima gde imamo mali broj jezgara *h* može biti malo, pa dobijamo nedovoljnu izglađenost. Ovo se može postići metodom sa promenljivim jezgrom gde je i širina intervala *h* promenljiva na osnovu jezgra.

LITERATURA

[1] Adamowski Kaz A Monte Carlo Comparison of Parametric and Nonparametric Estimation of Flood Frequencies [Časopis] // Journal of Hydrology. - 1989. - T. 108. - str. 295-308.

[2] Adamowski Kaz Nonparametric Estimation of Low-Flow Frequencies [Časopis] // Journal of Hydraulic Engineering. - 1996. - 1 : T. 122. - str. 46-49.

[3] Adamowski Kaz Nonparametric Kernel Estimation of Flood Frequencies [Časopis] // Water Resources Research. - 1985. - 11 : T. 21. - str. 1585-1590.

[4] Adamowski Kaz Nonparametric techniques for analysis of hydrological events [Časopis] // Water for the Future: Hydrology in Perspective (Proceedings of the Rome Symposium). - Rome : IAHS Publication No. 164, 1987. - str. 67-76.

[5] Alley William / Burns Alan Mixed-Station Extension of Monthly Streamflow Records [Časopis] // Journal of Hydraulic Engineering. - 1983. - 10 : T. 109. - str. 1272-1284.

[6] Cario Marne / Nelson Barry Autoregressive to anything: Time-series input processes for simulation [Časopis] // Operations Research Letters. - 1996. - T. 19. - str. 51-58.

[7] Dooge James Looking for Hydrologic Laws [Journal] // Water Resources Research. - 1986. - 9 : Vol. 22. - pp. 46S-58S.

[8] Faucher, D., Rasmussen, P.F., Bobee, B. A distribution function based bandwidth selection method for kernel quantile estimation [Časopis] // Journal of Hydrology. - 2001. - T. 250. - str. 1-11.

[9] Hardle Wolfgang / Linton Oliver Applied Nonparametric Methods [Časopis] // Cowles Foundation for Research in Economics at Yale University. - [s.l.] : Cowles Foundation Discussion Paper No. 1069., 1994.

[10] Hirsch Robert A Comparison of Four Record Extension Techniques [Časopis] // Water Resources Research. - 1982. - 4 : T. 18. - str. 1081-1088.

[11] Hui Dafeng [i drugi] Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations [Časopis] // Agricultural and Forest Meteorology. - 2004. - T. 121. - str. 93-111.

[12] Ilich Nesa / Despotovic Jovan A simple method for effective multi-site generation of stochastic hydrologic time series [Časopis] // Stohastic Environmental Research and Risk Assesment. - 2008. - 2 : T. 22. - str. 265-279.

[13] Ilich Nesa A matching algorithm for generation of statistically dependent random variables with arbitrary marginals [Časopis] // Europian Journal of Operational Research. - 2009. - T. 192. - str. 468-478.

[14] Iman Ronald / Conover William A Distribution-free Approach to Inducing Rank Correlation Among Input Variables [Časopis] // Communications in Statistics - Simulation and Computation. - 1982. - 3 : T. 11. - str. 311-334.

[15] Kim Kyung-Duk / Heo Jun-Haeng Comparative study of flood quantiles estimation by nonparametric models [Časopis] // Journal of Hydrology. - 2002. - T. 260. - str. 176-193.

[16] Kim Tae-Woong / Valdes Juan B. Synthetic Generation of Hydrologic Time Series Based on Nonparametric Random Generation [Časopis] // Journal of Hydrologic Engineering. - 2005. - 5 : T. 10. - str. 395-404.

[17] Koutroumanidis Theodoros [i drugi] Genetic modeling for the optimal forecasting of hydrologic time-series: Application in Nestos River [Časopis] // Journal of Hydrology. - 2009. - T. 368. - str. 156-164.

[18] Labatiuk Charles / Adamowski Kaz Application of Nonparametric Density Estimation to Computation of Flood Magnitude/Frequency [Časopis] // Stochastic Hydrology. - 1987. - str. 161-180.

[19] Lall Upmanu, Moon Young-Il / Bosworth Ken Kernel Flood Frequency Estimators: Bandwidth Selection and Kernel Choice [Časopis] // Water Resources Research. - 1993. - 4 : T. 29. - str. 1003-1015.

[20] Moon, Y.-I., Lall, U., Bosworth, K. A comparison of tail probability estimators for flood frequency analysis [Časopis] // Journal of Hydrology. - 1993. - T. 151. - str. 343-363.

[21] Schneider Tapio Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values [Časopis] // American Meteorological Society. - 2001. - str. 853-871.

[22] Schuster Eugene / Yakowitz Sidney Parametric/nonparametric mixture density estimation with application to flood-frequency analysis [Časopis] // Water Resources Bulletin. - 1985. - T. 21. - str. 797-804.

[23] Shao Quanxi [i drugi] Streamflow forecasting using functional-coefficient time series model with periodic variation [Časopis] // Journal of Hydrology. - 2009. - T. 368. - str. 88-95.

[24] Sharif Mohammed / Burn Donald H. Improved K-Nearest Neighbor Weather Generating Model [Časopis] // Journal of Hydrologic Engineering. - 2007. - 1 : T. 12. - str. 42-51.

[25] Sharif Mohammed / Burn Donald H. Simulating climate change scenarios using an improved K-nearest neighbor model [Časopis] // Journal of Hydrology. - 2006. - T. 325. - str. 179-196.

[26] Sharma Ashish / Lall Upmanu A nonparametric approach for daily rainfall simulation [Časopis] // Mathematics and Computers in Simulation. - 1999. - T. 48. - str. 361-371.

[27] Sharma Ashish Seasonal to interannual rainfall probabilistic forecast for improved water supply management: Part 3 – A nonparametric probabilistic forecast model [Časopis] // Journal of Hydrology. - 2000. - T. 239. - str. 249-258.

[28] Sharma Ashish, Lall Upmanu / Tarboton David Kernel bandwidth selection for a first order nonparametric streamflow simulation model [Časopis] // Statistic Hydrology and Hydraulics. - 1998. - T. 12. - str. 33-52.

[29] Sharma Ashish, Tarboton David / Lall Upmanu Streamflow simulation: A nonparametric approach [Časopis] // Water Resources Research. - 1997. - 2 : T. 33. - str. 291-308.

[30] Silverman B W / Jones M C E.Fix and J.L.Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation [Časopis]. - [s.l.] : International Statistical Review, 1989. - T. 3. - str. 233-247.

[31] Silverman Bernard W. Density Estimator for Statistics and Data Analysis [Časopis] // Statistics and Applied Probability. - London : Chapman & Hall, 1986.

[32] Smakhtin Vladimir Low flow hydrology: a review [Časopis] // Journal of Hydrology. - 2001. - T. 240. - str. 147-186.

[33] Srinivas V.V. / Srinivasan K. Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows [Časopis] // Journal of Hydrology. - 2005. - T. 302. - str. 307-330.

[34] Turlach B A Bandwidth Selection in Kernel Density Estimation: A Review [Časopis]. - [s.l.] : C.O.R.E. and Institut de Statistique, Universite Catholique de Louvain, 1993.

[35] Wang Q.J. / Nathan R.J. A method for coupling daily and monthly time scales in stochastic generation of rainfall series [Časopis] // Journal of Hydrology. - 2007. - T. 346. - str. 122-130.

[36] Wasserman Larry All of Nonparametric Statistics [Sveska]. - New York : Springer Science and Business Media, Inc., 2006.

NONPARAMETRIC ESTIMATION OF DENSITY FUNCTIONS IN HYDROLOGY

by

Mr Đurica MARKOVIĆ
Faculty of Technical Sciences, University of Priština at Kosovska Mitrovica
Dr Jasna PLAVŠIĆ, Dr Miloš STANIĆ
Faculty of Civil Engineering, University of Belgrade
Dr Goran SEKULIĆ
Faculty of Civil Engineering, University of Podgorica, Montenegro

Summary

As an alternative procedure for parametric methods in statistical analysis, non-parametric methods are emerging mainly because of their flexibility and ability to completely automate the estimation of probability density functions. One of the main features imposed by this method, in regard to parametric methods, is the absence of need to choose distribution based on samples or combination of two theorectical distribution functions for distribution tails.

The paper presents advantages and disadvantages of kernel based probability distribution estimates on an example of fitting distributions of weekly flows for the purpose of simulating long hydrologic series. In this context, the non-parametric approach is particularly useful because of the possibilities to automate the whole process.

Keywords: nonparametric estimation of density functions, kernel density, bandwidth selector

Redigovano 19.07.2011.