

A sűrűségfüggvény becsléséről

Dávid Gábor

I. Legyen a ξ valószínűségi változó eloszlásfüggvénye $F(x)$, sűrűségfüggvénye $f(x)$. Ebben a cikkben X a valós számegyenes, $\mathcal{U}$ a Borelmérhető halmazok σ -algebrája, P a $\mathcal{U}$ -n az $F(x)$ által generált mérték, λ a Lebesgue-mérték.

A feladat: a ξ -re végzett megfigyelések segítségével becslést adni az $f(x)$ függvényre.

Jelöljük a továbbiakban $\xi_1, \xi_2, \dots, \xi_n$ -el a ξ -re végzett n megfigyelt értéket, amelyek tehát függetlenek, azonos $F(x)$ eloszlásúak. ξ^* jelölje a megfigyelés sorrendjében vett mintaelemeket

$$\xi^* = \{\xi_1, \xi_2, \dots, \xi_n, \dots\}$$

A ξ^* eloszlását és az általa generált mértéket a $(\bar{X}_n, \bar{X}_n \mathcal{U})$ mérhető téren ugyancsak $F(x)$ -el és P -vel fogjuk jelölni. n -elemű mintára definiálhatjuk $\mathcal{U}$ -n a tapasztalati mértéket: tetszőleges $A \in \mathcal{U}$ Borel-halmazra legyen

$$P_n(A) = \frac{k_A}{n}$$

ahol $k_A =$ (az A -ba eső mintapontok száma). Ekkor a tapasztalati eloszlásfüggvény:

$$F_n(x) = P_n((-\infty, x])$$

A sűrűségfüggvény becslésével kapcsolatban M. Rosenblatt és E. Parzen érték el eredményeket. Közlöm Parzen eredményét: Legyen $K(x)$ egy egységnormájú, tetszőleges $L_1(X, \lambda)$ -beli függvény, amely korlátos és $k(x) = o(\frac{1}{x})$ ha $x \rightarrow \infty$ tetszőleges $h(n)$ sorozatra, amelyre $\lim_{n \rightarrow \infty} nh^2 = \infty$, ekkor definiáljuk a tapasztalati sűrűségfüggvényt a következőképpen:

$$f_n(x) = \int_{-\infty}^{\infty} \frac{1}{n(n)} K\left(\frac{y-x}{n(n)}\right) dF_n(y) = \frac{1}{nh(n)} \sum_{j=1}^n K\left(\frac{x-\xi_j}{n(n)}\right) \quad /1/$$

Ha $f(x)$ egyenletesen folytonos valamely $[a, b]$ intervallumban, akkor

$$\lim_{n \rightarrow \infty} P\left[\sup_{x \in [a, b]} |f_n(x) - f(x)| < \varepsilon\right] = 1 \quad /2/$$

tetszőleges $\varepsilon > 0$ -ra.

Parzen eredményét Crasswel arra az esetre bizonyította, amikor X topologikus csoport, λ a Haár-mérték; D.O. Loftsgaarden és C.P. Quesenberry a többváltozós esetben bizonyították. Nadarya a sokkal erősebb

$$P\left[\lim_{n \rightarrow \infty} \sup_{x \in [a, b]} |f_n(x) - f(x)| = 0\right] = 1 \quad /3/$$

állítását igazolta.

Ebben a cikkben a tapasztalati sűrűségfüggvényt az /1/-től eltérő módon definiálom, amiről először kimutatom, hogy 1 valószínűséggel egyenletesen konvergál (1.Tétel), a konvergencia gyorsaságára bizonyítok egy tételt (2.Tétel). A III-ban egy egyszerűbben számolható függvényt fogok definiálni, amelynek a sztochasztikus konvergenciáját bizonyítom be (3.Tétel).

A következőkben gyakran fogom alkalmazni a matematikai statisztika alaptételét, a Glivenko-Cantelli-tételt, amit itt idézek:

$$\lim_{n \rightarrow \infty} \sup_{x \in (-\infty, +\infty)} |F_n(x) - F(x)| = 0$$

1 valószínűséggel.

II. Jelölje ξ_k^* a $\xi_1, \xi_2, \dots, \xi_n$ nagyság szerint k -adik elemét: $\xi_k^* \leq \xi_{k+1}^*$, $k = 1, 2, \dots, n-1$. Legyen tetszőleges n és m -re ($n < m$)

$$f_n^m(x) = \begin{cases} \frac{P_m(I_{(x)}^n)}{\lambda(I_{(x)}^n)}, & \text{ha } x \in [\xi_1^*, \xi_n^*] \\ 0, & \text{ha } x \in [a, b] \setminus [\xi_1^*, \xi_n^*] \end{cases} \quad /4/$$

ahol $I_{(x)}^n$ jelöli a (ξ_k^*, ξ_{k+1}^*) $k = 1, \dots, n-1$ intervallumok közül azt, amelyik x -et tartalmazza.

1. Tétel. Legyen $[a, b]$ olyan intervallum, amelyben $f(x)$ folytonos és az $[F(a), F(b)]$ intervallumban $F^{-1}(x)$ folytonos.

Ekkor

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \sup_{x \in [a, b]} |f_n^m(x) - f(x)| = 0$$

1 valószínűséggel.

Bizonyítás: Jelölje $A(n, \varepsilon)$ azt az eseményt, hogy

$$\overline{\lim}_{m \rightarrow \infty} \sup_{x \in [a, b]} |f_n^m(x) - f(x)| > 2\varepsilon$$

A Borel-Cantelli-lemma értelmében tételünket igazoltuk, ha belátjuk, hogy tetszőleges $\varepsilon > 0$ -ra a $\sum_{n=1}^{\infty} P(A(n, \varepsilon))$ sor konvergens. Jelölje

$$f_n(x) = \begin{cases} \frac{P(I_n^*(x))}{\lambda(I_n^*(x))} & \text{ha } x \in [a, b] \cap [\xi_n^*, \xi_n^*] \\ 0 & \text{különben} \end{cases}$$

az elméleti sűrűségfüggvény módosítását. Minthogy

$$\begin{aligned} [\overline{\lim}_m \sup_x |f_n^m(x) - f(x)| > 2\varepsilon] &\subseteq [\lim_m \sup_x |f_n^m(x) - f_n(x)| > \varepsilon] \cup \\ &\cup [\overline{\lim}_m \sup_x |f_n(x) - f(x)| > \varepsilon] \end{aligned} \quad /5/$$

- ahol a jobboldal második tagja a $\sup_x |f_n(x) - f(x)| > \varepsilon$ esemény - így elegendő a $\sum_{n=1}^{\infty} P(A(n, \varepsilon))$ sor konvergenciájához belátni a következő két állítást:

$$1/ \sum_{n=1}^{\infty} P(\sup_{x \in [a,b]} |f_n(x) - f(x)| > \varepsilon) < +\infty \quad \text{minden} \quad \varepsilon > 0 \text{ -ra}$$

$$2/ \sum_{n=1}^{\infty} P(\lim_m \sup_{x \in [a,b]} |f_n^m(x) - f(x)| > \varepsilon) < +\infty \quad \text{minden} \quad \varepsilon > 0 \text{ -ra.}$$

A 2/ állítás igazolása: Legyen $\delta_n = \inf_{x \in [a,b]} \lambda(I^n(x))$. $f(x)$ folytonossága miatt 1 valószínűséggel $\delta_n > 0$. Így

$$\begin{aligned} P(\lim_m \sup_{x \in [a,b]} |f_n^m(x) - f_n(x)| > \varepsilon) &= \\ &= P(\lim_m \sup_x \left| \frac{P_m(I^n(x))}{\lambda(I^n(x))} - \frac{P(I^n(x))}{\lambda(I^n(x))} \right| > \varepsilon) \\ &\leq P(\lim_m \sup_x |P_m(I^n(x)) - P(I^n(x))| > \varepsilon \delta_n) \\ &\leq P(\lim_m \sup_{x \in [a,b]} |P_m(I^n(x)) - P(I^n(x))| > 0) \end{aligned}$$

Az utóbbi a Glivenko-Cantelli-alaptétel miatt 0. Kaptuk, hogy a 2/-beli sor nemcsak konvergens, hanem az összege 0 is.

Az 1/ állítás igazolása: Vizsgáljuk rögzített $I^n = (\xi_k^*, \xi_{k+1}^*)$ intervallumra a

$$[\sup_{x \in I^n} |f_n(x) - f(x)| > \varepsilon] = [\sup_{x \in I^n} \frac{1}{\lambda(I^n)} \left| \int_{I^n} (f(t) - f(x)) dt \right| > \varepsilon]$$

eseményt. Tetszőleges $\varepsilon' > \varepsilon$ -hoz válasszuk meg $\delta = \delta(\varepsilon')$ -t, majd $\delta(\varepsilon')$ -höz ε' -t úgy, hogyha $x, y \in [a, b]$, akkor

a/ ha $|x-y| < \delta$, akkor $|f(x) - f(y)| < \varepsilon'$ és

b/ ha $|F(x) - F(y)| < \varepsilon'$ akkor $|x-y| < \delta$ legyen.

A tétel feltevései miatt - a/-ban az $f(x)$ folytonosságát, b/-ben az $F^{-1}(x)$ folytonosságát felhasználva a zárt $[a,b]$ illetőleg a zárt $[F(a), F(b)]$ intervallumban - ε'' , δ és ε' megválaszthatók így.

Ezekkel a jelölésekkel:

$$\begin{aligned} P\left[\sup_{x \in I^n} \frac{1}{\lambda(I^n)} \left| \int f(x) - f(t) dt \right| > \varepsilon\right] &\leq P\left[\sup_{x,y \in I^n} |f(x) - f(y)| > \varepsilon''\right] \\ &\leq 1 - P\left[\sup_{x,y \in I^n} |f(x) - f(y)| < \varepsilon''\right] = \\ &= 1 - P\left[\sup_{x,y \in I^n} |f(x) - f(y)| < \varepsilon'' \mid \lambda(I^n) < \delta\right] P[\lambda(I^n) < \delta] \\ &\quad - P\left[\sup_{x,y \in I^n} |f(x) - f(y)| < \varepsilon'' \mid \lambda(I^n) \geq \delta\right] P[\lambda(I^n) \geq \delta] \end{aligned} \quad /6/$$

A második tag első tényezője 1, hiszen a feltétel bekövetkezése a/ miatt maga után vonja, hogy

$$\sup_{x,y \in I^n} |f(x) - f(y)| < \varepsilon''$$

A b/ miatt és mert

$$I^n = (\xi_k^*, \xi_{k+1}^*)$$

$$P(F(\xi_{k+1}^*) - F(\xi_k^*) < \varepsilon') \leq P(\xi_{k+1}^* - \xi_k^* < \delta)$$

Az $\eta = F(\xi)$ transzformációval kapott, a $[0,1]$ -ben egyenletes eloszlású η valószínűségi változó $\eta_k^* = F(\xi_k^*)$ rendezett mintáira ismert:

$$P(\eta_{k+1}^* - \eta_k^* < \alpha) = B_{i,n}(\alpha)$$

ahol $B_{i,n}(x)$ az (i,n) -paraméterű béta-eloszlás, így

$$P(\xi_{k+1}^* - \xi_k < \delta) = B_{i,n}(\varepsilon') = \frac{\Gamma(1+n)}{\Gamma(1)\Gamma(n)} \int_0^{\varepsilon'} (1-t)^{n-1} dt = 1 - (1-\varepsilon')^n$$

A /6/ harmadik tagját 0-val becsüljük alulról. Összegezve

$$P\left(\sup_{x \in I^n} |f_n(x) - f(x)| > \varepsilon\right) \leq (1-\varepsilon')^n \quad /7/$$

ahol ε' független n -től, az egész $[a,b]$ -re univerzális állandó, így

$$P\left[\sup_{x \in [a,b]} |f_n(x) - f(x)| > \varepsilon\right] \leq n(1-\varepsilon')^n$$

amiből

$$\sum_{n=1}^{\infty} P\left[\sup_{x \in [a,b]} |f_n(x) - f(x)| > \varepsilon\right] \leq \sum_{n=1}^{\infty} n(1-\varepsilon')^n < +\infty$$

azaz 1/ valóban fennáll, és ezzel a tétel állítását igazoltuk.

Nyilvánvaló az $\tilde{F}_{n,m}(x) = \int_{-\infty}^x f_n(x) dx$ tapasztalati eloszlásfüggvényre a Glivenko-Cantelli-tétel.

Következmény

$$P\left[\lim_n \lim_m \sup_{x \in [-\infty, +\infty]} |\tilde{F}_{n,m}(x) - F(x)| = 0\right] = 1$$

A következő tétel az $f_n^m(x)$ konvergenciájának gyorsaságára vonatkozik és egyben rávilágít n és m viszonyára is.

Legyen f differenciálható, pozitív és az $x \in [a, b]$ pontban

$$f(x) > c > 0; |f'(x)| < C_1, f'(x) \neq 0$$

alkalmas C és C_1 konstansokkal.

Legyen tetszőleges $\alpha > 0$ -ra

$$\omega_1(n, \alpha, x) = \frac{c}{4C_1} (1 - e^{-\frac{1}{n}(\alpha \log n)})^{-1}$$

$$\omega_2(n, m, x) = \lambda(I^n(x)) \left[\frac{2}{m} P(I^n(x)) (1 - P(I^n(x)) \log \log m) \right]^{-\frac{1}{2}}$$

$$\omega(n, m, \alpha, x) = \frac{1}{2} \min(\omega_1(n, \alpha), \omega_2(n, m, x)) \quad /8/$$

A definíció szerint tehát $\omega(n, m, \alpha, x)$ függ a véletlentől az $I^n(x)$ -en keresztül! Bebizonyítjuk a következő tételt:

2. Tétel. Ha x -nek van olyan $[a, b]$ környezete, amelyre teljesülnek az 1. Tétel feltételei, valamint $f(x)$ pozitív, az x pontban differenciálható, deriváltja: $f'(x) \neq 0$, akkor a /8/-ben definiált $\omega(n, m, \alpha, x)$ függvényre

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \omega(n, m, \alpha, x) |f_n^m(x) - f(x)| \leq 1$$

1 valószínűséggel.

Bizonyítás.

Hasonlóan az előző tétel bizonyításához, elegendő belátni, hogy a

$$\sum_{n=1}^{\infty} P(\overline{\lim}_{m \rightarrow \infty} \omega(n, m, \alpha, x) | f_n^m(x) - f(x) | > 1 + \varepsilon)$$

sor konvergens minden $\varepsilon > 0$ -ra. Ugyanazzal az átalakítással, mint /5/-ben

$$[\overline{\lim}_{m \rightarrow \infty} \omega(n, m, \alpha, x) | f_n^m(x) - f(x) | > 1 + \varepsilon]$$

/9/

$$\leq [\overline{\lim}_m \omega(n, m, \alpha, x) | f_n^m(x) - f_n(x) | > \frac{1+\varepsilon}{2}] \cup$$

$$\cup [\overline{\lim}_m \omega(n, m, \alpha, x) | f_n(x) - f(x) | > \frac{1+\varepsilon}{2}]$$

Most

$$[\overline{\lim}_m \omega(n, m, \alpha, x) | f_n(x) - f(x) | > \frac{1+\varepsilon}{2}] \leq$$

$$\leq [\overline{\lim}_m \omega_1(n, \alpha) | f_n(x) - f(x) | > 1 + \varepsilon] =$$

$$= [\omega_1(n, \alpha) | f_n(x) - f(x) | > 1 + \varepsilon]$$

Az utóbbi esemény valószínűségének becsléséhez:

$$\frac{|F(x) - F(y)|}{|f(x) - f(y)|} = \frac{|F(x) - F(y)|}{|x - y|} \cdot \frac{|x - y|}{|f(x) - f(y)|} = \frac{f(x) + O(x-y)}{|f'(x) + O(x-y)|} \approx \frac{f(x) + O(x-y)}{2C_1} \quad /10a/$$

(ahol az utóbbi becslés egy null-mértékű halmaz kivételével érvényes).

Másrészt

$$\{ |f_n(x) - f(x)| > \frac{1}{\omega_1(n, \alpha, x)} \} \leq \{ \sup_{x, y \in I^n} |f(x) - f(y)| > \frac{1}{\omega_1(n, \alpha, x)} \}$$

és ez utóbbi maga után vonja, hogy van olyan $x, y \in [\xi_{k-1}^*, \xi_k^*]$ hogy $|f(x) - f(y)| > \frac{1}{\omega_1(n, \alpha, x)}$, amiből - a /10a/ becslést alkalmazva -

$$|F(\xi_{k-1}^*) - F(\xi_k^*)| > |F(x) - F(y)| > \frac{f(x) + O(x-y)}{2c_1} \cdot \frac{1}{\omega_1(n, \alpha, x)}$$

következik. Összegezve

$$\begin{aligned} P(|f_n(x) - f(x)| > \frac{1}{\omega_1(n, \alpha, x)}) &\leq P(F(\xi_k^*) - F(\xi_{k-1}^*) > \frac{f(x) + O(x-y)}{2c_1} \cdot \frac{1}{\omega_1(n, \alpha, x)}) = \\ &= P(F(\xi_k^*) - F(\xi_{k-1}^*) > \frac{f(x) + O(x-y)}{2c_1} \cdot \frac{1}{c} (1 - e^{-\frac{1}{n}(1+\alpha)\log n})) \quad /10b/ \end{aligned}$$

Itt $[f(x) + O(x-y)]2/c > 1$ egy valószínűséggel, hiszen $(x-y) \leq |\xi_k^* - \xi_{k-1}^*|$ és $|\xi_{k-1}^* - \xi_k^*| \rightarrow 0$ egy valószínűséggel, ha $n \rightarrow \infty$, így - folytatva a /10b/-nél abbahagyott becslést -

$$\begin{aligned} &\leq P(F(\xi_k^*) - F(\xi_{k-1}^*) > (1 - e^{-\frac{1}{n}(1+\alpha)\log n})) = \\ &= [1 - (1 - e^{-\frac{1}{n}(1+\alpha)\log n})]^n = \frac{1}{n^{1+\alpha}} \end{aligned}$$

tehát

$$\sum_{n=1}^{\infty} P(|f_n(x) - f(x)| \omega_1(n, \alpha, x) > 1 + \varepsilon) < +\infty \quad /10/$$

minden $\varepsilon > 0$ -ra.

A /9/ jobboldalán álló első eseményre

$$\begin{aligned}
 & P[\lim_m \omega(n, m, \alpha, x) | f_n^m(x) - f_n(x) | > \frac{1+\varepsilon}{2}] \leq \\
 & \leq P[\lim_m \omega_2(n, m, x) | f_n^m(x) - f_n(x) | > 1+\varepsilon] \quad /11/ \\
 & = P[\lim_m \frac{\lambda(I^n(x))}{\sqrt{\frac{2}{m}} P(I^n)(1-P(I^n)) \log \log m} | \frac{P_m(I^n(x))}{\lambda(I^n(x))} - \frac{P(I^n(x))}{\lambda(I^n(x))} | > 1+\varepsilon]
 \end{aligned}$$

A $\lambda(I^n(x))$ -el egyszerűsítve, az iterált-logaritmus tételle miatt tetszőleges $\varepsilon > 0$ -ra az utóbbi valószínűség 0, azaz

$$\sum_{n=1}^{\infty} P[\lim_m \omega(n, m, \alpha) | f_n^m(x) - f_n(x) | > 1+\varepsilon] = 0 \quad /12/$$

A /10/ és /12/ bizonyítja a tételt.

Megjegyzés. Kézenfekvő lett volna a tapasztalati sűrűségfüggvényt a

$$g_n(x) = \begin{cases} \frac{P_n(I^n(x))}{\lambda(I^n(x))} & \text{ha } x \in [\xi_1^n, \xi_n^n] \\ 0 & \text{különben} \end{cases}$$

definícióval megadni. A $g_n(x)$ azonban nem konvergens:

A $g_n(x)$ pozitív valószínűséggel nem konvergál egyenletesen 0-hoz.

Tegyük ugyanis ennek ellenkezőjét fel valamely $\xi_1, \xi_2, \dots, \xi_n, \dots$ mintaelem-sorozatra. Létezik olyan $n_i, i = 1, 2, \dots$ indexsorozat, hogy

$$\sup_{x \in (-\infty, +\infty)} |g_{n_i}(x) - f(x)| < \frac{1}{2^i}$$

Jelölje γ a ξ_{n_i+1} , az η és ζ a ξ_k^* illetve a ξ_{k+1}^* mintaelemeket, ha a $\xi_{n_i+1} = \gamma$ a (ξ_k^*, ξ_{k+1}^*) intervallumba esik. (Hogy $\gamma \in (\xi_k^*, \xi_{k+1}^*)$ annak valószínűsége, $n \rightarrow \infty$ esetén 1-hez tart).

Ekkor

$$\begin{aligned} B &= \sup_x |g_{n_i+1}(x) - f(x)| = \\ &= \max \left\{ |g_{n_i+1}(x) - f(x)|; x \notin (\gamma, \zeta), \left| \frac{1}{(n_i+1)(\zeta-\eta)} - f(x) \right|; x \in (\gamma, \zeta) \right\} \\ &\quad \left| \frac{1}{(n_i+1)(\zeta-\eta)} - f(x) \right|; x \in (\eta, \gamma) \} \end{aligned}$$

Az

$$A = \max \left(\left| \frac{1}{(n_i+1)(\zeta-\eta)} - \frac{1}{n_i(\zeta-\eta)} \right|, \left| \frac{1}{(n_i+1)(\zeta-\eta)} - \frac{1}{n_i(\zeta-\eta)} \right| \right)$$

akkor minimális, ha $(\vartheta - \eta) = (\zeta - \vartheta) = \frac{\zeta - \eta}{2}$

és

$$A \geq \left(\frac{2}{(n_i + 1)(\zeta - \eta)} - \frac{1}{n_i(\zeta - \eta)} \right)$$

Igy elég nagy i -re

$$B \geq \max \left\{ \frac{1}{2}, |A - g_{n_i}(x) - f(x)| \right\} \geq \frac{1}{2} f(x)$$

ha $f(x) > 0$

$$\lim_{i \rightarrow \infty} \sup_x |g_{n_i+1}(x) - f(x)| > 0$$

azaz valóban divergens.

III. A /4/-ben definiált tapasztalati sűrűségfüggvény nagy minta-elemszám esetén nehezen számítható. Egy olyan módosítást adom ebben a részben, amit Révész Pál vetett fel.

A gyakorlatban jó lenne, ha a tapasztalati sűrűségfüggvényt - mondjuk - csak minden $\sqrt{n}$ -edik rendezett minta által meghatározott intervallumokban adnánk meg n -elemű minta esetén, és egy-egy ilyen intervallumban úgy definiáljuk, hogy legyen ezen intervallumban a tapasztalati mérték - azaz $\frac{\sqrt{n}}{n}$ - és az intervallum hosszának

hányadosa. Itt /13/-ban a most heurisztikusan elmondott definíciót precizírozom és a /13/-ban adott tapasztalati sűrűségfüggvény sztochasztikus konvergenciáját vizsgálom.

A továbbiakban $[y]$ az y egész részét jelöli.

Tetszőleges $\varepsilon_n > 0$ -ra az $I_k = (\xi_k^* [\varepsilon_n n], \xi_{k+1}^* [\varepsilon_n n])$ jelöléssel legyen rögzített n -re

$$f_n^{\varepsilon_n}(x) = \begin{cases} \frac{P_n(I_k)}{\lambda(I_k)} & \text{ha } x \in I_k \\ 0 & \text{különben} \end{cases} \quad \begin{matrix} /13/ \\ k = 1, 2, \dots, [\frac{1}{\varepsilon_n}] - 1 \end{matrix}$$

3.Tétel. Legyen a $\xi(\omega)$ valószínűségi változó $F(x)$ eloszlásfüggvénye szigorúan monoton, sűrűségfüggvénye: $f(x)$ egyenletesen folytonos, korlátos és pozitív: $0 < f < c$ és a /13/-ban szereplő $\{\varepsilon_n\}$ sorozatra legyen

$$\varepsilon_n = O\left(\frac{1}{\sqrt{n^{1-\alpha}}}\right)$$

ahol α tetszőleges kis pozitív szám. Ekkor

$$\lim_{n \rightarrow \infty} P\{|f_n^{\varepsilon_n}(x) - f(x)| > \delta\} = 0$$

tetszőleges $\delta > 0$ -ra

Bizonyítás: Legyen

$$f^{\varepsilon_n}(x) = \begin{cases} \frac{1}{\lambda(I_k)} \int_{I_k} f(t) dt & \text{ha } x \in I_k, k=1, 2, \dots, [\frac{1}{\varepsilon_n}] - 1, \\ 0 & \text{különben} \end{cases}$$

Vizsgáljuk az $|f_n^{\varepsilon_n}(x) - f^{\varepsilon_n}(x)|$ valószínűségi változó várható értékét.

Felhasználva a Schwartz-egyenlőtlenséget

$$M(|f_n^{\varepsilon_n}(x) - f^{\varepsilon_n}(x)|) \leq M^{\frac{1}{2}}\left(\frac{1}{\lambda(I_k)^2}\right) M^{\frac{1}{2}}\left(\left(\frac{[\varepsilon_n \cdot n]}{n} - P(I_k)\right)^2\right) = /14/$$

Felhasználva a könnyen igazolható

$$M(|P_n(-\infty, \xi_k^*) - P(-\infty, \xi_k^*)|^2) = O\left(\frac{k}{n^2}\right)$$

becslést, folytatva /14/-et

$$= M^{\frac{1}{2}}\left(\frac{1}{\lambda(I_k)^2}\right) O\left(\frac{1}{\sqrt{n}}\right)$$

Igy tetszőleges $\delta > 0$ mellett a Markov-egyenlőtlenséget felhasználva

$$P\{|f_n^{\varepsilon_n}(x) - f^{\varepsilon_n}(x)| > \delta\} \leq M^{\frac{1}{2}}\left(\frac{1}{\lambda(I_k)^2}\right) O\left(\frac{1}{\sqrt{n}}\right) \quad /15/$$

Itt

$$M\left(\frac{1}{\lambda(I_k)^2}\right) = M\left(\frac{P_n^2(I_k)}{\lambda(I_k)^2} \cdot \frac{1}{P_n^2(I_k)}\right) = O\left(\frac{1}{\varepsilon_n^2}\right) M^2\left(\frac{P_n^4(I_k)}{\lambda(I_k)^4}\right)$$

A Glivenko-Cantelli tétel miatt 1 valószínűséggel

$$P_n(I_k) \approx \int_{I_k} f(t) dt \leq C \lambda(I_k)$$

azaz

$$M\left(\frac{1}{\lambda(I_k)^2}\right) = O\left(\frac{1}{\varepsilon_n^2}\right)$$

Behelyettesítve /15/-be

$$P\{|f_n^{\varepsilon_n}(x) - f(x)| \geq \delta\} = O\left(\frac{1}{\varepsilon_n} \cdot \frac{1}{\sqrt{n}}\right) \quad /16/$$

Válasszuk δ és δ' -t úgy, hogyha $|x-t| < \delta$, akkor $|f(x) - f(t)| < \delta'$ legyen. Mivel

$$P\left\{\xi_{[k+1), [\varepsilon_n n]}^* - \xi_{[k, \varepsilon_n n]}^* \geq \delta'\right\} \leq O\left(\frac{[\varepsilon_n n]}{n+1}\right) = O(\varepsilon_n)$$

következik - a /6/-nál használt átalakítás megismétlésével és egyszerű becslésekkel

$$P\{|f_n^{\varepsilon_n}(x) - f(x)| > \delta\} = O(\varepsilon_n) \quad /17/$$

Mostmár /17/ és /16/-ből a tett feltevések miatt következik, hogy

$$\lim_{n \rightarrow \infty} P \{ |f_n^{\xi_n}(x) - f(x)| > \delta \} = 0$$

Mivel $f(x) > 0$, így $\xi_1^* \rightarrow -\infty$ és $\xi_n^* \rightarrow +\infty$
1 valószínűséggel, ha $n \rightarrow \infty$, így a tétel bármely x -re igaz.

Befejezésül köszönetet szeretnék mondani Csáki Péternek, aki a dolgozat lektorálása során sok hasznos megjegyzést tett.

Irodalom:

- Rosenblatt, M./1956/: Remarks on some nonparametric estimates of density function. Ann.Math.Stat. Vol.27. /832-835/
Parzen, E./1962/: On estimates of a probability density and mode. Ann.Math.Stat.Vol 33 (1065-1076)
Loftsgaarden, D.C. and Quesenberry, C.P. /1965/. A nonparametric estimate of a multivariate density function. Ann.Math.Stat. Vol 36. (1049-1051).

Craswel /1965/. Density estimation in a topological group. Ann.Math.Stat. Vol. 36 (1047-1048).

Nadarya /1965/. On nonparametric estimates of density functions and regression curve. Theory of Probability and its Applications 10. (186-190) (In Russian).

S u m m a r y

The author in (4) gives a new definition of empirical density function for which he proves the following theorems:

Theorem 1. Let $[a, b]$ be such an interval, in which $f(x)$ and the inverse of $F(x)$ in $[F(a), F(b)]$ is continuous then

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \sup_{x \in [a, b]} |f_n^m(x) - f(x)| = 0$$

with probability one.

If the function $\omega(n, m, \alpha, x)$ is defined in (8), then

Theorem 2. Under conditions of Theorem 1, if $x \in [a, b]$ and if $f(x) > c > 0$, and $f(x)$ has first derivate $f'(x)$, for which $|f'(x)| < c$, and non-zero, then

$$\overline{\lim}_n \overline{\lim}_m \omega(n, m, \alpha, x) |f_n^m(x) - f(x)| \leq 1$$

with probability one.

For brevity of computation the author gives an other definition of empirical density function in (13), for which the following theorem holds.

Theorem 3. If the density function $f(x)$ is uniformly continuous, bounded and positive: $0 < f(x) < c$ and for all ε_n (in (13))

$$\varepsilon_n = O\left(\frac{1}{\sqrt{n^{1-\alpha}}}\right)$$

(where α is an arbitrary small positive number) holds then

$$\lim_{n \rightarrow \infty} P[f_n^{\varepsilon_n}(x) - f(x) > \delta] = 0$$

for arbitrary $\delta > 0$.