

VALÓSZÍNŰÉGSZÁMÍTÁSI MÓDSZEREK ALKALMAZÁSA BIZONYOS METEOROPATOLÓGIAI JELENSÉGEK VIZSGÁLATÁNÁL

TAKÁCS LAJOS

Bevezetés

H. v. Schelling [1] munkájában, amely 1941-ben jelent meg, eljárást közöl annak megállapítására, hogy egy bizonyos eseménysorozat (biológiai jelenségek) úgy tekinthető-e, mint amely egy másik eseménysorozat (klimatikai jelenségek) hatására iött létre. *Schelling* módszert közöl annak megállapítására, hogy a hatás jelentős (szignifikáns) vagy jelentéktelen-e. *Schelling* eme módszerét hazai orvoskutatók, *dr Kérdő Imre* [2], [3] és *dr Takácsy László* [4] is átvették kísérleti adataik feldolgozására. Ezen munkáikban *Kérdő* és *Takácsy* azzal a kérdéssel foglalkoznak, hogy meteorológiai frontok előfordulásai hatással vannak-e bizonyos biológiai elváltozásokra. *Schelling* módszerét a szakemberek általában általános érvényűnek tekintik, bár *Schelling* nem foglalkozott a megengedhető alkalmazási terület vizsgálatával. Ezek a tények szükségessé tették annak megvizsgálását, hogy *Schelling* módszere milyen körülmények között alkalmazható, illetve adott körülmények között milyen módszert kell követni a kérdéses hatás kimutatására. Ezen célokat szem előtt tartva *dr Juvancz Iréneusz* kezdeményezett egy megbeszélést, amelyen *dr Kérdő Imre* és *dr Takácsy László* felvetették a megoldandó problémákat, és megadták a szükséges tájékoztatást, továbbá *Takács Lajos* a Meteorológiai Intézet részéről adott értékes felvilágosításokat a meteorológiai frontok törvényszerűségére vonatkozóan. E helyen is köszönetet mondunk mindannyiuknak a problémák felvetéséért és értékes felvilágosításaikért.

A most következő tárgyalás *Schelling* matematikai módszerének kritikájával foglalkozik, és a sztochasztikus folyamatok elméletére támaszkodva a kérdésnek az eddiginél mélyebb analizisét adja meg. Ennek eredményeképpen egy olyan kiértékelési eljárást adunk meg, amely a jelenségek széles körére alkalmazható.

Külön is köszönetet mondunk *dr Juvancz Iréneusz*nak, akivel jelen dolgozat elkészülte során a felmerülő kérdéseket megvitattuk, és aki értékes tanácsaival segítette a munka elkészültét. Továbbá köszönettel tartozunk *dr Takácsy Lászlónak*, aki [5] munkájában közölt kísérleti adatait rendelkezésünkre bocsátotta az elmélet illusztrálásához. Végül köszönet illeti *Fischer Jánost* a numerikus számítások elvégzésének irányításáért, valamint *Koós Erzsébetet* és *Major Imrénét* a számítások elvégzéséért.

1. §. A probléma kitűzése

Jelen értekezésünkben a problémát tisztán matematikai szempontból tárgyaljuk, és annak megállapításával kívánunk foglalkozni, hogy két eseménysorozat, mondjuk $\{t_n\}$ és $\{u_n\}$ előfordulási pontjai között fennáll-e valamilyen kapcsolat vagy sem, azaz az egyik sorozat eseményei hatást gyakorolnak-e a másik sorozat eseményeinek bekövetkezési időpontjaira, vagy az utóbbiak az előzőktől függetleneknek tekinthetők-e. A problémát természetesen nem ebben az általános alakjában tárgyaljuk, hanem speciális feltevéseket teszünk a $\{t_n\}$ és $\{u_n\}$ sorozat törvényszerűségére. Mégpedig feltesszük, hogy: a) $\{t_n\}$ Poisson-folyamatot alkot, b) $\{t_n\}$ rekurrens folyamatot alkot és $\{u_n\}$ mindkét esetben Poisson-folyamatot alkot. Ezután megállapítjuk függetlenséget feltételezve a két sorozatnak milyen viszonyban kell állnia egymással, és megnézzük, hogy a valóságos helyzet ettől jelentősen vagy jelentéktelenül tér-e el. Az előbbi esetben hatás feltételezésével élhetünk.

Kérdő és Takácsy példáiban a $\{t_n\}$ sorozat a meteorológiai frontok előfordulási pontjait jelöli az $\{u_n\}$ sorozat pedig a biológiai hatások bekövetkezésének időpontjait.

Meggondolásainkban a $\{t_n\}$ és $\{u_n\}$ sorozat adatait pontosaknak fogjuk tekinteni, és nem foglalkozunk azzal a kérdéssel, hogy ezek egyáltalában megállapíthatók-e, illetve milyen hibák lépnek fel ezek meghatározásánál.¹⁾

2. §. H. v. Schelling módszeréről

Schelling eljárása a következő: tekinti az időtengelyt, annak origójául rendre a $\{t_n\}$ időpontokat választja, és minden egyes t_n -hez viszonyítva ezen az időtengelyen feltünteti az $\{u_n\}$ időpontok helyzetét. Így az időtengelyen egy pontosságú nyel. Ha ezen pontok az origó körül csoportosulnak, akkor beszélhetünk hatásról, különben nem. Schelling az időtengelyt α nagyságú szakaszokra osztja (Schellingnél $\alpha = 1$ nap, Kérdőnél és Takácsynál $\alpha = 8$ óra), mégpedig úgy, hogy az origó egy szakasz közepére esik, és ezen középső szakasztól jobbra és balra ugyanannyi szakaszt tekint. Legyen az ilyen módon tekintett szakaszok száma r . (r páratlan, mondjuk $r = 2m + 1$.) Jelölje v_i^* ($i = 0, \pm 1, \pm 2, \dots, \pm m$) az egyes szakaszokra eső pontok számát. A számozás v_{-m}^* -tól v_m^* -ig történik sorrendben haladva. Schelling most a maximális és minimális v_i^* különbségét veszi: $\eta_r^* = \max_{i,j} (v_i^* - v_j^*)$, ($i, j = 0, \pm 1, \dots, \pm m$) és elosztja $\sqrt{2N_r/r}$ -rel, ahol $N_r = \sum_{i=-m}^m v_i^*$ az összes tekintetbe vett észlelések száma. Schelling a hatást jelentősnek minősíti, ha $\eta_r^* \sqrt{2N_r/r} > f_r$, egyébként pedig jelentéktelennek. Schelling f_r értékeire a következő táblázatot közli:

¹⁾ Megjegyezzük, hogy a Magyar Tudományos Akadémia IV. Osztályának Meteorológiai Főbizottsága 1955. szeptember 2-i ülésén felvette munkatervébe a meteoropatológiai kérdések tanulmányozásához szükséges módszerek kidolgozását, és e munkával dr. Aujezky Lászlót, dr. Juwancz Iréneusz és dr. Kérdő Imrét bízta meg.

r	f_r	r	f_r	r	f_r	r	f_r
2	3,00	7	3,51	12	3,67	17	3,76
3	3,20	8	3,55	13	3,69	18	3,78
4	3,32	9	3,59	14	3,71	19	3,79
5	3,40	10	3,62	15	3,73	20	3,80
6	3,46	11	3,64	16	3,75	21	3,82

Schelling nem közli azt, hogy miként jutott a fenti eredményre. Az a tény, hogy nem követeli meg, hogy v_0^* legyen a maximális v_i^* , arra utal, hogy megengedi, hogy a hatás állandó időtartammal késsen, illetve siessen a kiváltó okhoz képest. Továbbá valószínűleg feltételezi, hogy függetlenség esetén az egyes szakaszokba eső észlelések számai, a v_i értékek, egymástól független, egyforma várható értékű Poisson-eloszlású változók. Ekkor v_i várható értékének becslése: $\mathbf{M}\{v_i\} \cong N_r/r$, és mivel két tetszőleges v_i és v_j érték ($i \neq j$) különbségének a szórásnégyzete, $\mathbf{D}^2\{v_i - v_j\} = 2\mathbf{D}^2\{v_i\} = 2\mathbf{M}\{v_i\}$, tehát $\mathbf{D}\{v_i - v_j\} \cong \sqrt{2N_r/r}$. *Schelling* ehhez viszonyítja a tényleg észlelt η_r^* különbséget, és ha ez a hányados f_r -nél nagyobb, akkor a hatást jelentősnek (szignifikánsnak) minősíti.

Schelling megfontolásának részleteit nem ismerjük, de látszik, hogy a tényleg észlelt η_r^* értéket a függetlenség hipotézisével számított $\mathbf{D}\{v_i - v_j\}$ értékkel kívánja összehasonlítani. Ez az eljárás helyes és szokásos. Hiba azonban az, hogy $\mathbf{D}\{v_i - v_j\}$ becslésére olyan mennyiségeket (N_r) is felhasznál, amelyeknek értéke függ a hatás milyenségétől, és így nem tekinthető a függetlenség esetére vonatkozó érték becslésének. Továbbá nem tudjuk, hogy miképpen nyerte *Schelling* az f_r értékeket, de látszik, hogy rögzített valószínűségi szint mellett f_r -nek gyorsabban kellene növekednie r -rel. Végül *Schelling*-nél nem történik utalás α és r helyes megválasztására. Továbbá módszeréből nem derül ki alkalmazásának érvényességi köre, ami pedig annak veszélyét rejti magában, hogy olyankor is alkalmazzák, amikor nem szabad.

Most következő tárgyalásunkban igyekszünk a fent felsorolt hiányosságokat kiküszöbölni és *Schelling* közelítő eljárásánál általánosabban alkalmazható módszert megadni.

3. §. A probléma megoldása

A következőkben feltesszük, hogy az $\{u_n\}$ időpontokat a $\{t_n\}$ időpontok környezetében *Schelling* diagrammjához hasonlóan ábrázoljuk, és megállapítjuk az egyes α nagyságú közökbe eső v_i ($i = 0, \pm 1, \pm 2, \dots, \pm m$) mennyiségeket. Mégpedig vizsgálatunk vonatkozzék egy T hosszúságú időszakaszra, azaz csak azon $\{u_n\}$ időpontokat tekintjük, amelyek beleesnek ebbe a szakaszba. Legyen ismét a tekintetbe vett szakaszok száma $r = 2m + 1$. A következőkben bizonyos feltevésekkel élünk a $\{t_n\}$ és $\{u_n\}$ sorozatokra. Feltesszük, hogy a két sorozat egymástól független, és így megállapítjuk, hogy a $\{v_i\}$ változók eloszlása milyen törvényszerűségeket követ. Ezután megnezzük, hogy a $\{v_i^*\}$ konkrét észlelési adatok jelentős eltérést mutatnak-e a függetlenség feltételezésével indokolható értékektől. Ha az eltérés jelentős, úgy hatásról beszélünk, különben pedig nem.

A meteorológiai frontok törvényszerűségére vonatkozóan a következő feltevésekkel élünk :

a) Bármely időpontban a korábban előforduló frontok semilyen hatást sem gyakorolnak a jövőben keletkező frontokra, és a frontok folyamata bármely időponttól számítva ugyanazon sztochasztikus törvénynek van alávetve.

Matematikailag ez a feltevés ekvivalens azzal, hogy feltesszük, hogy a frontok folyamata λ eseménysűrűségű Poisson-folyamat. Ekkor a frontok közötti távolságok átlaga, $\tau = 1/\lambda$ és szórása $\sigma = 1/\lambda$.

b) Az egyes frontok létrejötte után a megelőző frontok semilyen hatást sem gyakorolnak a jövőben keletkező frontokra, és a frontok folyamata minden egyes front előfordulásától számítva ugyanazon sztochasztikus törvénynek van alávetve.

Matematikailag ez a feltevés a következővel ekvivalens : A $\{t_n\}$ sorozatban a szomszédos t_n időpontok közötti távolságok egyforma eloszlású független valószínűségi változók közös $F(x)$ eloszlásfüggvényével. Feltesszük, hogy $F(x)$ nem rácso eloszlás, átlaga, τ , és szórása, σ , végesek.

A biológiai hatások előfordulási pontjairól, az $\{u_n\}$ sorozatról feltesszük, hogy μ (állandó) eseménysűrűségű Poisson-folyamat eseményeinek előfordulási pontjaival egyeznek meg.

A meteorológiai frontok törvényszerűségére tett a), illetve még inkább a b) feltevés gyakorlatilag elég jól megállja a helyét. A fentieknél általánosabb feltevések már igen nagy bonyodalmakra vezetnének a matematikai tárgyalásban. Ami a biológiai hatások folyamatát illeti, ha a $\{t_n\}$ sorozattól való függetlenséget tételezzük fel, úgy csakis Poisson-folyamat feltevésének van létjogosultsága.

Feltesszük végül, hogy mindezen folyamatok végtelen hosszú ideje tartanak már. Ekkor stacionárius állapot valósul meg, és vizsgálatainkban stacionárius folyamat T hosszúságú időintervallumára szorítkozunk.

A következőkben a $\{v_i\}$ ($i = 0, \pm 1, \pm 2, \dots, \pm m$) változók eloszlásának a tanulmányozásával foglalkozunk, majd ismertetjük a kiértékelés módszerét. Először csak az eredményeket közöljük, a bizonyításokat pedig később adjuk meg. Tárgyalásunkat két részre bontjuk aszerint, amint a $\{t_n\}$ sorozat a) Poisson-folyamatot, b) rekurrens folyamatot alkot.

a) *Poisson-folyamat esete.* Tegyük fel, hogy a meteorológiai frontok $\{t_n\}$ sorozata λ eseménysűrűségű Poisson-folyamatot, a biológiai megfigyelések pedig az előzőtől független μ eseménysűrűségű Poisson-folyamatot alkot. Ebben az esetben a Schelling-féle diagramm egyes oszlopaiban levő észlelések számai, a $\{v_i\}$ valószínűségi változók egyforma eloszlásúak. Közös várható értékük

$$(1) \quad \mathbf{M} \{v_i\} = \lambda \mu \alpha T,$$

és szórásnégyzetük

$$(2) \quad \mathbf{D}^2 \{v_i\} = \lambda \mu \alpha T (1 + \lambda \alpha + \mu \alpha) + \dots,$$

T -nél kisebb nagyságrendű tagoktól eltekintve.

A v_i és v_j változók ($i \neq j$) korrelációs koefficiensének határértéke $T \rightarrow \infty$ esetben

$$(3) \quad \rho(v_i, v_j) = \frac{(\lambda + \mu) \alpha}{1 + (\lambda + \mu) \alpha}.$$

Ezen formulákban a T és α mennyiségek előre rögzített értékek, a λ és μ ismeretleneket pedig a vizsgált észlelések alapján kell meghatározni. Ha a T hosszúságú időszak alatt N_1 meteorológiai front fordult elő és N_2 biológiai megfigyelés, úgy a következő becsléssel élhetünk :

$$\lambda \cong \frac{N_1}{T} \quad \text{és} \quad \mu \cong \frac{N_2}{T}$$

Ezen becslések alkalmazásával

$$(4) \quad \mathbf{M} \{v_i\} \cong \frac{N_1 N_2}{T} \alpha,$$

$$(5) \quad \mathbf{D}^2 \{v_i\} \cong \frac{N_1 N_2}{T} \alpha \left[1 + \frac{(N_1 + N_2) \alpha}{T} \right],$$

és

$$(6) \quad \varrho(v_i, v_j) \cong \frac{(N_1 + N_2) \alpha / T}{1 + (N_1 + N_2) \alpha / T}$$

írható jó közelítéssel. Ezen becslések javára irandó, szemben *Schellingé*vel, hogy nem tartalmaznak olyan mennyiségeket, amelyek a hatás-diagrammból lettek véve.

Tulajdonképpen a $\{v_i\}$ változók együttes eloszlásának a meghatározása lenne a fő cél, ami azonban igen nehéz feladat. Ezért közelítő megoldáshoz folyamodunk, és a $\{v_i\}$ változók együttes eloszlására jobb híján normális eloszlást tételezünk fel a megadott paraméterekkel. Mint ismeretes, a többdimenziós normális eloszlás egyértelműen meg van határozva az átlagok, szórások és a korrelációs koeficiensek ismeretében. Egyébként a többdimenziós normális eloszlás feltételezése indokolt is, ugyanis a $\{v_i\}$ változóknak a T időtartamtól való függését vizsgálva, jó közelítéssel független növekményű sztochasztikus folyamatot nyerünk, ahonnan következik, hogy a $\{v_i\}$ változók együttes eloszlása közelítőleg aszimptotikusan normális. A következőkben tehát tegyük fel, hogy a $\{v_i\}$ változók együttes eloszlása normális a (4), (5) és (6) által meghatározott paraméterekkel. A $\varrho(v_i, v_j)$ korrelációs koeficiienst jelölje egyszerűen ϱ , és legyen $\mathbf{D} \{v_i\} = D$.

A függetlenség feltevésének eldöntésére azt kívánjuk megállapítani, hogy az észlelt $\{v_i^*\}$ értékek tekinthetők-e a fenti törvénynek eleget tevő $\{v_i\}$ változók megfigyelt értékeinek. Ugyanis minél kisebb ennek a valószínűsége, annál biztosabban következtethetünk hatás létezésére. Indokoltnak látszik az $\eta_r = \max_{i,j} (v_i - v_j)$ változó vizsgálata, ugyanis ez a mennyiség érzékenyen tükrözi a hatást. Ha hatás mutatkozik, akkor bizonyos v_i értékek megnőnek, míg mások lecsökkennek. Az η_r változó eloszlásfüggvénye explicite nem határozható meg, de táblázatozva megtalálható abban az esetben, ha $\varrho = 0$. (Lásd például [6].)

I. táblázat

r	$P_\lambda = 0,50$	$P_\lambda = 0,60$	$P_\lambda = 0,70$	$P_\lambda = 0,80$	$P_\lambda = 0,90$	$P_\lambda = 0,95$	$P_\lambda = 0,99$	$P_\lambda = 0,999$	$P_\lambda = 0,9995$
2	0,95	1,20	1,47	1,81	2,33	2,77	3,64	4,65	4,92
3	1,59	1,83	2,09	2,42	2,90	3,31	4,12	5,06	5,31
4	1,98	2,21	2,47	2,78	3,24	3,63	4,40	5,31	5,56
5	2,26	2,48	2,73	3,04	3,48	3,86	4,60	5,48	5,72
6	2,47	2,69	2,94	3,23	3,66	4,03	4,76	5,62	5,86
7	2,65	2,86	3,10	3,39	3,81	4,17	4,88	5,73	5,96
8	2,79	3,00	3,24	3,52	3,93	4,29	4,99	5,82	6,04
9	2,92	3,12	3,35	3,63	4,04	4,39	5,08	5,90	6,12
10	3,02	3,23	3,46	3,73	4,13	4,47	5,16	5,97	6,19
11	3,12	3,32	3,55	3,82	4,21	4,55	5,23	6,04	6,25
12	3,21	3,41	3,63	3,90	4,29	4,62	5,29	6,09	6,31
13	3,29	3,48	3,70	3,97	4,35	4,69	5,35	6,14	6,36
14	3,36	3,55	3,77	4,03	4,41	4,74	5,40	6,19	6,40
15	3,42	3,62	3,83	4,09	4,47	4,80	5,45	6,23	6,45
16	3,48	3,67	3,89	4,14	4,52	4,85	5,49	6,28	6,49
17	3,54	3,73	3,94	4,19	4,57	4,89	5,54	6,32	6,52
18	3,59	3,78	3,99	4,24	4,61	4,93	5,57	6,35	6,56
19	3,64	3,83	4,03	4,29	4,65	4,97	5,61	6,38	6,59
20	3,69	3,87	4,08	4,33	4,69	5,01	5,65	6,41	6,62

Az I. táblázat a

$$(7) \quad P\{\eta_r \geq \lambda D\} = P_\lambda$$

valószínűség összetartozó λ és P_λ értékeit tünteti fel különböző r -ekre.

Arra az esetre, ha a ρ korrelációs koefficiens zérustól különbözik, nincsenek táblázatok, de szerencsére az előbbi táblázat összetartozó λ és P_λ értékei felhasználhatók η_r eloszlásának meghatározására ebben az általánosabb esetben is, ugyanis $\rho \neq 0$ -ra fennáll, hogy

$$(8) \quad P\left\{\eta_r \geq \lambda D \frac{\sqrt{\gamma^2 - 2\gamma(r-1)\rho + (r-1)[1 + (r-2)\rho]}}{\gamma + 1}\right\} = P_\lambda,$$

ahol

$$(9) \quad \gamma = \frac{1 + (r-2)\rho + \sqrt{1 + (r-2)\rho - (r-1)\rho^2}}{\rho}.$$

Amint látjuk, $\rho = 0$ és $\rho \neq 0$ esetben egyaránt meghatározható η_r eloszlása. Ezen eloszlás ismeretében pedig eldönthetjük, hogy a kísérletileg talált $\eta_i^* = \max_{i,j}(\nu_i^* - \nu_j^*)$ érték tekinthető-e a fenti feltevéssel nyert η_r változó megfigyelt értékének. A döntés a hipotézis-vizsgálat szokásos módszere szerint hajtható végre.

Az elmondottak alapján a következő vizsgálati módszert követhetjük: először megválasztjuk r és α értékeit. Mint említettük, $r = 2m + 1$. r megválasztásával kapcsolatban az a helyzet, hogy minél nagyobbra választjuk, annál szigorúbbnak mondható eljárásunk, ugyanis nagy r érték esetén elég nagy η_r^* értékeket sem tekinthetünk jelentősen kiugrónak. Nem célszerű azonban r -et úgy megválasztani, hogy $m\alpha > \tau$ legyen, ugyanis ekkor már távolabbi frontok hatáskörébe esnek az észlelések. Ezért javasoljuk az $m \approx \tau/\alpha$

választást. Az α érték megválasztására a következő szempontok mérvadók. Ha kiértékelésre a (7) képletet akarjuk felhasználni, amely a legegyszerűbb, de amely csak $\varrho = 0$ esetén érvényes, úgy α -t úgy kell megválasztani, hogy ϱ értéke lehetőleg kicsiny legyen. Mivel ϱ α -val együtt nő, ezért α -t elegendő kicsinyre kell választani. A (8) alatti képlet alkalmazásánál ϱ miatt nem kell kikötést tenni α -ra, de α -t elég kicsire kell választani τ -hoz képest, és pedig annál kisebbre, minél szigorúbbnak vesszük a megengedett hatótávolságot. Miután r -et és α -t megválasztottuk, az (5) és (6) képlet alapján meghatározzuk D és τ értékét. A Schelling-féle diagrammból leolvassuk az $\eta_r^* = \max_i v_i^* - \min_i v_i^*$ különbséget és képezzük az

$$(10) \quad \frac{\eta_r^*}{D}$$

illetve

$$(11) \quad \frac{\eta_r^*}{D} \frac{\gamma + 1}{\sqrt{\gamma^2 - 2\gamma(r-1)\varrho + (r-1)[1 + (r-2)\varrho]}}$$

hányadost, aszerint, amint ϱ elhanyagolható, vagy sem. A P_λ -táblázat r -edik sorában megkeressük a fenti értéknek megfelelő λ -t, és a hozzátartozó P_λ szolgáltatja azt a valószínűségi szintet, amellyel a hatás feltételezhető.

A (7) és (8) képletek nyeréséhez a következő gondolatmenet vezet. Ismeretes, hogy korrelálatlan normális együttes eloszlású változók egyúttal függetlenek is. Így, ha $\varrho = 0$, úgy a $\{v_i\}$ változók ugyanazon normális eloszlású független valószínűségi változók, azaz egy r - elemű minta elemei. Az η_r változó jelöli a minta terjedelmét, amelynek eloszlása normális alapsokaság esetén ismeretes. Ha $\varrho \neq 0$, akkor a $\{v_i\}$ változók egyforma normális eloszlású változók ugyan, de nem függetlenek. Ha azonban a

$$\xi_i = (\gamma + 1)v_i - \sum_{j=-m}^m v_j \quad i = 0, \pm 1, \dots, \pm m$$

új változókat vezetjük be, ahol γ -t (9) definiálja, akkor a ξ_i változók is egyforma normális eloszlású változók lesznek, de már korrelálatlanok és következőleg függetlenek is lesznek. Így az új változókra (7) alapján fennáll, hogy

$$P \left\{ \frac{\max(\xi_i - \xi_j)}{D(\xi_i)} \geq \lambda \right\} = P_\lambda.$$

Ha most tekintetbe vesszük, hogy

$$\frac{\max(\xi_i - \xi_j)}{D(\xi_i)} = \frac{\gamma + 1}{\sqrt{\gamma^2 - 2\gamma(r-1)\varrho + (r-1)[1 + (r-2)\varrho]}} \frac{\max(v_i - v_j)}{D(v_i)}$$

úgy nyerjük a (8) képletet.

b) *Rekurrens folyamat esete.* Tegyük fel most, hogy a meteorológiai frontok $\{t_n\}$ sorozatában a $t_n - t_{n-1}$ időkülönbségek egyforma eloszlású, független pozitív valószínűségi változók, ugyanazon nem-rácsos $F(x)$ eloszlásfüggvénnyel. Legyen $F(x)$ átlaga τ és szórása σ . A biológiai megfigyelésekről pedig tegyük fel, hogy az előzőtől függetlenül μ eseménysűrűségű Poisson-folyamatot alkotnak. Jelölje most is $\{v_i\}$ a Schelling-diagrammban szereplő

értékeket. A $\{v_i\}$ valószínűségi változók most is egyforma eloszlásúak. Közös várható értékük

$$(12) \quad \mathbf{M}\{v_i\} = \frac{\mu\alpha T}{\tau},$$

és szórásnégyzetük

$$(13) \quad \mathbf{D}^2\{v_i\} = \frac{\mu\alpha T}{\tau} \left[1 + \frac{\mu\alpha\sigma^2}{\tau^2} + \frac{2}{\alpha} \int_0^{\alpha} m(t) dt \right] + \dots,$$

eltekintve T -nél kisebb nagyságrendű tagoktól. Itt $m(t)$ jelenti egy meteorológiai fronttól számított t időtartamon belül előforduló újabb frontok várható számát. Ennek értéke

$$m(t) = \sum_{n=1}^{\infty} F_n(t),$$

ahol $F_n(t)$ jelöli az $F(t)$ eloszlásfüggvénynek önmagával való n -szeres kompozícióját.

A v_i és v_j változók ($i < j$) korrelációs koefficiense $T \rightarrow \infty$ esetben:

$$(14) \quad \varrho(v_i, v_j) = \frac{\frac{\alpha\mu\sigma^2}{\tau^2} + \frac{1}{\alpha} \int_{(i-i)\alpha}^{(j-i+1)\alpha} m(t) dt - m((j-i)\alpha)}{1 + \frac{\alpha\mu\sigma^2}{\tau^2} + \frac{2}{\alpha} \int_0^{\alpha} m(t) dt}.$$

Ezen formulákban rögzített T és α értékek mellett a τ, σ, μ paraméterektől és az $m(t)$ átlagfüggvényt kell megbecsülni. A meteorológiai frontok analíziséhez megadja τ, σ és $m(t)$ értékeit, és $\mu \cong N_2/T$, ha T idő alatt N_2 biológiai megfigyelés történt. A fenti képleteknek ily módon becsült értékei is azzal sa előnnyel rendelkeznek, hogy nem tartalmaznak olyan mennyiségeket, amelyeket a hatás-diagramm alapján kellene számítani. A Poisson-folyamat tárgyalásánál említett normális eloszlással való közelítés most is alkalmazható, és így az előbbi tárgyalás szinte szószó szerint megismételhető. Ha α -t olyan kicsinyre választjuk, hogy $\varrho(v_i, v_j)$ elhanyagolható, úgy az

$$\frac{\eta_r^*}{\mathbf{D}\{v_i\}}$$

mennyiség kiszámításával a (7) formula alapján dönthetünk a hatás létezéséről. Ha a $\varrho(v_i, v_j)$ mennyiségek nem hanyagolhatóak el, akkor a (7) formulára való visszavezetés bonyolult transzformációk alkalmazását kívánja meg, mivel ekkor $\varrho(v_i, v_j)$ nem állandó, hanem függ a $j - i$ különbségtől. A legtöbb esetben azonban $\varrho(v_i, v_j)$ közel állandó, és ekkor egy átlagos ϱ bevezetésével a (8) kiértékelés is alkalmazható.

4. §. Az eredmények bizonyítása

Az előző fejezetben tárgyalt *a*) eset a *b*) eset speciális esetét képezi, mégpedig, ha *b*)-ben $m(t) = \lambda t$, $\tau = 1/\lambda$, $\sigma = 1/\lambda$, úgy nyerjük az *a*) esetet. Ezért külön *a*)-val nem foglalkozunk.

Az előrebozsított eredmények bizonyításához felhasználunk néhány segédtelet, amelyet [7] dolgozatunkban bizonyítottunk be.

Tegyük fel, hogy $\{t_n\}$ ($n = 0, \pm 1, \pm 2, \dots$) időpontokban előfordul egy-egy esemény, ahol a $t_n - t_{n-1}$ ($n = 0, \pm 1, \pm 2, \dots$) időkülönbségek egyforma eloszlású független, pozitív valószínűségi változók ugyanazon $F(x)$ nem-rácsos eloszlásfüggvénnyel. Tegyük fel, hogy $F(x)$ átlaga, τ , és szórása, σ , végesek.

Feltesszük, hogy a $\{t_n\}$ sorozat minden egyes eseményéhez hozzá van rendelve egy $f(t - t_n)$ amplitúdójú jel, ahol $f(u)$ csak véges hosszúságú intervallumban különbözik zérustól. Feltesszük továbbá, hogy az egyes jelek lineárisan szuperponálódnak (összeadódnak). Ilymódon az

$$\eta_t = \sum_{(n)} f(t - t_n)$$

sztochasztikus folyamatot nyerjük, amely 1 valószínűséggel létezik, ha $\int_{-\infty}^{\infty} |f(t)| dt < \infty$. [7] dolgozatunkban megmutattuk, hogy η_t egy stacionárius sztochasztikus folyamatot alkot. η_t várható értéke

$$(15) \quad \mathbf{M} \{ \eta_t \} = \frac{1}{\tau} \int_{-\infty}^{\infty} f(t) dt,$$

és második momentuma

$$(16) \quad \mathbf{M} \{ \eta_t^2 \} = \frac{1}{\tau} \left[\int_{-\infty}^{\infty} [f(t)]^2 dt + 2 \int_0^{\infty} f_{11}(t) dm(t) \right],$$

ahol

$$f_{11}(t) = \int_{-\infty}^{\infty} f(u) f(u + t) du$$

és

$$m(t) = \sum_{h=1}^{\infty} F_n(t),$$

ahol $F_n(t)$ jelöli az $F(t)$ függvénynek önmagával való n -szeres kompozícióját. Itt $m(t)$ jelenti egy $\{t_n\}$ sorozatbeli eseménytől számított t időtartam alatt előforduló újabb események várható számát.

Megjegyezzük, hogy ha az $f(u)$ függvény argumentumát lineárisan eltoljuk, úgy ezáltal az η_t folyamat sztochasztikus viselkedése nem változik meg; ez kifejezésre jut a fenti képletekben is. Ezen észrevétel alapján megengedhető, hogy az egyes jelek kezdőpontja a hozzájuk tartozó esemény idő-

pontjába toljuk, azaz, úgy tekinthető a helyzet, mintha az egyes események tényleg elindították volna a jeleket, és így az eseményeknek csak utóhatásuk van.

Ha speciálisan $\{t_n\}$ λ eseménysűrűségű Poisson-folyamatot alkot, amikor is $m(t) = \lambda t$ és $\tau = 1/\lambda$, úgy

$$\mathbf{M}\{\eta_i\} = \lambda \int_{-\infty}^{\infty} f(t) dt$$

és

$$\mathbf{D}^2\{\eta_i\} = \lambda \int_{-\infty}^{\infty} [f(t)]^2 dt.$$

A fenti segédtelemek segítségével kívánjuk a v_i változók várható értékét, szórásnégyzetét és korrelációs koefficiensét meghatározni.

Tekintsük a $\{t_n\}$ sorozatot, $\{u_n\}$ pedig legyen egy, a $\{t_n\}$ időpontoktól független, μ eseménysűrűségű Poisson-folyamat, amelyet $(0, T)$ intervallumban tekintünk. Készítsük el Schelling diagrammját, és jelölje v_i ($i = 0, \pm 1, \dots, \pm m$) az egyes α hosszúságú közőkbe eső észlelések számát.

Célunk a v_i változó tanulmányozása. Észrevevesszük, hogy v_i a következőképpen is megkapható. Tegyük fel, hogy minden egyes $\{t_n\}$ sorozatbeli esemény elindít egy α ideig tartó történést, és pedig a történést t_n -től $(i - 1/2)\alpha$ távolságra van, és a $(0, T)$ intervallumban minden egyes $\{u_n\}$ sorozatbeli eseményt annyszor számolunk, ahány történést éppen folyamatban van. Így megkapjuk v_i értékét. Az elmondottak viszont tovább egyszerűsíthetők, ugyanis egyszerűen tekinthetünk olyan történéseket, amelyek a t_n időpontokban kezdődnek és α ideig tartanak, és ekkor v_i -vel azonos eloszlást mutató valószínűségi változót nyerünk. A történéseket pedig definiálhatjuk

$$f(u) = \begin{cases} 1, & \text{ha } 0 \leq u \leq \alpha \\ 0 & \text{egyébként} \end{cases}$$

időbeli lefolyású jelekkel, és azt mondhatjuk, hogy t időpontban E_k állapot van, ha $\eta_t = k$. Ha az $\{u_n\}$ Poisson-folyamatban $(0, T)$ időközben előforduló eseményeket annyszorosan számítjuk, amennyi abban a pillanatban η_t értéke, úgy az események összege v_i -vel ekvivalens változót szolgáltat.

Jelölje most a $\{t_n\}$ sorozatban $(0, T)$ időközben előforduló események számát ξ_t változó, az E_0, E_1, E_2, \dots állapotokban való tartózkodási időket pedig $\tau_0, \tau_1, \tau_2, \dots$ változók.

Ekkor nyilvánvalóan fennáll

$$(17) \quad \tau_0 + \tau_1 + \tau_2 + \dots = T$$

és

$$(18) \quad \tau_1 + 2\tau_2 + 3\tau_3 + \dots = \alpha\xi_T + \varepsilon_T.$$

Az utóbbi egyenlőség, jobboldalán csak $\alpha\xi_T$ -vel, nem pontos a $(0, T)$ intervallum szélein levő speciális helyzet folytán, ezért kell az ε_T hibtagot is szerepeltetni. Erre mindenesetre fennáll, hogy $\mathbf{M}\{\varepsilon_T\} = 0$ és $\mathbf{D}^2\{\varepsilon_T\}$ T -től független korlát alatt marad.

$\mathbf{M}\{v_i\}$ meghatározása. v_i összetevődik az $\{u_n\}$ Poisson-folyamat $(0, T)$ időközben előforduló k -szoros $(k = 1, 2, 3, \dots)$ multiplicitással számított eseményeinek számából. Ha τ_k állandó, úgy ezen feltétel mellett a Poisson-folyamatban τ_k idő alatt előforduló események várható száma $\mu\tau_k$, és mivel minden ilyen esemény k -szoros multiplicitással veendő, tehát a k -szoros multiplicitású eseményekből adódó adalék $k\mu\tau_k$. A keresett várható érték pedig

$$(19) \quad \mathbf{M}\{v_i\} = \mathbf{M}\{\mu\tau_1 + 2\mu\tau_2 + 3\mu\tau_3 + \dots\} = \mu T \mathbf{M}\{\eta_t\} = \frac{\mu\alpha T}{\tau}.$$

Ugyanis

$$\mathbf{M}\{\tau_k\} = T \mathbf{P}\{\eta_t = k\}, \quad \sum_{k=0}^{\infty} k \mathbf{P}\{\eta_t = k\} = \mathbf{M}\{\eta_t\}$$

és (15) alapján

$$\mathbf{M}\{\eta_t\} = \frac{\alpha}{\tau}.$$

Megjegyezzük, hogy ez az eredmény (18) alapján is belátható, ha tekintetbe vesszük, hogy $\mathbf{M}\{\xi_T\} = \frac{T}{\tau}$ és $\mathbf{M}\{\varepsilon_T\} = 0$.

$\mathbf{D}^2\{v_i\}$ meghatározása. A szórásnégyzet a feltételes várható érték ismert definíciója alapján határozható meg:

$$(20) \quad \mathbf{D}^2\{v_i\} = \mathbf{M}\left\{\mu \sum_{k=0}^{\infty} k^2 \tau_k\right\} + \mathbf{D}^2\left\{\mu \sum_{k=0}^{\infty} k \tau_k\right\}.$$

Itt most (16) alapján

$$\mathbf{M}\left\{\mu \sum_{k=0}^{\infty} k^2 \tau_k\right\} = \mu T \mathbf{M}\{\eta_t^2\} = \frac{\mu T}{\tau} \left[\alpha + 2 \int_0^{\alpha} m(t) dt \right],$$

ugyanis

$$\mathbf{M}\{\tau_k\} = T \mathbf{P}\{\eta_t = k\} \quad \text{és} \quad \mathbf{M}\{\eta_t^2\} = \sum_{k=0}^{\infty} k^2 \mathbf{P}\{\eta_t = k\}.$$

Továbbá

$$\begin{aligned} \mathbf{D}^2\left\{\mu \sum_{k=0}^{\infty} k \tau_k\right\} &= \mu^2 \mathbf{D}^2\{a\xi_T + \varepsilon_T\} = \mu^2 a^2 \mathbf{D}^2\{\xi_T\} + \mu^2 \mathbf{D}^2\{\varepsilon_T\} + \\ &+ 2\mu^2 a \varrho(\xi_T, \varepsilon_T) \mathbf{D}\{\xi_T\} \mathbf{D}\{\varepsilon_T\}, \end{aligned}$$

ahol $\varrho(\xi_T, \varepsilon_T)$ a korrelációs együttható

Most, mint ismeretes

$$\mathbf{D}^2\{\xi_T\} = \frac{2}{\tau} \int_0^T \left[m(t) - \frac{t}{\tau} + \frac{1}{2} \right] du = \frac{\sigma^2}{\tau^3} T + \dots,$$

ahol jobboldalon elhagytuk a T -nél kisebb nagyságrendű tagokat. $\mathbf{D}^2\{v_i\}$ korlátos ($m(a) a^2$ - nagyságrendű). Így tehát végül

$$(21) \quad \mathbf{D}^2\{v_i\} = \frac{\mu T}{\tau} \left[a + \frac{\mu a^2 \sigma^2}{\tau^2} + 2 \int_0^a m(t) dt \right] + \dots,$$

eltekintve a T -nél kisebb nagyságrendű tagoktól.

Abban a speciális esetben, midőn $\{t_n\}$ λ eseménysűrűségű Poisson-folyamatot alkot, érvényes, hogy $\mathbf{D}^2\{v_i\} = \frac{2}{3} \lambda a^3$ és

$$\mathbf{D}^2\{v_i\} = T \lambda \mu a (1 + \lambda a + \mu a) + \dots$$

$\rho(v_i, v_j)$ meghatározása. A korrelációs koefficiens meghatározását úgy végezzük el, hogy először $v_i + v_j$ várható értékét és szórásnégyzetét határozzuk meg, és abból következtetünk $\rho(v_i, v_j)$ -re. $v_i + v_j$ meghatározására pontosan azt az utat követjük, mint v_i meghatározásánál tettük, csak az $f(u)$ függvényt definiáljuk másképpen, mégpedig most feltesszük, hogy

$$f(u) = \begin{cases} 1, & \text{ha } 0 \leq u \leq a, (j-i)a \leq u \leq (j-i+1)a, \\ 0 & \text{egyébként.} \end{cases}$$

Az E_k állapotot most is $\eta_k = k$ definiálja, és a $(0, T)$ időközben E_k állapotban töltött időt most is τ_k -val jelöljük. Ekkor (18)-hoz hasonlóan felírható, hogy

$$\tau_1 + 2\tau_2 + 3\tau_3 + \dots = 2a\xi_T + \varepsilon'_T$$

ahol most $\mathbf{M}\{\varepsilon'_T\} = 0$ és $\mathbf{D}\{\varepsilon'_T\}$ T -ről független korlát alatt marad.

Ekkor (19)-hez hasonlóan azt nyerjük, hogy

$$\mathbf{M}\{v_i + v_j\} = \frac{2\mu a T}{\tau},$$

és (20) szerint

$$\mathbf{D}^2\{v_i + v_j\} = \mathbf{M}\left\{\mu \sum_{k=0}^{\infty} k^2 \tau_k\right\} + \mathbf{D}^2\left\{\mu \sum_{k=0}^{\infty} k \tau_k\right\},$$

ahol most

$$\begin{aligned} \mathbf{M}\left\{\mu \sum_{k=0}^{\infty} k^2 \tau_k\right\} &= \mu T \mathbf{M}\{\eta_i^2\} = \\ &= \frac{2\mu a T}{\tau} + \frac{2\mu T}{\tau} \left[2 \int_0^a m(t) dt + \int_{(j-i)a}^{(j-i+1)a} m(t) dt - a m[(j-i)a] \right], \\ \mathbf{D}^2\left\{\mu \sum_{k=0}^{\infty} k \tau_k\right\} &= \frac{4\alpha^2 \mu^2 \sigma^2}{\tau^3} T + \dots \end{aligned}$$

ahol elhagytuk a T -nél kisebb nagyságrendű tagokat.

Ha tekintetbe vesszük, hogy

$$\mathbf{D}^2\{v_i + v_j\} = \mathbf{D}^2\{v_i\} + 2\mathbf{D}\{v_i\} \mathbf{D}\{v_j\} \rho(v_i, v_j) + \mathbf{D}^2\{v_j\},$$

és hogy $\mathbf{D}\{v_i\} = \mathbf{D}\{v_j\}$, úgy azt nyerjük, hogy

$$\varrho(v_i, v_j) = \frac{\mathbf{D}^2\{v_i + v_j\} - 2\mathbf{D}^2\{v_i\}}{2\mathbf{D}^2\{v_i\}}$$

A korrelációs koefficienszt mindjárt $T \rightarrow \infty$ esetre felírva azt kapjuk, hogy

$$(22) \quad \varrho(v_i, v_j) = \frac{\frac{\alpha\mu\sigma^2}{\tau^2} + \frac{1}{\alpha} \int_{(j-i)\alpha}^{(j-i+1)\alpha} m(t) dt - m[(j-i)\alpha]}{1 + \frac{\alpha\mu\sigma^2}{\tau^2} + \frac{2}{\alpha} \int_0^{\alpha} m(t) dt}$$

Ha $\{t_n\}$ λ eseményűrűségű Poisson-folyamatot alkot, úgy speciálisan

$$\varrho(v_i, v_j) = \frac{(\lambda + \mu)\alpha}{1 + (\lambda + \mu)\alpha}$$

5. §. Példa

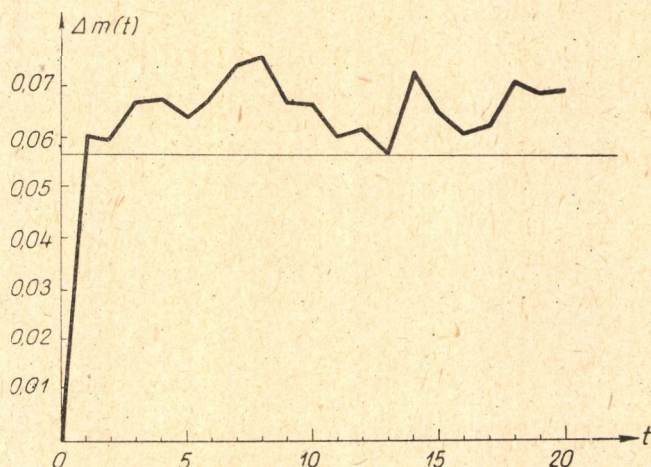
A következő illusztráló példát *Takácsy L.* [5] értekezésének adatai alapján számítottuk. *Takácsy* adatai 4 évi megfigyelésen alapulnak, amelyek az 1948. január 1-től 1951. december 31-ig terjedő időre vonatkoznak. $\{t_n\}$ jelöli a meteorológiai frontok előfordulási pontjait, $\{u_n\}$ pedig a biológiai jelenségek (hirtelen szívhalálozások) időpontjait.

Ennél a példánál $T = 4$ év = 1461 nap = 35064 óra. A Meteorológiai Intézet által az *Időjárás* című folyóiratban közzétett adatok szerint ez idő alatt $N_1 = 1979$ meteorológiai front fordult elő. A frontok átlagos távolsága $\tau = 17,64$ óra volt ($\tau \cong \frac{T}{N_1}$ becsléssel $\tau = 17,72$ órának adódik), és szórása

$\sigma = 21,09$ óra. Az $m(t)$ átlagfüggvény mellett a $\Delta m(t) = m(t) - m(t-1)$ különbséget is feltüntetjük, ahol t órában értendő. Ezekre a 4 évi adatok alapján a következő adódott (lásd: 1. ábra és 2. táblázat):

2. táblázat

t óra	$m(t)$	$\Delta m(t)$	t óra	$m(t)$	$\Delta m(t)$
1	0,0602	0,0602	11	0,7385	0,0602
2	0,1194	0,0592	12	0,8002	0,0617
3	0,1861	0,0667	13	0,8569	0,0567
4	0,2534	0,0673	14	0,9297	0,0728
5	0,3273	0,0639	15	0,9949	0,0652
6	0,3950	0,0677	16	1,0556	0,0607
7	0,4689	0,0739	17	1,1179	0,0623
8	0,5443	0,0754	18	1,1887	0,0708
9	0,6110	0,0667	19	1,2575	0,0688
10	0,6783	0,0673	20	1,3268	0,0693



1. ábra

A következőkben az $m(t)$ átlagfüggvényt a tapasztalati megállapított függvénnyel helyettesítjük. Első közelítésként azonban $\Delta m(t)$ közel állandónak tekinthető, és a $\Delta m(t) \cong \frac{N_1}{T} = \frac{1979}{35064} = 0,0564$ becsléssel élhetünk. Pontosabb számításnál az $F(x)$ eloszlásfüggvényt jó közelítéssel vehetjük Γ -eloszlásnak, mikor is

$$F'(x) = \frac{e^{-\omega x} \omega^\beta x^{\beta-1}}{\Gamma(\beta)},$$

ahol $\omega = \tau/\sigma^2$ és $\beta = \tau^2/\sigma^2$, és ekkor

$$m'(t) = e^{-\omega t} \omega \sum_{h=1}^{\infty} \frac{(\omega t)^{h\beta-1}}{\Gamma(h\beta)}.$$

A megfigyelt halálozások száma ez alatt a 4 év alatt $N_2 = 1004$ volt. Így a μ eseménysűrűsége $\mu \cong N_2/T = \frac{1004}{35064} = \frac{1}{34,8}$ adódik.

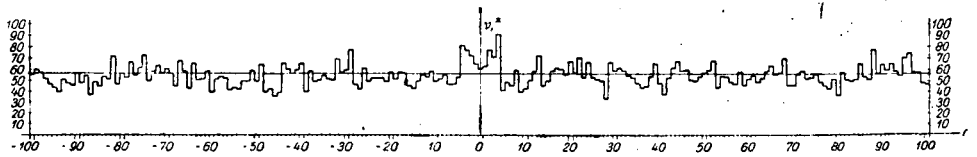
Schelling diagrammját $\alpha = 1$ órás közökre a következőkben adjuk meg (lásd: 2. ábra és 3. táblázat). Most

$$\mathbf{M}\{v_i\} \cong \frac{N_1 N_2}{T} \alpha$$

alapján az $\mathbf{M}\{v_i\} = 56,7$ becslést nyerjük az átlagra függetlenség feltételezése mellett.

3. táblázat

i	-100	-99	-98	-97	-96	-95	-94	-93	-92	-91	-90	-89	-88	-87	-86
v_i^*	56	60	57	52	47	44	40	51	48	46	57	48	55	37	49
i	-85	-84	-83	-82	-81	-80	-79	-78	-77	-76	-75	-74	-73	-72	-71
v_i^*	45	54	52	72	47	57	53	67	56	62	73	50	59	64	57
i	-70	-69	-68	-67	-66	-65	-64	-63	-62	-61	-60	-59	-58	-57	-56
v_i^*	61	57	45	68	59	43	66	51	52	59	40	52	54	53	42
i	-55	-54	-53	-52	-51	-50	-49	-48	-47	-46	-45	-44	43	-42	-41
v_i^*	44	43	50	50	59	50	65	50	53	46	49	66	61	57	60
i	-40	-39	-38	-37	-36	-35	-34	-33	-32	-31	-30	-29	-28	-27	-26
v_i^*	66	40	59	50	51	55	52	51	70	58	60	78	47	43	62
i	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11
v_i^*	50	53	53	53	50	58	52	58	57	46	43	50	56	54	59
i	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	-0	1	2	3	4
v_i^*	50	52	56	47	47	53	82	78	73	66	61	64	78	72	92
i	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
v_i^*	41	49	47	60	40	43	51	59	73	46	50	59	62	61	56
i	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
v_i^*	68	57	71	53	67	53	52	50	34	67	59	62	59	55	53
i	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
v_i^*	48	44	45	54	66	48	38	53	62	68	59	60	51	50	54
i	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
v_i^*	57	60	68	44	55	54	49	47	58	46	52	55	49	52	59
i	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
v_i^*	64	56	57	70	46	46	56	59	52	53	56	49	46	43	52
i	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
v_i^*	37	58	51	50	52	66	54	51	78	57	65	60	66	59	55
i	95	96	97	98	99	100									
v_i^*	71	75	58	58	48	47									



2. ábra

Az $\alpha = 1$ órás diagramm ismeretében az $\alpha = 3$ órás, $\alpha = 5$ órás, ... diagrammok összevonással egyszerűen elkészíthetők.

A fent elmondottak ismeretében hozzákezdhetünk a kiértékeléshez. A kiértékelésnél célszerűen $\alpha = 3$ óra és $m = 6$, azaz $r = 2m + 1 = 13$ választással élhetünk. A kiértékelést először az a) alatti Poisson-féle közelítéssel, majd a b) alatti rekurrens folyamatokra vonatkozó közelítéssel végezzük. A fenti esetben $\eta_r^* = \max_i v_i^* - \min_i v_i^* (i = 0, \pm 1, \dots, \pm m)$ értéke $r = 13$ mellett $\eta_{13}^* = 106$.

$$a) \text{ Poisson-folyamat esete. Ekkor } \lambda \cong \frac{N_1}{T} = \frac{1}{17,72 \text{ óra}}$$

$$\text{és } \mu \cong \frac{N_2}{T} = \frac{1}{34,79 \text{ óra}}, \text{ és így (5) szerint}$$

$$D = \mathbf{D}\{v_i\} \cong \sqrt{\frac{N_1 N_2 \alpha}{T} \left[1 + \frac{(N_1 + N_2) \alpha}{T} \right]} = 14,25$$

és (6) szerint

$$\varrho \cong \frac{(N_1 + N_2) \alpha / T}{1 + (N_1 + N_2) \alpha / T} = 0,2032.$$

Most

$$\frac{\eta_r^*}{D} \cong 7,44$$

és

$$\frac{\eta_r^*}{D} \frac{\gamma + 1}{\sqrt{\gamma^2 - 2\gamma(r-1)\varrho + (r-1)[1 + (r-2)\varrho]}} \cong 7,44 \cdot 1,11 \cong 8,26;$$

ugyanis most (9) szerint $\gamma = 24,0$.

Mindkét esetben a $P_\lambda > 0,9995$ valószínűségi szintet nyerjük az 1. táblázat 13. sorából ($\lambda = 6,36$ mellett $P_\lambda = 0,9995$).

b) Rekurrens folyamat esete. Ekkor $\tau \cong 17,64$ óra, $\sigma \cong 21,09$ óra, $\mu \cong 1/34,79$ óra és $m(t)$ értékeit a 2. táblázat tünteti fel. Ezen értékek segítségével $\alpha = 3$ órára és $r = 13$ esetre számítva

$$D = \mathbf{D}\{v_i\} \cong \sqrt{\frac{\mu \alpha T}{\tau} \left[1 + \frac{\mu \alpha \sigma^2}{\tau^2} + \frac{2}{\alpha} \int_0^\alpha m(t) dt \right]} \cong 14,83;$$

és a $\rho(v_i, v_j)$ korrelációs koefficiensekre az adódik, hogy

$j-i$	1	2	3	4	5
$\rho(v_i, v_j)$	0,1925	0,1978	0,1884	0,1836	0,1827

Most

$$\frac{\eta_r^*}{D} \cong 7,15$$

és (11) szerint $\rho \cong 0,189$ középértékkel számolva

$$\frac{\eta_r^*}{D} \frac{\gamma + 1}{\sqrt{\gamma^2 - 2\gamma(r-1)\rho + (r-1)[1 + (r-2)\rho]}} \cong 7,15 \cdot 1,11 \cong 7,49;$$

ugyanis most (9) szerint $\gamma = 25,3$.

Most is mindkét esetben a hatás feltételezésének valószínűségére a $P_\lambda > 0,9995$ valószínűségi szintet nyerjük az 1. táblázat $r = 13$ -adik sorából.

Amint látjuk, az egyes kiértékelési módok nem adnak lényegesen különböző eredményeket. Ha a meteorológiai frontokra érvényes, hogy $\sigma \sim \tau$ és $\Delta m(t)$ közel állandó, úgy számolhatunk az a) alatti Poisson-féle közelítéssel. Ha ez nem áll fenn, akkor a b) alatti, rekurrens folyamatokra vonatkozó közelítéssel kell számolni. Az η_r^*/D szorzótényezőjének figyelembevétele mindkét esetben nem lényeges, de némileg finomítja az eredményt.

Schelling módszerénél ebben az esetben $\eta_r^* = 106$, $N_r = 2242$, és

$$\frac{\eta_r^*}{\sqrt{2N_r/r}} \cong 5,65$$

nagyobb a megadott $f_r = 3.69$ értéknél, és így a hatás Schelling szerint is jelentős.

IRODALOM

- [1] H. v. SCHELLING: „Die Bedeutung der statistischen Methodik für die Biologie” *Ergebnisse der Hygiene, Bakteriologie, Immunitätsforschung und Experimentelle Therapie* 24 (1941) 87—149.
- [2] KÉRDŐ I.: „Újabb vizsgálatok a frontátvonulásoknak a halálózásra való hatásáról” *Orvosi Hetilap* 90 (1949) 430.
- [3] KÉRDŐ I.: „Az orvos meteorológiai vizsgálatok módszertani kérdései és új feladatai” *Időjárás* 53 (1949) 232—237.
- [4] TAKÁCSY L.: „Tapasztalatok a frontátvonulások hatásáról, három évi boncolási anyag alapján”. *Időjárás* 55 (1951) 199—209.
- [5] TAKÁCSY L.: „Hirtelen szívhalál orvos-meteorológiai vonatkozásai” *Nép-egészségügy* (sajtó alatt).
- [6] A. HALD: *Statistical tables and formulas*. Wiley, New-York, 1952.
- [7] TAKÁCSY L.: „Rekurrens folyamatok által származtatott másodlagos sztochasztikus folyamatokról.” *A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának Közleményei* 5 (1955) 187—197.

ПРИМЕНЕНИЕ ТЕОРЕТИКО-ВЕРОЯТНОСТНЫХ МЕТОДОВ
ПРИ ИССЛЕДОВАНИИ НЕКОТОРЫХ МЕТЕОРО-ПАТОЛОГИЧЕСКИХ ЯВЛЕНИЙ

Л. Такач

Резюме

Настоящая работа задается целью установить, оказывает ли один ряд событий $\{t_n\}$ (климатические явления) влияние на точки происхождения другого ряда событий $\{u_n\}$ (биологические явления). Исследование основывается на том обычном принципе что мы предполагаем, что влияния нет и установим, с какой вероятностью оправдается этим предположением некоторая вычисленная на основе данных опыта характеристическая величина. Если эта вероятность мала, то можно говорить о влиянии.

Наши рассуждения относятся к следующей модели: о ряде $\{t_n\}$ мы предполагаем, что он является или а) процессом Пуассона с плотностью событий λ или б) рекуррентным процессом, в котором разности времени $t_n - t_{n-1}$ являются независимыми положительными случайными величинами одинакового распределения с общей функцией распределения $F(x)$; распределение $F(x)$ не решетчатое, его среднее значение τ и дисперсия σ^2 конечны. Относительно ряда $\{u_n\}$ мы предполагаем в обоих случаях, что он является процессом Пуассона плотности μ . Все эти явления мы предполагаем продолжающимися с бесконечно давнего времени, т. е. мы предполагаем стационарное состояние.

Следуя теперь методу, изложенному в работе [1] *Х. фон Шеллинга*, мы изобразим расположение точек $\{u_n\}$ в окрестности точек $\{t_n\}$ следующим образом. Рассмотрим ось времени и принимая за начало координат последовательно точки t_n , расположим в их окрестности точки u_n . Пусть наши исследования относятся к промежутку времени T и разделим отрезок $\left[-\left(m + \frac{1}{2}\right)a, \left(m + \frac{1}{2}\right)a\right]$ на $r = 2m + 1$ частей длины a и подсчитаем точки в каждом интервале. Пусть будут эти величины в случае конкретных данных $v_{-m}^*, \dots, v_0^*, \dots, v_m^*$, в то время как в случае, получающемся при предположении независимости они были бы случайными величинами $v_{-m}, \dots, v_0, \dots, v_m$.

Для исследования влияния мы будем рассматривать величины $\eta_r^* = \max_i v_i^* - \min_i v_i^*$ соответственно случайные величины $\max_i v_i - \min_i v_i$.

На основе результатов, содержащихся в работе [7] мы определяем математическое ожидание, дисперсию и коэффициент корреляции переменных v_i ($i = 0, \pm 1, \dots, \pm m$). Эти величины даются в случае а) формулами (1), (2) и (3) соответственно,

а в случае б) формулами (12), (13) и (14) соответственно, где $m(t) = \sum_{n=1}^{\infty} F_n(t)$. Через $F_n(t)$

мы обозначаем n -кратную композицию функции распределения $F(t)$ с самым собой. Если в промежутке времени длины T произошло N_1 событий, принадлежащих к ряду $\{t_n\}$ и N_2 событий, принадлежащих к ряду $\{u_n\}$, то в случае а) мы можем пользоваться оценками $\lambda \cong N_1/T$ и $\mu \cong N_2/T$ и получить тем самым для (1), (2) и (3) оценки (4), (5) и (6).. В случае б) надо еще оценить σ и $m(t)$. В обоих случаях а) и б) можно положить с хорошим приближением, что величины $\{v_{-m}, \dots, v_0, \dots, v_m\}$ имеют совместное нормальное распределение. Но тогда в случае а) можно определить распределение η_r , именно для него имеет место (8), где $D = \mathbf{D}\{v_i\}$, ρ — коэффициент корреляции, а γ определяется через (9). Соответствующие друг другу значения вероятности P_λ и параметра λ даются таблицей 1. (см. А. Халд [6]). На основе сказанного в случае а) мы можем следовать такому способу: оценим значение D и ρ , из (11) определим λ и вероятность гипотезы влияния дается уровнем вероятности P_λ , соответствующим найденному значению λ , фигурирующему в r -ой строке таблицы 1. В случае б) распределение η_r уже сложнее, но с помощью введения среднего коэффициента корреляции ρ сказанное об а) можно применить в качестве приближения. На практике целесообразно выбрать $m \sim \tau/a$, а a следует выбрать тем меньшим, чем более точным предполагается значение дальности действия; иначе говоря a — допускаемая неопределенность дальности действия.

Для иллюстрации принципиальных результатов мы выбрали пример, приведенный в работе Л. Такачи [5], занимающемся воздействием метеорологических фронтов от 1. января 1948 г. до 1. декабря 1951 г. на случаи внезапной смерти от сердца.

Настоящая статья возникла так, что некоторые (И. Керде [2], [3], Л. Такачи [4]) считали, что изложенный в статье [1] метод для исследования вышеописанного явления Х. фон Шеллинга имеет общую силу. Однако, доктор Иренеус Ювану обратил внимание на то, что метод Шеллинга следует принимать с критикой и надо исследовать возможности его применения. В результате относящегося к этому разговору возникла настоящая статья, показывающая, что метод Шеллинга недостаточно основан математически, в нем скрываются источники ошибок и он не имеет общей силу. Напротив, изложенный в настоящей статье метод может быть применен к широкому кругу явлений и принимает во внимание также индивидуальные особенности исследуемого явления.

ANWENDUNG WAHRSCHEINLICHKEITSTHEORETISCHER METHODEN BEI DER UNTERSUCHUNG GEWISSER METEOROPATHOLOGISCHER ERSCHEINUNGEN

L. TAKÁCS

Zusammenfassung

Es wird das Problem untersucht, ob die Aufttrittspunkte einer Ereignissfolge $\{u_n\}$ (biologische Erscheinungen) durch eine andere Ereignissfolge $\{t_n\}$ (meteorologische Ereignisse) beeinflusst werden oder nicht? Die Untersuchung beruht auf dem üblichen Prinzip der Annahme, dass keine Beeinflussung besteht; sodann die Wahrscheinlichkeit bestimmt mit welcher der Wert einer gewissen auf Grund der Versuchsergebnisse berechneten charakteristischen Grösse durch diese Hypothese begründet werden kann. Ist diese Wahrscheinlichkeit klein, so kann von einer Beeinflussung gesprochen werden.

Die Überlegungen beziehen sich auf das folgende Modell: Von der Folge $\{t_n\}$ wird angenommen, dass sie a) ein Poissonscher Prozess von Ereignisdichte λ , bzw. b) ein rekurrenter Prozess ist, in welchem die Zeitdifferenzen $t_n - t_{n-1}$ unabhängige positive Zufallsveränderliche von gleicher Verteilung mit der gemeinsamen Verteilungsfunktion $F(x)$ sind; $F(x)$ ist keine gitterartige Verteilung und ihr Mittelwert τ und Streuungsquadrat σ^2 existieren. Von der Folge $\{u_n\}$ wird es in beiden Fällen angenommen, dass sie ein Poissonscher Prozess von Ereignisdichte μ ist. Es wird angenommen, dass beide Prozesse stationär sind.

Dem Verfahren von H. v. Schelling [1] befolgend wird die Lagerung der Punkte $\{u_n\}$ in der Umgebung der Punkte $\{t_n\}$ folgendermassen dargestellt: Man betrachte die Zeitachse; die Punkte t_n der Reihe nach als Nullpunkt wählend, werden die Zeitpunkte u_k in deren Umgebung betrachtet, d. h. die Punkte mit Koordinaten $u_k - t_n$ in Bezug auf t_n . Die Untersuchungen sollen sich auf das Zeitintervall T beziehen. Man teile das Intervall $\left[-\left(m + \frac{1}{2}\right)a, \left(m + \frac{1}{2}\right)a\right]$ der Zeitachse in $r = 2m + 1$ Strecken von der Grösse a und zähle die Anzahl der in die einzelnen Strecken fallenden Punkte ab. Im Falle konkreter Angaben seien diese Anzahlen der Reihe nach $\nu_{-m}^*, \dots, \nu_0^*, \dots, \nu_m^*$ im Falle aber, dass eine Unabhängigkeit angenommen wurde, die Zufallsveränderlichen $\nu_{-m}, \dots, \nu_0, \dots, \nu_m$.

Zur Untersuchung der Beeinflussung seien die Grösse $\eta_r^* = \max_i \nu_i^* - \min_i \nu_i^*$ bzw. die Zufallsveränderliche $\eta_r = \max_i \nu_i - \min_i \nu_i$ betrachtet.

Weiterhin werden der Mittelwert, das Streuungsquadrat und die Korrelationskoeffizienten der Veränderlichen ν_i ($i = 0, \pm 1, \dots, \pm m$) auf Grund der in der Abhandlung [7] publizierten Ergebnisse bestimmt. Die Grössen werden im Falle a) der Reihe nach durch die Formeln (1), (2) und (3), im Falle b) aber durch die Formeln (12), (13),

(14) geliefert, wobei $m(t) = \sum_{m=1}^{\infty} F_n(t)$. Durch $F_n(t)$ wird n -fache Faltung der Verteilungsfunktion $F(t)$ mit sich bezeichnet. Wenn während des Zeitintervalls der Länge T , N_1 bzw. N_2 Ereignisse, die der Folge $\{t_n\}$ bzw. $\{u_n\}$ angehören, vorkommen, dann kann man im Falle a) die Abschätzungen $\lambda \cong N_1/T$ und $\mu \cong N_2/T$ anwenden; man erhält die Abschätzungen (4), (5) und (6) für (1), (2) und (3). Im Falle b) sind auch σ

und $m(t)$ abzusätzen. Sowohl im Falle $a)$ als in $b)$ darf man mit guter Annäherung annehmen, dass die kumulative Verteilung der Veränderlichen $\{\nu_{-m}, \dots, \nu_0, \dots, \nu_m\}$ normal ist. Dann kann aber die Verteilung von η_r im Falle $a)$ bestimmt werden, und zwar gilt (8), wobei $D = D\{\nu_i\}$, ρ der Korrelationskoeffizient ist und γ durch (9) definiert ist. Die zusammengehörigen Werte der Wahrscheinlichkeit P_λ und des Parameters λ sind in der Tabelle 1. gegeben (siehe *A. Hald* [6]). Daher verfährt man im Falle $a)$ folgendermassen: man schätzt die Werte von D und ρ ab, rechnet λ (11) aus, und sucht dessen Wert in der r -ten Zeile der Tabelle 1. auf. Das zu diesem λ Wert gehörige Wahrscheinlichkeitsniveau liefert die Wahrscheinlichkeit der Hypothese einer Beeinflussung. Im Falle $b)$ ist die Verteilung von η_r bereits verwickelter; die Einführung eines mittleren Korrelationskoeffizienten ρ ermöglicht aber die angenäherte Anwendung dessen, was unter $a)$ festgelegt wurde. In der Praxis darf man zweckmässig $m \sim \tau/\alpha$ wählen und α soll desto kleiner gewählt werden, je mehr die Grösse der Reichweite als pünktlich angenommen wird, d. h. α ist die erlaubte Unsicherheit der Reichweite.

Zur Illustration der theoretischen Resultate wird auf ein Beispiel in *L. Takácsy's* hingewiesen, welcher sich auf die Frage des Einflusses der zwischen 1. I. 1948. und 1. XII. 1951. aufgetretenen meteorologischen Fronten auf das Vorkommen von plötzlichen Todesfällen bei Herzkranken bezieht.

Die für die Untersuchung der obigen Erscheinung in [1] vorgeschlagene Methode von *H. v. Schelling* wurde von mehreren Verfassern (*I. Kérdó* [2], [3], *L. Takácsy* [4]) angewandt. *I. Juvancz* machte darauf aufmerksam, dass *Schelling's* Methode mit Vorsicht angewendet werden sollte. Als Folge einer hieraufbezüglichen Besprechung ist die vorliegende Abhandlung entstanden. Sie zeigt, dass *Schelling's* Methode mathematisch ungenügend begründet ist, Fehlerquellen in sich verbirgt und keine Allgemeingültigkeit besitzt. Dagegen kann die in dieser Arbeit veröffentlichte Methode auf einen breiteren Kreis von Erscheinungen angewandt werden; sie zieht auch die individuellen Besonderheiten der untersuchten Erscheinung in Betracht.