# Az exponenciális valószínűség-eloszlás illesztése az empirikus gyakoriság-görbékhez

*KARDEVÁN PÉTER*

*A dolgozat azzal a kérdéssel foglalkozik, hogyan lehet az empirikus gyakorisági értékeket, melyeket egy folytonos valószínűségi változó $\tau$ hibával megmért értékeinek alapján határozunk meg, az elméleti valószínűség-értékekkel közelíteni abban az esetben, amikor a valószínűségi változó exponenciális eloszlást követ. Ekkor egy* $\exp\left(\alpha\dfrac{\tau}{2}\right)$ *alakú korrekciós tag lép fel, mely általában nem hanyagolható el.*

*Статья говорит о том, как можно с определенной на основе измеренных данных ошибкой эмпирическую относительную частоту случайной величины вероятностей показательного расспределения приблизить с теоретическами величинами. В этом случае существует* $\exp(\alpha\tau/2)$ *поправка, которой нельзя забросить.*

*The problem is dealt with, how the empirical frequency values — determined basing on the values of a continuous probability variable measured with an error* $\pm\tau/2$ *— can be approximated by theoretical probability values in the case, when the probability variable follows an exponential distribution. In that case a correction term of the form* $\exp\left(\alpha\dfrac{\tau}{2}\right)$ *appears which can not be neglected in general.*

## *1. Bevezetés*

A dolgozatban tárgyalt probléma a városi talajnyugtalanság-mérések feldolgozásakor kapott empirikus gyakoriság-görbéknek elméleti valószínűség-eloszlással való közelítése során merült fel. A $t$ valószínűségi változó e konkrét problémánál a talajelmozdulás időbeli függvényén található egymást követő lokális szélső értékek bekövetkezési időkülönbségeit jelenti. A tapasztalat szerint exponenciális eloszlással közelíthető e valószínűségi változó megmért értékeiből meghatározott gyakoriság-eloszlás.

*1.* A fenti módon értelmezett valószínűségi változó megmért értékeit $t_k^*$-gal jelöltük, s a mérés pontossága $\pm\tau/2$ sec.

A dolgozatban megtartottuk a $t$ valószínűségi változó konkrét fizikai jelentését, hogy a későbbiek során a kapott eredmények közvetlenül alkalmazhatók legyenek a talajnyugtalanság további statisztikai vizsgálatánál.

A közölt eredmények, ill. észrevételek azonban akkor is érvényesek, ha $t$ fizikai jelentése más.

*2.* Az exponenciális eloszlás eredendően folytonos eloszlás, mert a valószínűségi változó — mely jelen esetben a lokális szélsőértékek alapján megállapított időkülönbség (periódus) — szóba jöhető értékeinek halmaza a pozitív valós számokból álló halmaz. A valóságban természetesen nem fordulnak elő tetszőlegesen nagy periódusok.

Legyen $t$ a periódus, folytonos valószínűségi változó, melyről feltesszük, hogy exponenciális eloszlású, pontosabban:

$$P(t < T) = e^{-\alpha T}, \tag{1}$$

ahol $\alpha$ az átlagperiódus reciproka és $T$ egy tetszőlegesen rögzített periódusérték.

Végezzünk $N$ számú mérést $t$ értékére vonatkozólag, s legyen $T_m$ az a legnagyobb periódus, melyre teljesül, hogy

$$N(1-e^{-\alpha T}) \ll 1.$$

Ha $N$ elég nagy, akkor a $t \in \mathfrak{T}$ folytonos valószínűségi mezőnek a $\mathfrak{T} = (0, T_m)$ intervallumot tekinthetjük, azaz annak valószínűsége, hogy $t > T_m$ periódusérték előfordul $N$ mérés során, nagy biztonsággal zérónak tekinthető. Itt jegyezzük meg, hogy ha $t$ az (1) eloszlást követi, akkor $N$ mérés során várhatóan az a periódusérték lesz a legnagyobb, melyre

$$N(1-e^{-\alpha T^*_m}) \approx 1.$$

A statisztikus ingadozás következtében természetesen különböző $N$ darabból álló méréssorozatok esetén $T_m$-tól különbözhetnek az előforduló legnagyobb periódusok.

Az exponenciális eloszlás folytonos jellegéből következik, hogy annak valószínűsége, hogy a $t$ valószínűségi változó értéke éppen $T_0$ – bár nem lehetetlen esemény – mégis zéró. Véges valószínűsége annak van, hogy a $t$ valószínűségi változó valamely $(T_k, T_{k+1})$, $k = 1, 2 \ldots, n$ intervallumba esik.

Az eredetileg folytonos $\mathfrak{T}$ valószínűségi mezőt a $T_k$ osztópontokkal $n$ számú intervallumra bonthatjuk és ily módon diszkrét valószínűségi mezővé alakíthatjuk, ha a mező elemeit úgy értelmezzük, hogy a $t$ valószínűségi változó értéke valamelyik $(T_k, T_{k+1})$, $k = 1, 2 \ldots n$ intervallumba esik. Jelöljük ezeket az eseményeket $\mathfrak{T}_k$-val. A $\mathfrak{T}_k$, $k = 1, 2 \ldots n$ diszkrét mező tekinthető a $\mathfrak{T}$ mező kontrakciójának; teljesülni kell tehát, hogy a $\mathfrak{T}$ mező egy expozíciója során bekövetkezett esemény maga után vonja a $\{\mathfrak{T}_k\}$ mező valamely eseményének bekövetkezését és megfordítva.

Tehát

$$\bigcup_{k=1}^{n} \mathfrak{T}_k = \mathfrak{T}.$$

Ha méréseket végzünk, hogy a $t$ valószínűségi változó értékeit megállapítsuk és meghatározzuk az empirikus gyakorisággörbét, tulajdonképpen a fentiekben részletezett módon diszkrétté alakított mező elemeit exponáljuk minden egyes mérés alkalmával.

A mérések során kapott $t_k^*$ periódusértékeknek meg kell feleltetni a $\mathfrak{T}_k$, $k = 1, 2 \ldots n$ eseményeket, ha az empirikus gyakoriságértékeket az elméleti eloszlással össze kívánjuk hasonlítani.

Végezzünk $N$ mérést, melyek során a $t_k^*$ periódus $K$-szor fordul elő. Ekkor a $t_k^*$ periódusérték előfordulásának empirikus gyakoriságértéke $K/N$, míg a megfelelő elméleti valószínűség:

$$P(T_k < t^* < T_{k+1}) = \int_{T_k}^{T_{k+1}} \alpha e^{-\alpha t} dt, \tag{2}$$

ha minden esetben

$$t_k^* \in (T_k, T_{k+1}).$$

Az intervallum-skatulyarendszer $T_k$ osztópontjait csak abban az esetben választhatnánk meg tetszőlegesen, ha a $t_k$ periódusértékek $\tau$ kimérési hibája sokkal kisebb a $(T_{k+1}-T_k) = \Delta T$, $k = 1, 2 \ldots n$ intervallum-szélességnél. Nem szabad figyelmen kívül hagynunk azonban, hogy a kimérési hibák eleve meghatároznak egy $\{(T_k, T_{k+1})\}$ intervallum-rendszert. Ha a periódusok kimérési hibája $\tau$, akkor:

$$t_k = k \cdot \tau, \quad (k = 1, 2, \ldots n), \tag{3}$$

tehát a $t_k$ értékek diszkrét értéksorozatot alkotnak. Tegyük fel, hogy a kimérés során olyan hibákat követünk csak el, hogy ha a valódi $t$ periódusérték egy $(T_k, T_{k+1})$ intervallumba esik, akkor a helyette kimért $t_k^*$ periódusérték is eleme a $(T_k, T_{k+1})$ intervallumnak. Ha feltesszük tehát, hogy a kimérés során lényegileg csak kerekítési hibákat követünk el, akkor feltételezhetjük, hogy valahányszor $t_k^*$ értéket mérünk ki, a tényleges periódusértékek a $\left(t_k^* - \frac{\tau}{2}, t_k^* + \frac{\tau}{2}\right)$ intervallumba esnek.

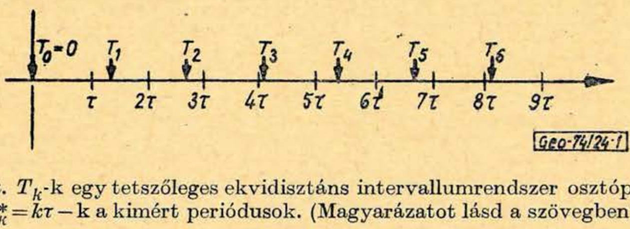
(3)-t felhasználva a fentiekben tárgyalt kimérési hibák a

$$(T_k, T_{k+1}) = \left((2k-1)\frac{\tau}{2}, (2k+1)\frac{\tau}{2}\right), \tag{4}$$

$$k = 1, 2, \ldots\ldots n$$

intervallumokat határozzák meg.

Anélkül, hogy a részletekbe belemennénk, az *1. ábrán* illusztráljuk, hogy a gyakoriságértékek milyen torzításához vezet, ha az ekvidisztáns intervallumrendszer $T_k$ osztópontjait tetszőlegesen választjuk és $T_{k+1}-T_k = \Delta T$ $\tau$-val összemérhető nagyságú. Az 1. ábrán a $(T_2, T_3)$, ill. $(T_5, T_6)$ intervallumokba eső periódusok gyakoriságértékei pusztán amiatt is nagyobbak lesznek, mert ezekbe nem csak egy, hanem két $t_k^* = k\tau$ periódusérték is beleesik.




*1. ábra.* $T_k$-k egy tetszőleges ekvidisztáns intervallumrendszer osztópontjai, $t_k^* = k\tau$ – k a kimért periódusok. (Magyarázatot lásd a szövegben)

*Рис. 1.* $T_k$ – точки деления в любой системе эквидистантных интервалов; $t_k = k\tau$ – замеренные периоды. (Объяснение см. в тексте)

*Fig. 1.* $T_k$-s denote the division points of an equidistant interval-systems, $t_k^* = \tau k$ are the periodes measured

Az empirikus gyakoriságértékkel tehát (2), (3) és (4) alapján a

$$P\left((2k-1)\frac{\tau}{2} < t < (2k+1)\frac{\tau}{2}\right) = \int\limits_{(2k-1)\frac{\tau}{1}}^{(2k+1)\frac{\tau}{2}} \alpha e^{-\alpha t} dt = 2e^{-\alpha k \tau} sh\left(\alpha \frac{\tau}{2}\right) \tag{5}$$

elméleti valószínűség-értékeket kell összehasonlítani.

3. A fenti (5) képlet azonban csak akkor érvényes szigorúan, ha olyan kísérleti berendezéssel mérjük ki a periódusokat, mely tetszőlegesen kis periódusértékeket is ki tud mérni. Célunk azonban olyan gyakoriságértékek közelítése elméleti valószínűségértékekkel, melyeket $\tau$ hibával kimért periódusértékek alapján határoztunk meg.

Tekintsük a (4) intervallumrendszer legelső intervallumát, melyet $k=1$ helyettesítéssel kapunk:

$$(T_1, T_2) = \left(\tau/2, \frac{3\tau}{2}\right).$$

(2) szerint a legkisebb kimért periódusérték $\tau$.

A tárgyalt értelmezés szerint mindazok a tényleges periódusértékek, melyek helyett a kimérési hiba folytán $\tau$ értéket mértünk ki, a $\left(\frac{\tau}{2}, \frac{3\tau}{2}\right)$ intervallumba esnek.

$\frac{\tau}{2}$-nél kisebb periódusokat egyáltalán nem tudunk kimérni. Így annak valószínűsége, hogy $\frac{\tau}{2}$-nél kisebb periódust mérünk ki, zéró.

Számítsuk ki annak valószínűség-sűrűségét, hogy a $t$ valószínűségi változó értéke, melyről feltételezzük, hogy (1) eloszlású, éppen $T$ legyen, feltéve, hogy csak $\frac{\tau}{2}$-nél nagyobb értékű lehet. Ha tetszőlegesen kis periódusokat is ki tudnánk mérni, akkor annak valószínűség-sűrűsége, hogy a $t$ értéke éppen $T$ legyen:

$$\varphi(T) = \alpha e^{-\alpha T}.$$

Annak a valószínűség-sűrűsége, hogy egy periódus értéke éppen $T$ legyen, feltéve, hogy $T > \frac{\tau}{2}$, a következő feltételes valószínűség:

$$P\left(t = T/t > \frac{\tau}{2}\right) = \frac{p(t = T)}{P\left(t > \frac{\tau}{2}\right)} = \frac{\varphi(T)}{1 - \int\limits_0^{\tau/2} \varphi(t)dt} =$$

$$= \frac{\alpha e^{-\alpha T}}{1 - \int\limits_0^{\tau/2} \alpha e^{-\alpha t} dt} = \alpha e^{-\alpha(T-\tau/2)} = \varphi(T) e^{\alpha\tau/2}.$$

A $\pm\tau/2$ pontossággal kimért periódusértékekhez tehát a

$$\varphi(T, \tau/2) = \begin{cases} 0 & \text{ha } T \leq \tau/2 \\ \alpha e^{-\alpha(T-\tau/2)} & \text{ha } T > \tau/2 \end{cases}$$

valószínűség-sűrűséget rendelhetjük.

Ekkor:

$$P_k\left(t_k^* - \frac{\tau}{2} < t \leq t_k^* + \frac{\tau}{2}\right) = \int\limits_{t^*_k-\tau/2}^{t^*_k+\tau/2} \alpha\, e^{-\alpha(t-\tau/2)}\, dt = 2e^{-\alpha t^*_k} sh\left(\alpha \frac{\tau}{2}\right) e^{\alpha\tau/2}, \tag{6}$$

ahol $t_k^*$-k a kimért periódusértékek a (2) képlettel állíthatók elő. Ha a $t_k^*$ periódusértékek kimérési hibáját figyelembe vesszük, a (6) képletet alkalmazhatjuk, mely (5)-től csak az exp $(\alpha\tau/2)$ korrekciós taggal különbözik. Minél kisebb $\tau/2$, exp $(\alpha\tau/2)$ annál inkább *1*-hez közeledik. Egy bizonyos kimérési pontosság esetén tehát (5) és (6) képletek nem különböznek lényegesen egymástól.

$\tau/2$ *sec*-ban megadott értéke függ a digitalizálási távolság *mm*-ben megadott értékétől, és a papírtovábbítási sebességtől. Minél kisebb az előbbi és minél nagyobb az utóbbi, annál kisebb hibát követünk el, ha (6) helyett az (5) képletet alkalmazzuk.

A talajnyugtalanság-regisztrátumok kiértékelésekor például a digitalizálási távolság *0,25 mm*, a papírtovábbítás sebessége $\approx$ *30 mm/sec* volt. Ebben az esetben $\alpha$ értékétől függően exp $(\alpha\tau/2) = 1{,}07 \div 1{,}08$ nem elhanyagolható korrekciót jelent az elméletileg várható darabszámok kiszámításánál.

*4.* A $\Delta T$ skatulyaszélességet úgy kell választanunk, hogy egyrészt elég keskeny legyen ahhoz, hogy részletes képet kapjunk egy viszonylag szélesebb intervallumba eső periódusok eloszlásáról, másrészt ne legyen túl keskeny sem, azaz minden intervallumba elég sok periódusérték essen. Ha a kimérési pontosság nagy, akkor az utóbbi feltétel nem feltétlenül teljesül, ha a mérések $N$ száma korlátozott. Így célszerű szélesebb intervallumokat választani. Minthogy azonban megfontolásaink szerint az intervallumrendszert a figyelembe vett kimérési hibák (4) szerint határozzák meg, a skatulya szélessége is csak

$$\Delta T = m\tau \qquad \textit{lehet}\ (m = 1, 2 \ldots).$$

(2) alapján

$$t_{k+m}^* = t_k^* + m\,\tau\,.$$

Ha a skatulya szélességét $\tau$-ról $m\tau$-ra növeljük, akkor az $l$-dik $m\tau$ szélességű intervallumba eső periódusok előfordulásának valószínűsége:

$$P_l\left(t_k^* - \tau/2 < t \leq t_{k+(m-1)}^* + \frac{\tau}{2}\right) = \int\limits_{t^*_k-\tau/2}^{t^*_{k+(m-1)}+\tau/2} \alpha\, e^{-\alpha(t-\tau/2)}\, dt =$$

$$= e^{-\alpha t_k^*} e^{\alpha\tau/2} \left(e^{\alpha\tau/2} - e^{-\alpha(2m-1)\tau/2}\right) = e^{-\alpha t'_k} e^{\alpha\tau/2} \left(e^{m\alpha\tau/2} - e^{-m\alpha\tau/2}\right) =$$

$$= e^{-\alpha t'_l} e^{\alpha\tau/2} \cdot 2sh\left(m\,\alpha\frac{\tau}{2}\right),$$

ahol $t'_l = t_k^* + (m-1)\tau/2$ az $l$-dik intervallum felezőpontjához tartozó periódusérték, $ml = k$.

*5.* Végezetül megemlítjük még, hogy a talajnyugtalanság-regisztrátumok feldolgozása során a talajelmozdulás időbeli függvényén az egymást követő lokális maximum- és minimumhelyek bekövetkezési időkülönbségeit mértük ki elsődlegesen. Ezek a $\xi_k$ számértékek tekinthetők a $\xi$ folytonos valószínűségi változó megmért értékeinek, melyek a tapasztalat szerint exponenciális eloszlást követnek. Ezeket az időközöket célszerű a szokásos periódus-fogalomnak megfelelően félperiódusoknak tekinteni. Ily módon a $t$ periódus, mely szintén valószínűségi változó, a $\xi$ folytonos valószínűségi változó függvénye. A függvénykapcsolatot a

$$t = 2\xi, \quad \text{ill.} \quad t_k^* = 2\xi_k \tag{7}$$

egyenletek határozzák meg.

Megmutatjuk, hogy a $\xi$ valószínűségi változóra érvényes (6) valószínűség ugyanazt az értéket szolgáltatja, mint a $t$ valószínűségi változóra.

Tegyük fel tehát, hogy a $\xi_k$ értékekre vonatkozólag a (6) képlet érvényes. Ekkor $\alpha$ a $\xi_k$ értékek átlagának reciproka [1]. Ha (7) alapján bevezetjük a $t$ valószínűségi változót, akkor új $\alpha$' paramétert kell bevezetnünk az

$$\alpha' = \frac{\alpha}{2} \tag{8}$$

egyenletnek megfelelően, továbbá

$$\tau' = 2\tau. \tag{9}$$

A (6) képletben azonban $\alpha\xi_k$, ill. $\alpha\tau/2$ szorzatok fordulnak elő, s (7), (8), ill. (9) alapján:

$$\alpha\,\xi_k = 2\alpha' \cdot \frac{t_k^*}{2} = \alpha' t_k^*, \tag{10}$$

$$\alpha\,\tau/2 = 2\alpha' \cdot \frac{\tau'}{2} \cdot \frac{1}{2} = \alpha' \frac{\tau'}{2}. \tag{11}$$

(10) és (11) egyenlőségek állításunkat bizonyítják. A fenti tétel $t = v\xi$ alakú transzformáció esetén is igaz, ahol $v$ tetszőleges pozitív valós szám. Ha tehát az elméleti darabszámokat a $t_k^*$ értékek alapján számítjuk ki, a (6) képletet alkalmazhatjuk, $\alpha$ jelentése ekkor az átlagperiódus reciproka.

## IRODALOM

[1] *Kardeván, P.:* 1970. Az exponenciális eloszlás alkalmazása az altalaj rezonancia-frekvenciáinak megkeresésére. Magyar Geofizika XII. évf. 2–3. sz.