Dávid Gábor:

Az eloszlás- és sürüségfüggvény becslésének konvergenciája.

0.§.

Bevezetés: A szerző [1] dolgozatában foglalkozott olyan csoportos mintavétel segitségével történő eloszlás- és sürüségfüggvénybecsléssel, amikor a csoportositás a minta alapján történik.

Az uj definicióra azért volt szükség, mert a Glivenkó-Cantelli-tétel alkalmazásához a minta egészének ismerete szükséges, ami a Monte-Carlo-módszerek esetén a számitógép memóriájának nagysága miatt lehetetlen. Az [1] dolgozatban az 1.Tétel, a Következmény a véletlentől függő tapasztalati sürüség- és eloszlásfüggvény egyenletes konvergenciájára vonatkoznak. Felmerül azonban az a kérdés, hogy ez a konvergencia milyen gyors. Egy rögzitett pontban erre a 2.Tétel válaszolt. Az irodalomban régóta foglalkoznak ezzel a problémával, amit a következőképpen fogalmazunk meg:

ha $\alpha(x)$ az elméleti függvény, $\alpha_N(x)$ a tapasztalati, akkor

$$P\left(\overline{\lim_{N\to\infty}}\, \Omega_N \sup_{x\in[a,b]} |\alpha_N(x) - \alpha(x)| \leq 1\right) = 1$$

milyen Ω_N mellett áll fent ($\Omega_N \to \infty$ ha $N \to \infty$).

Ezideig - és csak a tapasztalati eloszlásfüggvényre - a

$$P\left(\sup_{x\in[a,b]} |F_n(x) - F(x)| \geq \varepsilon\right)$$

függvény konvergenciájával kapcsolatban értek el eredményeket. Itt csak Sanov [2] és Sethuraman [3] dolgozataira utalok, ahol bőséges irodalomjegyzék található.

Az 1.§-ban a minta nagyságát vizsgálom, a 2.§-ban a konvergencia gyorsaságával foglalkozom az [1]-ben definiált sürüség - és eloszlásfüggvényre, majd a 3.§-ban az egyszerü csoportositott tapasztalati eloszlásfüggvényre.
A jelölések megegyeznek az [1] dolgozat jelöléseivel.

1.§.

Megjegyzés a tapasztalati sürüségfüggvény konvergencia-gyorsaságához.

Az 1 -ben definiált $\omega(n,m,\alpha,x)$ gyorsasági-modulus vizsgálatával kezdjük:

$$\omega(n,m,\alpha,x) = \frac{1}{2}\min\left(\omega_1(n,\alpha),\ \omega_2(n,m,x)\right)$$

ahol

$$\omega_1(n,\alpha) = \frac{C}{4C_1}\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^{-1},\quad f(x)>C>0,\ |f'(x)|<C_1,\ f'(x)\neq 0$$

és

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

ahol $I^n_{(x)}$ az X -et tartalmazó, az n-elemü rendezett minta $[\xi^*_k,\ \xi^*_{k+1}]$ intervalluma. Jelölje egyenlőre $\lambda(I^n_{(x)}) = \delta(x)$ és $P(I^n_{(x)}) = \eta(x)$. Igy

$$\frac{\omega_2(n,m,x)}{\omega_1(n,\alpha)} = \frac{\delta\sqrt{\frac{m}{2\eta(1-\eta)\log\log m}}}{\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^{-1}}\frac{4C_1}{C} = \delta\sqrt{\frac{m\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^2}{2\eta(1-\eta)\log\log m}}\cdot\frac{4C_1}{C}$$

azaz ha

$$\frac{m\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^2}{\log\log m} \geq \frac{2P(I^n_{(x)})(1-P(I^n_{(x)}))}{(\xi^*_{k+1}-\xi^*_k)^2}\cdot\left(\frac{C}{4C_1}\right)^2 \tag{/1/}$$

reláció teljesül n és m-re, ugy

$$\omega(n,m,\alpha,x) = \frac{1}{2}\,\omega_1(n,\alpha) = \frac{C}{8C_1}\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^{-1}$$

Az /1/ egyenlőtlenséget még a következőképpen alakithatjuk át: Mivel $\frac{P(I^n(x))}{\lambda(I^n(x))} \to f(x)$ ha $n \to \infty$ egy λ-0 -mértékü halmaz kivételével /az integrálszámitás középértéktétele/, igy az asszimptotikus

$$/2/ \qquad \frac{m\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^2}{\log\log m} \geq 2f(x)\left(\frac{1}{\xi^*_{k+1}-\xi^*_k} - f(x)\right)\left(\frac{C}{4C_1}\right)^2$$

reláció fennállása esetén is a gyorsasági modulus az x-től független $\frac{1}{2}\omega_1(n,\alpha)$ tag. Előnye a /2/ egyenlőtlenségnek, hogy az n és m viszonyára viszonylag egyszerü összefüggést ad, hiszen a /2/ jobboldalán /a C és C_1 konstansokkal jól becsülhető f(x) sürüségfüggvényen kivül/ az n-elemü mintából könnyen kiszámitható $\xi^*_{k+1} - \xi^*_k$ intervallumhosszak segitségével adtuk meg.

<u>1.Tétel.</u> Ha n és m-re teljesül az /1/, és az f(x) teljesiti az [1] 2. Tétel feltételeit, akkor az $\omega(n,\alpha) = \frac{C}{8C_1}\left(1-e^{-\frac{1}{n}(1+\alpha)\log n}\right)^{-1}$ jelöléssel:

$$P\left(\overline{\lim_n}\ \overline{\lim_m}\ \omega(n,\alpha)\,|f_n^m(x)-f(x)| \leq 1\right) = 1$$

<u>Bizonyitás:</u> Ha /1/-ben egyenlőség esetén m=g(n), akkor m≥g(n)-re méginkább fennáll /1/, továbbá $\varepsilon > 0$-ra

$$P\left(\overline{\lim_{m\geq g(n)}}\ \omega(n,\alpha)\,|f_n^m(x)-f(x)| > 1+\varepsilon\right) =$$

$$= P\left(\overline{\lim_m}\ \omega(n,m,\alpha,x)\,|f_n^m(x)-f(x)| > 1+\varepsilon\right)$$

Igy a

$$\sum_{n=1}^{\infty} P\left(\overline{\lim_{m\geq g(n)}}\ \omega(n,\alpha)\,|f_n^m(x)-f(x)| > 1+\varepsilon\right)$$

sor konvergens az [1] 2.Tétel miatt, ami a Borel-Cantelli-lemmával bizonyitja a tételt.

2.§.

<u>Egyenletes-konvergencia-gyorsasági tételek a tapasztalati eloszlás- és sürüségfüggvényre</u>

Legyen

$$\omega_1^*(n,\alpha) = \frac{C}{4C_1}\left(1 - e^{-\frac{1}{n}(2+\alpha)\log n}\right)^{-1} \qquad /3/$$

és

$$\omega^*(n,m,\alpha,x) = \frac{1}{2}\min\left(\omega_1^*(n,\alpha),\ \omega_2(n,m,x)\right)$$

tetszőleges $\alpha > 0$ -ra. n és m teljesitse a

$$\frac{m\left(1 - e^{-\frac{1}{n}(2+\alpha)\log n}\right)^2}{\log\log m} \geq \frac{2P(I_{(x)}^n)(1 - P(I_{(x)}^n))}{(\xi_{k+1}^* - \xi_k^*)^2}\left(\frac{C}{4C_1}\right)^2 \qquad /4/$$

egyenlőtlenséget. Bebizonyitjuk a következő tételt:

<u>2.Tétel.</u> Legyen [a,b] olyan intervallum, hogy az [F(a), F(b)]-ben $F^{-1}(x)$ folytonos, [a,b]-ben f(x) pozitiv, differenciálható, deriváltjára $f'(x) \neq 0$, akkor a /3/-ban definiált $\omega_1^*(n,\alpha)$ függvényre

$$\overline{\lim_{n}}\ \overline{\lim_{m}}\ \frac{1}{2}\,\omega_1^*(n,\alpha) \sup_{x \in [a,b]} \left| f_n^m(x) - f(x) \right| \leq 1$$

1 valószinüséggel, ha n és m kielégiti /4/-et.

<u>Bizonyitás:</u> Az [1]. 2.Tétel bizonyitásához hasonlóan elvégezve az ott használt felbontást, alkalmazzuk a

$$\sum_{n=1}^{\infty} P\Big(\overline{\lim_{m>g(n)}}\ \frac{1}{2}\,\omega_1^*(n,\alpha) \sup_{x\in[a,b]} |f_n^m(x)-f(x)| > 1+\varepsilon\Big) \le$$

/5/

$$\le \sum_{n=1}^{\infty} n \max_{i=1,2,\dots n} \Big\{P\Big(\overline{\lim_{m}}\ \omega^*(n,m,\alpha,\xi_i^*)\,|f_n^m(\xi_i^*)-f(\xi_i^*)| > 1+\varepsilon\Big)\Big\}$$

becslést, ahol felhasználtuk azt, hogy

$$P\Big(\overline{\lim_{m}}\ \frac{1}{2}\,\omega_1^*(n,\alpha) \sup_{x\in[a,b]} |f_n^m(x)-f(x)| > 1+\varepsilon\Big) \le$$

/6/

$$\le \sum_{i=1}^{n} P\Big(\overline{\lim_{m}}\ \omega^*(n,m,\alpha,\xi_i)\,|f_n^m(\xi_i)-f(\xi_i^*)| > 1+\varepsilon\Big)$$

ugyanis az $f(x)$ monotonitása miatt a szuprémum az osztópontokban vevődik fel $\big(f'(x) \neq 0\big)$.

Az /5/ jobboldalán álló valószinüségeket már megbecsültük az emlitett tétel bizonyitásában:

Az előző cikk jelöléseivel:

$$P\Big(\overline{\lim_{m>g(n)}}\ \frac{1}{2}\,\omega_1^*(n,\alpha)\,|f_n^m(\xi_i^*) - f(\xi_i^*)| > 1+\varepsilon\Big) \le$$

$$\le P\Big(\overline{\lim_{m\ge g(n)}}\ \frac{1}{2}\,\omega_2^*(n,\alpha)\,|f_n^m(\xi_i^*) - f_n(\xi_i^*)| > \frac{1+\varepsilon}{2}\Big) +$$

$$+ P\Big(\overline{\lim_{m>g(n)}}\ \frac{1}{2}\,\omega_1^*(n,\alpha)\,|f_n(\xi_i^*) - f(\xi_i^*)| > \frac{1+\varepsilon}{2}\Big) \le$$

$$\le P\Big(\overline{\lim_{m\ge g(n)}}\ \omega_2(n,m,\xi_i^*)\,|f_n^m(\xi_i^*) - f_n(\xi_i^*)| > \frac{1+\varepsilon}{2}\Big) +$$

$$+ P\Big(\overline{\lim_{m\ge g(n)}}\ \omega_1^*(n,\alpha)\,|f_n(\xi_i^*) - f(\xi_i^*)| > \frac{1+\varepsilon}{2}\Big) \le$$

$$\le P\Big(\overline{\lim_{m}}\ \omega_2(n,m,\xi_i^*)\,|f_n^m(\xi_i^*) - f_n(\xi_i^*)| > \frac{1+\varepsilon}{2}\Big) +$$

$$+ P\Big(\omega_1^*(n,\alpha)\,|f_n(\xi_i^*) - f(\xi_i^*)| > \frac{1+\varepsilon}{2}\Big)$$

Az első tag - mint [1]./11/-ben az iterált-logaritmus-tétel miatt 0, mig a második tagot az [1]. /10/ becsléséhez hasonlóan becsüljük meg;

$$\sum_{n=1}^{\infty} n P\left(\omega(n,m,\xi_i) \left| f_n^m(\xi_i^*) - f(\xi_i^*) \right| > \frac{1+\varepsilon}{2}\right) \leq \sum_{n=1}^{\infty} n\left[1-\left(1-e^{-\frac{1}{n}(2+\alpha)\log n}\right)\right]^n \leq \sum_{n=1}^{\infty} \frac{1}{n^{1+\alpha}} < +\infty$$

A tétel bizonyitását most már a szokásos módszerrel fejezhetjük be.

Legyen a tapasztalati eloszlásfüggvény: $F_n^m(x) = \int_a^x f_n^m(t)\,dt$, az $[a, b]$ véges intervallumban.

<u>3.Tétel.</u> A 2. Tétel feltételei mellett, ha $\Omega(n,\alpha) = \frac{\omega_1^*(n,\alpha)}{2(b-a)}$ akkor

$$P\left(\overline{\lim_{n}}\,\overline{\lim_{m \geq g(n)}}\, \Omega(n,\alpha) \sup_{x \in [a,b]} \left| F_n^m(x) - F(x) \right| \leq 1\right) = 1$$

<u>Bizonyitás</u> a következő becslés-sorozattal történik:

$$\begin{aligned}
&\left\{ \Omega(n,\alpha) \sup_x \left| F_n^m(x) - F(x) \right| > 1 \right\} \leq \\
&\leq \left\{ \Omega(n,\alpha) \sup_x \left| \int_a^x f_n^m(t) - f(t)\,dt \right| > 1 \right\} \leq \\
&\leq \left\{ \Omega(n,\alpha) \sup_x \int_a^x \left| f_n^m(t) - f(t) \right| dt > 1 \right\} \leq \\
&\leq \left\{ \Omega(n,\alpha)(b-a) \sup_x \left| f_n^m(x) - f(x) \right| > 1 \right\} \leq \\
&\leq \left\{ \frac{1}{2} \omega_1^*(n,\alpha) \sup_{x \in [a,b]} \left| f_n^m(t) - f(t) \right| > 1 \right\}
\end{aligned}$$

azaz

$$\begin{aligned}
&P\left(\overline{\lim_{n}}\,\overline{\lim_{m \geq g(n)}}\, \Omega(n,\alpha) \sup_{x \in [a,b]} \left| F_n^m(x) - F(x) \right| > 1\right) \leq \\
&\qquad \leq P\left(\overline{\lim_{n}}\,\overline{\lim_{m \geq g(n)}}\, \frac{1}{2} \omega_1^*(n,\alpha) \sup_{x \in [a,b]} \left| f_n^m(t) - f(t) \right| > 1\right) = 0
\end{aligned}$$

ami a tételt bizonyitja.

3.§.

A csoportositott tapasztalati eloszlásfüggvény egyenletes konvergenciájának gyorsaságáról

A 3. Tételben az $F_n^m(x)$ egy monoton növekedő függvény és gyorsaságát a sürüségfüggvény gyorsaságából vezettük le. Valószinüleg a 3. Tétel állitása nem éles, és a feltételek is erősek.

A csoportositott mintavétel az [1] cikkben és a jelenlegi 1.§ és 2.§-okban az előzőleg vett n-elemü minta alapján történt. Most a számegyenest osszuk fel a_i^n $i=1,\ldots,n$ osztópontokkal ugy, hogy $n\to\infty$ esetén $a_{i+1}^n - a_i^n \to 0$ legyen. /A véletlen felosztásnál ez csak 1 valószinüséggel áll fent./ Végezzünk m független kisérletet és $n \ll m$ legyen, $n\to\infty$ és jelölje m_i az i-edik intervallumba eső mintapontok számát, továbbá legyen

$$G_m(x)=\frac{\sum_{i=1}^{j(x)} m_i}{m} \qquad \text{ha} \qquad x\in \left(a_{j(x)}^n,\ a_{j(x)+1}^n\right] .$$

Tegyük fel, hogy az $[a,b]$ olyan intervallum, hogy $0<C_2<F(x)<C_3<1$.

Legyen $\frac{1}{\Omega(m)}<F(x)(1-F(x))$ minden $x\in[a,b]$-re.

Felhasználva az $F(x)$ monotonitását

$$P\left(\Omega(m)\sup_{x\in[a,b]}\left|G_m(x)-F(x)\right|>1+\varepsilon\right)\leq\sum_{k=1}^{n}P\left(\Omega(m)\left|\frac{\sum_{i=1}^{k-1}m_i}{m}-F(a_k^n)\right|>1+\varepsilon\right)$$

Ismert /Rényi Alfréd: Valószinüségszámitás /1954/ 405. oldal/, hogyha ζ az A esemény relativ gyakorisága és $P(A)=p$, akkor, ha $0<\varepsilon<p(1-p)$, ugy

$$P(|\zeta_m - p| \geq \varepsilon) \leq 2e^{-\frac{\varepsilon^2 m}{2pq(1+\frac{\varepsilon}{2pq})^2}}$$

Alkalmazva ezt a tételt, kapjuk, hogy

$$P\left(\left|\frac{\sum_{i=1}^{k-1} m_i}{m} - F(a_k^n)\right| \geq \frac{1+\varepsilon}{\Omega(m)}\right) \leq 2e^{-\frac{\left(\frac{1+\varepsilon}{\Omega(m)}\right)^2 m}{2F(a_k^n)(1-F(a_k^n))\left(1+\frac{\frac{1+\varepsilon}{\Omega(m)}}{2F(a_k^n)(1-F(a_k^n))}\right)^2}}$$

Mivel $\max F(x)(1-F(x)) = \frac{1}{4}$ és

$$\frac{1+\varepsilon}{\Omega(m) 2F(x)(1-F(x))} \leq \frac{1+\varepsilon}{2} < 1$$

így

$$P\left(\Omega(m)\left|\frac{\sum_{1}^{k-1} m_i}{m} - F(a_k^n)\right| > 1+\varepsilon\right) \leq 2e^{-\frac{m}{2\Omega^2(m)}}$$

azaz

$$\sum_{m=1}^{\infty} \sum_{k=1}^{n} P(\Omega(m)|G_m(a_k^n) - F(a_k^n)| > 1+\varepsilon) \leq 2\sum_{m=1}^{\infty} m e^{-\frac{m}{2\Omega^2(m)}} = \sum_{m=1}^{\infty} \frac{1}{m^{1+\alpha}} < +\infty$$

ha $e^{-\frac{m}{2\Omega^2(m)}} = \frac{1}{m^{2+\alpha}}$, amiből

$$\Omega(m) = \sqrt{\frac{m}{2(2+\alpha)\log m}}$$

és ekkor

$$\sum_{m=1}^{\infty} P\left(\Omega(m) \sup_{x \in [a,b]} |G_m(x) - F(x)| > 1+\varepsilon\right) < +\infty$$

minden $\varepsilon > 0$-ra.

Összegezve:

4. Tétel: Ha $[a,b]$ olyan intervallum, hogy $0<C_2<F(a)<F(b)<C_3<1$ és tetszőleges $\alpha>0$ -ra

$$\Omega(m)=\sqrt{\frac{m}{2(2+\alpha)\log m}}$$

akkor

$$P\left(\overline{\lim_{m}}\ \Omega(m)\sup_{x\in[a,b]}|G_m(x)-F(x)|\leq 1\right)=1$$

Továbbra is nyitott kérdés maradt az, hogy vajon ezek a gyorsaságimodulusok a leggyorsabbak-e.

4.§.

Egy alkalmazás:

A feladatunk a

$$\xi=\frac{\zeta\varrho_1-0{,}3\varrho_2-0{,}1\varrho_1\varrho_2}{\varrho_1+\varrho_2}$$

$$\eta=\frac{-8.4\beta}{\varrho_1}-\frac{\zeta+0{,}4}{\varrho_2}$$

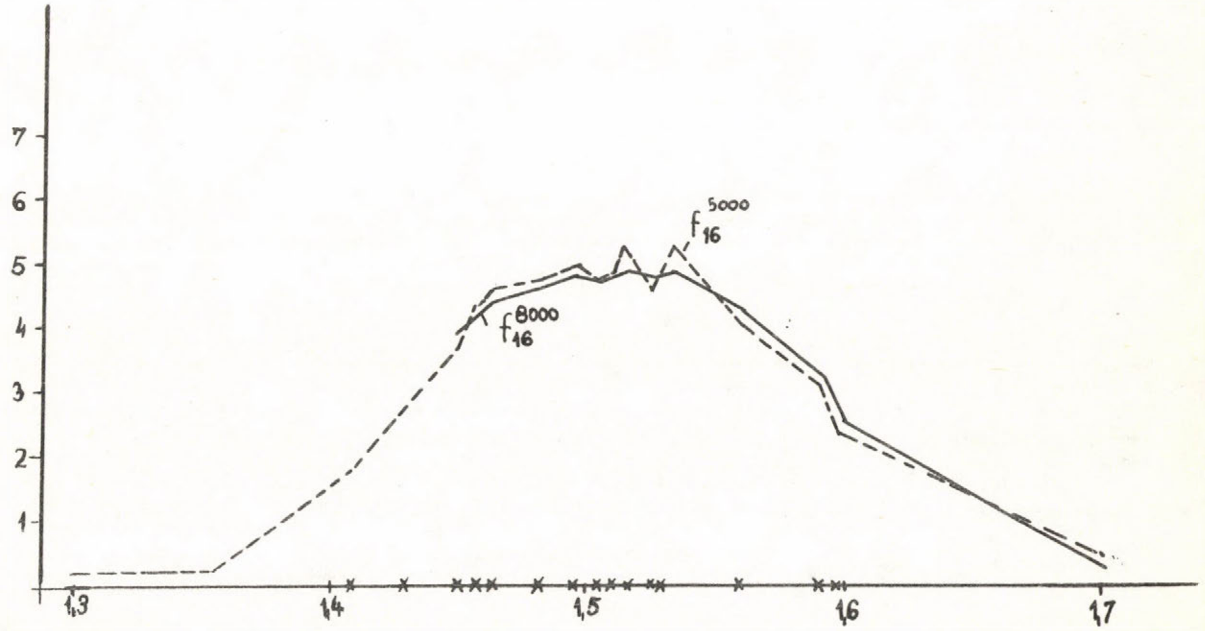
valószinüségi változók sürüségfüggvényeinek becslése volt, ahol a ϱ_1 ill. ϱ_2 csonkitott normális eloszlásu valószinüségi változók. $M\varrho_1=12$ ill. $M\varrho_2=18$ várható-értékekkel és $D\varrho_1=1{,}8$ ill. $D\varrho_2=2.7$ szórással, és a csonkitás a $[11,4,\ 12,6]$ illetve a $[17.1,\ 18.9]$ intervallumokon történt,

β normális eloszlásu valószinüségi változó, $M\beta=60$, $D\beta=45$, és a csonkitás $[44,\ 320]$ intervallumon történt,

ζ az $[5.4,\ 6.6]$ intervallumban egyenletes eloszlásu valószinüségi változó.

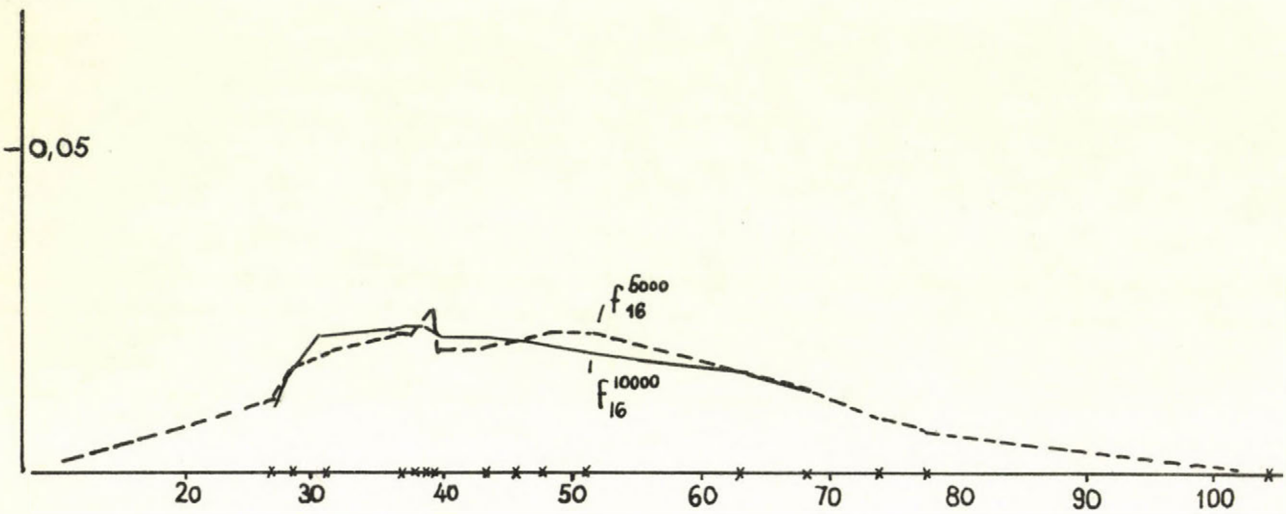
Az 1 ill. 2 ábra az $f_n^m(\xi)$ illetve az $f_n^m(\eta)$ függvények két becslését ábrázolja.

A 3 ill. 4 ábrán ugyanezen tapasztalati sürüségfüggvények szerepelnek, de $D\varrho_1 = 0.4$ és $D\varrho_2 = 0.6$ szórás mellett.



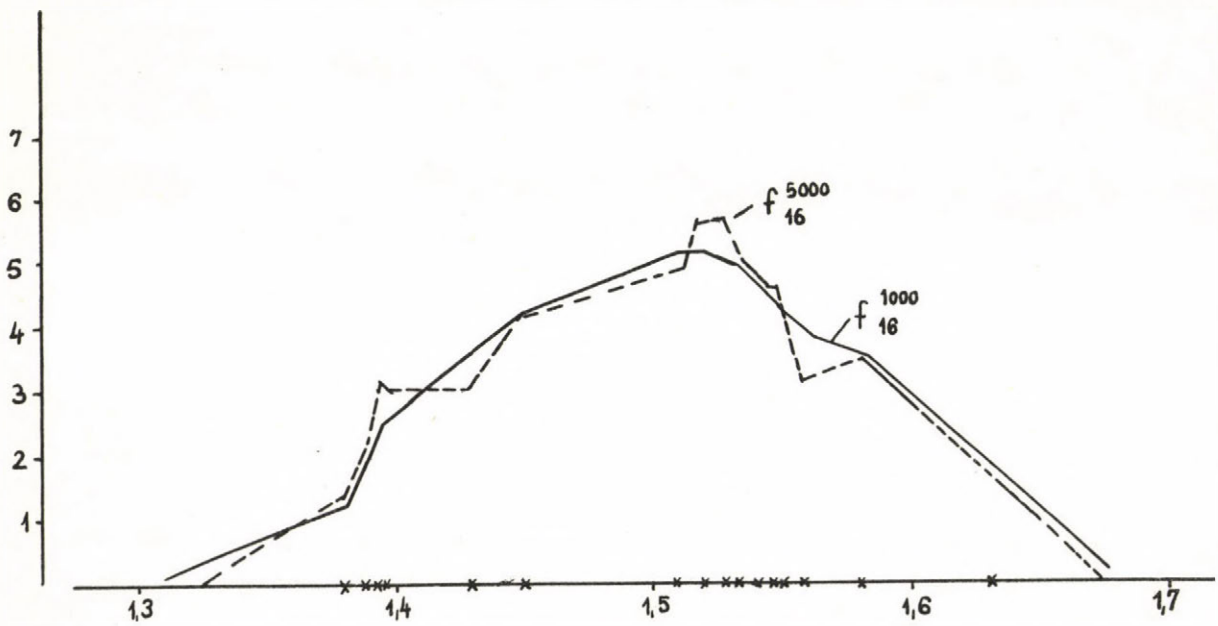
1. ábra: a ξ sürüségfüggvényének becslése

$n = 16$, $D\varrho_1 = 1,8$, $D\varrho_2 = 2.7$



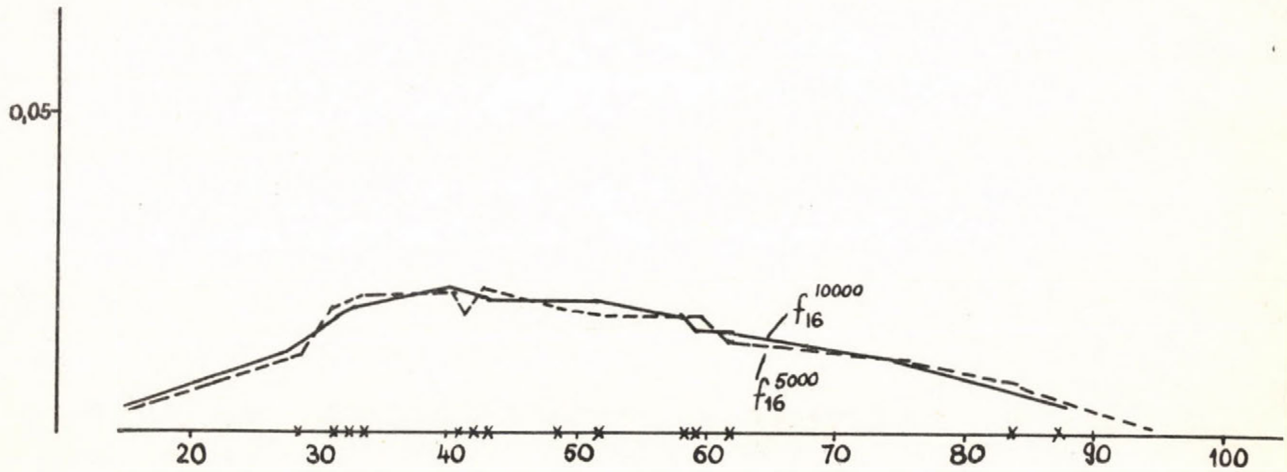
2. ábra: η sürüségfüggvényének becslése

$n = 16$, $D\varrho_1 = 1.8$, $D\varrho_2 = 2.7$



3. ábra: ξ sürüségfüggvényének becslése

$n = 16$, $D\varrho_1 = 0.4$, $D\varrho_2 = 0.6$



4. ábra: η sürüségfüggvényének becslése

n = 16, $D\,\varrho_1 = 0.4$, $D\,\varrho_2 = 0.6$

Irodalom:

[1] Dávid Gábor: A sürüségfüggvény becsléséről /MTA Számitástechnikai Központja Közlemények 3. 163-181. oldal/

[2] Sanov I.N.: /1957/ On the probability of large deviations /in Russian/Math. Sbornik 42 /89/ 11-44.

[3] Sethuraman: /1964/ On the probability of large deviations of sample mean. Annals of Mathematical Statistics. 1964. Vol 35 /pp. 1304-1316/

S u m m a r y

The main results of these paper for empirical density and distribution functions defined in author's paper [1] are the followings:
Theorem 2. Let $[a,b]$ be such an interval, in which $f(x)$ and $F^{-1}(x)$ in $[F(a), F(b)]$ are continuous, $f(x)$ has a first continuous derivate $f'(x)$, for which $0 \neq |f'(x)| < C_1$, $0 < f[x] < C$ then

$$\overline{\lim_{n}}\,\overline{\lim_{m}}\ \frac{C}{8C_1}\left(1-e^{-\frac{1}{n}(2+\alpha)\log n}\right)^{-1}\sup_{x\in[a,b]}\left|f_n^m(x)-f(x)\right| \leq 1$$

whith probability one.

A similar theorem holds for the empirical distribution function defined by the formula $F_n^m(x)=\int_a^x f_n^m(t)\,dt$ /3. Tétel/

In the applications of the Monte-Carlo-method it is costumary that the empirical distribution function is defined by the following manner: Let a_i^n be such points of the interval $[a,b]$ for which $a_i^n < a_{i+1}^n$, $i=0,1,\dots,n-1$, and $a_{i+1}^n - a_i^n \to 0$ if $n\to\infty$ for all values of i, and let us denote by m_i the number of sample, which are elemets of $(a_i^n, a_{i+1}^n]$, and

$$G_m(x)=\frac{\sum_{i=1}^{j(x)} m_i}{m} \qquad if \quad x\in\left(a_{j(x)}^n,\ a_{j(x)+1}^n\right]$$

where m is the number of sample.

If $0 < F(a) < F[b] < 1$ then for arbitrary $\alpha > 0$

$$P\left(\overline{\lim_{m}}\sqrt{\frac{m}{2(2+\alpha)\log m}}\ \sup_{x\in[a,b]}\left|G_m(x)-F(x)\right| \leq 1\right)=1$$

Р е з ю м е

Главные результаты этой статьи для эмпирической функции плотности и распределения, определённой автором в статье / I /, - это следующие:

Т е о р е м а 2.

Если функции $f(x)$ и $F^{-1}(x)$ непрерывные соответственно на отрезке $[a,b]$ и $[F(a),F(b)]$, и $f(x)$ обладает первой непрерывной производной $f'(x)$, для которой

$$0 \neq |f'(x)| < c_1, \quad 0 < f(x) < c;$$

тогда:
$$\overline{\lim_{n}}\ \overline{\lim_{m}}\ \frac{C}{8c_1}\left(1-e^{-\frac{1}{n}(2+\alpha)\log n}\right)^{-1} \sup_{x\in[a,b]} |f_n^m(x) - f(x)| \leq 1$$
с вероятностью I.

Подобная теорема имеет место для эмпирической функции распределения определённой формулой:

$$F_n^m(x) = \int_a^x f_n^m(t)\,dt \qquad / \text{ Теорема 3 } /$$

В методе статистических испытаний обычно используется эмпирическая функция распределения, определённая следующим образом:

Пусть a_i^n точки отрезка / а, в / для которых $a_i^n < a_{i+1}^n$ $i=0,1,\dots,n-1$ и $a_{i+1}^n - a_i^n \longrightarrow 0$, если $n \longrightarrow \infty$ для всех значений i . Обозначим через m_i число выборочных значений, попадающих в отрезок $(a_i^n \ a_{i+1}^n]$ и пусть эмпирическая функция распределения:

$$G_m(x) = \frac{\sum_{i=1}^{j(x)} m_i}{m} \qquad \text{если} \quad x \in (a_{j(x)}^n,\ a_{j(x)+1}^n)$$

если $0 < F(a) < F(b) < 1$, тогда для произвольного $\alpha > 0$

$$P\left(\lim_{m} \sqrt{\frac{m}{2(2+\ \ )\log m}} \sup_{x\in[a,b]} |G_m(x) - F(x)| \leq 1\right) = 1$$