

VALÓSZÍNŰSÉG-ELOSZLÁSFÜGGVÉNYEK KEVERÉKÉNEK FELBONTÁSA ÖSSZETEVŐIRE

MEDGYESSY PÁL

Összefoglalás

A szerző a következő problémával foglalkozik: Ha egy valószínűség-sűrűség-függvény (amely empirikusan van megadva) felfogható, mint különböző paraméterű normális sűrűségfüggvények keveréke, ahol a súlyok konstansok, hogyan lehet meghatározni a komponenseket? — A cikk hangsúlyozza a probléma fontosságát és spektroszkópiai valamint biológiai példákkal illusztrálja. Azután foglalkozik a megoldás grafikus és analitikus módszereivel. *G. Doetsch* a megoldást a keverék Fourier-transzformáltja segítségével adja meg, míg a szerző ugyanezt a keverék magasabb deriváltjaival képezett függvényt sor segítségével állítja elő. — A vizsgálatokat a szerző két példával illusztrálja, megemlíti ezen kívül saját vizsgálatait is, amelyek az általánosítás kérdéseire és a gépi megoldás lehetőségére vonatkoznak.

Kevert valószínűség-eloszlásfüggvényen a következőt értjük: Legyen $F^*(x, t)$ eloszlásfüggvényeknek a t paramétertől függő családja, $(-\infty < t < \infty)$, és tegyük fel, hogy $F^*(x, t)$ folytonos függvénye t -nek; legyen továbbá $p(t)$ egy valószínűség-sűrűségfüggvény. Könnyen belátható, hogy az

$$F(x) = \int_{-\infty}^{\infty} F^*(x, t) p(t) dt$$

függvény eloszlásfüggvény; ezt a függvényt a $F^*(x, t)$ eloszlásfüggvényeknek a $p(t)$ súlyfüggvénnyel képezett keverékének nevezzük. Ha $\frac{\partial F^*(x, t)}{\partial x} = f(x, t)$ létezik és folytonos x és t értelmezési tartományában, akkor a

$$\frac{dF}{dx} = \int_{-\infty}^{\infty} f^*(x, t) p(t) dt = f(x)$$

integrál is létezik; $f(x)$ -et, az $F(x)$ sűrűségfüggvényét, az $f^*(x, t)$ sűrűségfüggvények $p(t)$ súllyal képzett keverékének nevezzük.

Hasonlóképpen ha $t_1, t_2, \dots, t_k, \dots$ egy tetszőleges számsorozat, az $F^*(x, t_k)$, ill. $f^*(x, t_k)$ függvényeknek p_k súlyokkal képezett keverékén az

$$F_1(x) = \sum_{k=1}^{\infty} F^*(x, t_k) p_k \text{ ill. } f_1(x) = \sum_{k=1}^{\infty} f^*(x, t_k) p_k \text{ függvényt értjük,}$$

$$\left| \sum_{k=1}^{\infty} p_k = 1 \right|.$$

A definíció további kiterjesztésével nem foglalkozunk.

Eloszlásfüggvények keverékével a gyakorlatban gyakran találkozunk. Például vizsgáljuk egy ország (bizonyos korú) lakossága testmagasságának, mint valószínűségi változónak eloszlásfüggvényét. Ez a vizsgált sokaságba tartozó férfiak, ill. nők testmagasság-eloszlásfüggvényeinek keveréke, a férfiak, ill. nők számával arányos súlyokkal. Vagy például meteorológiai vizsgálatokból kiderült, hogy egy adott hely napi átlagos hőmérsékletének, mint valószínűségi változónak eloszlásfüggvénye jól közelíthető normális eloszlásfüggvények keverékével [1].

Találkozunk azonban a következő problémával is, amelyet a keverés, mint valószínűségszámítási »művelet« inverzének nevezhetünk. Az

$$F(x) = \int_{-\infty}^{\infty} F^*(x, t) p(t) dt,$$

$$F_1(x) = \sum_{k=1}^{\infty} F^*(x, t_k) p_k$$

ill.

$$f(x) = \int_{-\infty}^{\infty} f^*(x, t) p(t) dt,$$

$$f_1(x) = \sum_{k=1}^{\infty} f^*(x, t_k) p_k$$

összefüggésekből F (ill. F_1, f, f_1), és esetleg F^* (ill. f) ismeretében meg kell határoznunk a többi függvényt, melyek a képletekben szerepelnek.

A meghatározáskor alkalmazott eljárást az *eloszlás-, ill. sűrűségfüggvények keveréke összetevőkre bontásának* nevezzük.

Ebben a dolgozatban speciális sűrűségfüggvények keverékének összetevőkre bontásával fogunk foglalkozni. A kérdés gyakorlati jelentőségére alább még kitérünk.

A szóbanforgó keverék a következő:

$$f(x) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2\sigma_k^2}}}{\sqrt{2\pi} \sigma_k},$$

ahol

$$\sum_{k=1}^N A_k = 1, \quad A_k > 0, \quad 0 < m_1 < \dots < m_N, \quad \sigma_k > 0$$

konstansok. (V. ö. a fenti $f_1(x)$ -szel). $f(x)$ tehát normális (Gauss-féle) sűrűségfüggvények A_k súlyokkal vett keveréke. ($\sum_{k=1}^N A_k = 1$ megkötés különben lényegtelen; itt csak a valószínűségszámítási terminológia vonta maga után).

Analitikus fogalmazásban tehát *problémánk így hangzik:*

$$f(x) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2\sigma_k^2}}}{\sqrt{2\pi} \sigma_k}$$
 ismeretében *hogyan határozhatók meg egyértelműen az A_k , m_k , σ_k konstansok értékei?*

Ezzel a problémával többek közt a következő vizsgálatokban találkozunk:

1. Nem-folytonos spektrumok fényképezésénél diszkrét vonalak helyett a valóságban folytonosan változó, maximumokkal bíró intenzitáseloszlást kapunk. Ennek az az oka, hogy a Doppler-effektus következtében minden vonalhoz egy Gauss-függvény szerint lefutó intenzitáseloszlás tartozik és ezek szuperponálódnak [2]. Ennek folytán az eredő intenzitáseloszlást a következő függvény fejezi ki:

$$I(x) = \sum_{k=1}^N \frac{I_k}{2\sqrt{\pi\kappa T}} e^{-\frac{(x-m_k)^2}{4\kappa T}}$$

(Itt x a felvétel sávjának hossz-koordinátája, valamilyen alkalmas origótól számítva, I_k az egyes komponensek szolgáltatja összes intenzitás, m_k a spektrumvonalak elméleti helye, κ univerzális állandó, T a sugárzást kibocsátó anyag abszolút hőfoka). Minthogy a felvételtől a spektrumvonalak valóságos helyeit, az m_k -kat kell meghatároznunk, ismét problémánkkal van dolgunk. Itt minden σ_k ugyanaz; a spektroszkópiai gyakorlatban általában így van; — állócsillag-spektrum vizsgálatoktól kezdve luminiscens porok vizsgálatáig [3].

2. Oldatokban tartalmazott különböző molekulásúlyú fehérjék ultracentrifugálás vagy elektroforézis útján történő szétválasztásakor a komponens fehérjék koncentrációja Gauss-függvény szerint oszlik el bizonyos maximum körül, amelynek helye a kérdéses ion mozgékonyaságától függ. Minthogy pl. optikai (Tiselius-féle készülék) vagy kromatográfiai úton az egyes koncentrációeloszlások keveredésének eredőjét rögzíthetjük csak, ebből a felvételtől még meg kell határoznunk a komponens Gauss-függvények maximumhelyét és egyéb adatait [4]. Ennek kapcsán láthatjuk problémánk megoldásának fontosságát, ha arra gondolunk, hogy az emberi vér fehérjefrakcióinak vizsgálatát diagnosztikai célból gyakran kell elvégezni [5].

Problémánk alkalmazási területeinek ismertetése után most rátérünk a megoldás különböző módjaira.

1. §.

A kérdést néhány, speciális esetekkel foglalkozó cikk [6] után először G. Doetsch tárgyalta összefoglalóan és gyakorlatilag is kielégítően [7]. (A komponensekre bontást ő nevezte el a Fourier-analízis mintájára »Gauss-analízis«-nek). Eredményeit sikerrel alkalmazták a spektroszkópiában [8].

Más területen azonban nem használták fel (bizonyára félve a matematikai nehézségektől), úgyhogy legtöbbször csak a következő, elég pontatlan grafikus eljárással találkozunk [9]:

Tekintsük a komponensekre bontandó függvényt. Ennek görbéje általában néhány Gauss-görbe szuperpozíciója. Feltehető, hogy a görbe jobb, ill. bal szélén az utolsó, ill. első Gauss-görbe dominál, minthogy a szomszédosak legtöbbször olyan távol vannak, hogy az előbbieket alig befolyásolják. Előre elkészített, különböző szórású és magasságú Gauss-görbe sablonokat illesztve a görbe bal, ill. jobb széléhez, találunk egy olyat, amely a legjobban egybeesik a görbeszélekkel. Ennek az adatai megadják az első, ill. utolsó komponens adatait. Az odaillesztett sablont teljes egészében is berajzoljuk az analizálandó görbébe és grafikusan levonjuk. Ily módon két komponenszt leválasztottunk. A maradék-görbén ismét alkalmazzuk az eljárást, mindaddig, míg egyetlen komponens sem marad. (Újabbban a sablonokat rávetítik a görbére, ez lényegesen megkönnyíti a munkát.)

Ez a módszer sok pontatlansággal jár és sokszor nem dönthető el, megkaptuk-e az összes komponenseket? (Ennek illusztrálására később még visszatérünk.)

Mindezekkel szemben *Doetsch* két pontos módszert is adott. Az egyikben grafikusan sorban kiszámítja az analizálandó függvény momentumait. Ezek A_k , m_k és σ_k polinomjai, tehát elegendő számú momentum ismeretében kellő számú egyenletet kapunk az ismeretlenek meghatározására. Az egyenletek azonban magasabb fokúak és így kezelésük nagyon nehéz.

Doetsch másik módszere lett a kiindulópontja a későbbi idevágó vizsgálatoknak. Ezt a módszert a »szóráscsökkentés módszerének« nevezhetjük, mert alapgondolata a következő:

Legyen az analizálandó függvény

$$f(x) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2\sigma_k^2}}}{\sqrt{2\pi} \sigma_k}$$

Tegyük fel, hogy ennek az ismeretében elő tudjuk állítani az

$$f^*(x) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2(\sigma_k^2 - \lambda^2)}}}{\sqrt{2\pi} \sqrt{\sigma_k^2 - \lambda^2}}$$

függvényt. (λ valós paraméter; λ kisebb a legkisebb szórásnál.) Ez olyan Gauss-függvények szuperpozíciója, amelyek maximumhelye ugyanaz, mint $f(x)$ komponenseinek, csak a szórásnégyzetek csökkentek ugyanazzal a λ^2 értékkel. $f^*(x)$ komponensei tehát keskenyebbek és magasabbak, mint $f(x)$ -éi, következésképp a szuperpozíció után is kevésbé keverednek össze, mint $f(x)$ -ben. Legjobban az a komponens válik ki, amelyiknek a legkisebb a szórása, vagyis ahol $\sigma_k^2 - \lambda^2$ a legkisebb. Minthogy ez erősen kiemelkedik a görbéből, mint külön Gauss-görbe kezelhető, amelynek adatai grafikusan megállapíthatók. (Legbiztosabban a maximum helye határozható meg, s épp ez kell a legtöbbször. A többi adat meghatározásához legjobb a teljes Gauss-görbét extrapolálni és a maximumhely magasságából, valamint a görbe alatti

területből a Gauss-függvény ismert tulajdonságai alapján az A_k és σ_k értékeket kiszámítani.) Általában egyszerre több komponens is különválnak, ha nem így volna, az egyetlen különvált komponens grafikusán levonva, a maradékon újból elvégezzük az eljárást. Minél jobban megközelíti λ^2 a legkisebb σ_k^2 -et, annál jobban különválnak a megfelelő komponens. Így tehát λ^2 kellő választásával az egyes komponenseket egyre pontosabban kiemelhetjük, ill. leválaszthatjuk. Az eljárás nem tételezi fel a komponensek számának ismeretét, csak azt, hogy az m_k számok különbözők. Ha azonban ez teljesül, az eljárás egyértelmű is, minthogy az elkeskenyedett és túszerűen kiemelkedő maximumokat az m_k -knál kapjuk meg. Egyenlő szórás esetén egyszerre keskenyednek el a komponensek és így egyszerre válnak szét.

A főfeladat tehát az $f(x)$ -ből $f^*(x)$ -be vivő transzformáció megkeresése. Erre több módszert dolgoztak ki.

2. §.

G. Doetsch eredetileg csak egyenlő szórások esetével foglalkozott [10]. Az alapgondolat azonban különböző szórásoknál is alkalmazható.

Képezzük az

$$f(x) = \sum_{k=1}^N A_k \frac{e^{-\frac{(x-m_k)^2}{2\sigma_k^2}}}{\sqrt{2\pi} \sigma_k}$$

Fourier-transzformáltját, $\varphi(t)$ -t. Ismeretes, hogy [11]

$$\varphi(t) = \sum_{k=1}^N A_k \cdot e^{im_k t - \frac{\sigma_k^2}{2} t^2}.$$

$$f^*(x) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2(\sigma_k^2 - \lambda^2)}}}{\sqrt{2\pi} \sqrt{\sigma_k^2 - \lambda^2}}$$

Fourier-transzformáltja $\varphi^*(t)$ pedig :

$$\varphi^*(t) = \sum_{k=1}^N A_k \cdot e^{im_k t - \frac{(\sigma_k^2 - \lambda^2)}{2} t^2} = e^{\frac{\lambda^2}{2} t^2} \cdot \varphi(t)$$

Az unicitási tétel értelmében tehát

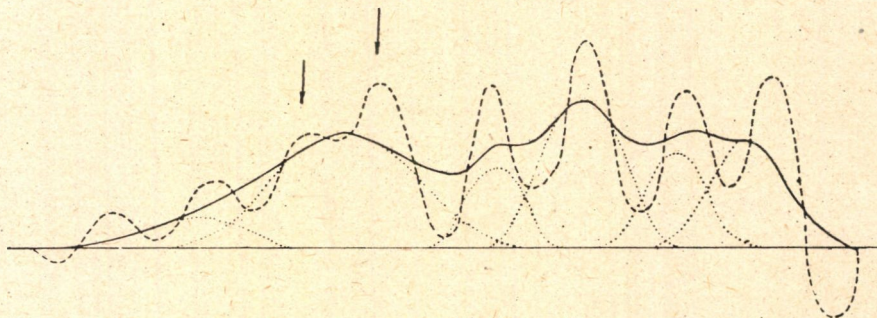
$$f^*(x) = \mathfrak{F}^{-1} \left\{ e^{\frac{\lambda^2}{2} t^2} \cdot \varphi(t) \right\},$$

ahol $\mathfrak{F}^{-1}$ az inverz Fourier-transzformációt jelenti. (A mellékfeltételek teljesülése könnyen ellenőrizhető.)

Ennek alapján tehát $f^*(x)$ $f(x)$ -ből a következőképpen kapható meg :

Képezzük $f(x)$ Fourier-transzformáltját, megszorozzuk $\frac{\lambda^2}{e^2} t^2$ -tel (ahol λ^2 kisebb a legkisebb σ_k^2 -nél) és az így kapott függvényt visszatranszformáljuk. Ha λ^2 -re támpontunk nincs, különböző értékekkel próbálkozunk. Ha a legkisebb σ_k -t túlléptük, az integrálok divergensnek és az eljárás értelmét veszti.

Az eljárás gyakorlatban persze nem végezhető el pontosan, már csak a végtelen határu integrálások miatt sem. Közelfekvő gondolat azonban a Fourier-integrálok helyett Fourier-sorokkal próbálkozni. Elég nagy periódust választva, az $f(x)$, ill. $f^*(x)$ Fourier-együtthatóit kifejező integrálok a Gauss-függvény rohamos csökkenése folytán tetszőleges keveset térnek el a Fourier-transzformált értékétől, és így a transzformáltak közti összefüggés közelítőleg az együtthatók közt is felírható. Így közelítőleg megkaphatók $f^*(x)$ Fourier-együtthatói, $f^*(x)$ -et pedig ezen együtthatókkal elvégzett Fourier-szintézis adja meg.



1. ábra

Ezt a módszert *Doetsch* is alkalmazta egyenlő szórások esetén [12]. Intézetünkben emberi vér fehérjefrakcióinak elektroforetikus vizsgálatakor nyert görbét dolgoztunk fel hasonló módon (1. ábra). A vastag vonal az eredeti görbe, a szaggatottat bizonyos λ^2 paraméterértéknél kaptuk. Az első 18 Fourier-együtthatóval dolgoztunk. Jól látható a komponensek szétválása. Pontozott vonallal a fentebb említett grafikus eljárás eredményét tüntettük fel. Látható, hogy a grafikus módszer két különböző komponenst egyetlennek állapított meg (nyilak). A negatív értékek a Fourier-analízisből adódnak ; ha 18 együtthatónál többet vettünk volna, az ilyen kis ingadozások kevésbé mutatkoznának.

Ez a közelítő módszer tág teret nyújt matematikai gépek alkalmazására. A Fourier-analizátorok és szintetizátorok irodalma igen nagy. Modern elektronikus készülékekkel az eljárást gyorsan végre lehetne hajtani. Mi Mader—Ott-analizátorral dolgoztunk, a szintézist egyenlő távolságban felvett pontokban numerikusan végeztük el. A feldolgozás számológép nélkül cca 6 órát vett igénybe ; számológéppel fele idő is elég.

Érdekes, hogy a kezdeti nekilendülés után ezt az eljárást többet nem alkalmazták. Valószínűleg a nem matematikus kutatók idegenkedtek tőle.

Csak éppen megemlítjük, hogy szerzőnek sikerült az alapötletet más (pl. Cauchy-féle) sűrűségfüggvények keverékének felbontására is alkalmazni.

3. §.

Egy másik eljárást, amely azonban csak egyenlő σ szórások esetén alkalmazható, *H. Freudenthal* ismertette az 1953. évi varsói matematikai kongresszuson [13]. Az eljárás kiindulópontja *Doetsch* korábbi dolgozatában [14] is szerepel: az

$$f(x, t) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2t}}}{\sqrt{2\pi} \sqrt{t}} \quad (t > 0)$$

függvény kielégíti a

$$\frac{\partial f}{\partial t} = \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}$$

hővezetési differenciálegyenletet. $t = \sigma^2$ -nél ismerjük ennek az egyenletnek a megoldását: ez ugyanis éppen az adott függvény. Ennek alapján határozzuk meg a megoldást valamilyen $t = \sigma^2 - k \cdot \Delta t > 0$ értékre ($\Delta t > 0$ tetszőleges kis szám, $k = 1, 2, \dots$). Ebben a megoldásban a komponensek arányosan csökkentett szórással jelentkeznek, vagyis különválva láthatók.

A *Freudenthal* által ismertett módszer abban áll, hogy a hővezetés differenciálegyenletét differenciaegyenletként írjuk fel az (x, t) síkon felvett Δx , ill. Δt ponttávolságú hálón. A háló (x_i, t) pontjaihoz ($i = 1, 2, \dots$) a kiinduló függvény értékei tartoznak az x_i pontokban. E pontokra a differenciaegyenlet

$$\frac{f(x_i, t) - f(x_i, t - \Delta t)}{\Delta t} = \frac{f(x_i + \Delta x, t) + f(x_i - \Delta x, t) - 2f(x_i, t)}{2 (\Delta x)^2}$$

Itt $f(x_i + \Delta x, t)$, $f(x_i - \Delta x, t)$, $f(x_i, t)$, Δt , Δx ismert, így ezekből $f(x_i, t - \Delta t)$ meghatározható. Pontról pontra haladva sorban meghatározhatjuk az $f(x_i, t - \Delta t)$ értékeket, vagyis végül megkapjuk a $t = \sigma^2 - \Delta t$ -hez tartozó megoldást bizonyos hibával, amely a háló sűrítésével csökken. Az eljárást megismételve megkapjuk a megoldást $t = \sigma^2 - 2 \cdot \Delta t$ -nél, é. i. t., $\sigma^2 - k \cdot \Delta t$ -ig.

Az eljárás minden gépi segédeszköz nélkül is elvégezhető, mert csak a négy alapművelet szerepel. Hátránya, hogy csak egyenlő szórásoknál alkalmazható.

4. §.

Az $f^*(x)$ transzformált függvény más módon is előállítható explicite. Állítsuk elő a keresett

$$f^*(x) = f^*(x, \mu) = \sum_{k=1}^N A_k \cdot \frac{e^{-\frac{(x-m_k)^2}{2(\sigma_k^2 - \mu)}}}{\sqrt{2\pi} \sqrt{\sigma_k^2 - \mu}} \quad (\mu = \lambda^2)$$

függvényt μ hatványsoraként. Minthogy $f^*(x, \mu)$ kielégíti a

$$\frac{\partial f^*}{\partial \mu} = -\frac{1}{2} \cdot \frac{\partial^2 f^*}{\partial x^2}$$

hővezetési differenciálegyenletet, $\frac{\partial^n f^*}{\partial \mu^n} = \left(-\frac{1}{2}\right)^n \frac{\partial^{2n} f^*}{\partial x^{2n}}$,

$$\left[\frac{\partial^n f^*}{\partial \mu^n} \right]_{\mu=0} = \left[\left(-\frac{1}{2}\right)^n \frac{\partial^{2n} f^*}{\partial x^{2n}} \right]_{\mu=0} = \left(-\frac{1}{2}\right)^n \cdot \frac{\partial^{2n} f}{\partial x^{2n}},$$

ahol $f(x) = f^*(x, 0)$ a kiindulási függvény és így $f^*(x, \mu)$ Taylor-sora

$$f^*(x, \mu) = \sum_{\nu=0}^{\infty} \left[\frac{\partial^{\nu} f^*}{\partial \mu^{\nu}} \right]_{\mu=0} \cdot \frac{\mu^{\nu}}{\nu!}$$

felírható $f(x)$ x szerinti deriváltjaival:

$$f^*(x, \mu) = \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu} \mu^{\nu}}{2^{\nu} \cdot \nu!} \frac{\partial^{2\nu} f}{\partial x^{2\nu}}$$

A deriváltak ismeretében tehát $f^*(x, \mu)$ közelítőleg meghatározható. A fő nehézség az, hogy $f(x)$ deriváltjait görbéjéből kell meghatározni; ez a véges differenciák segítségével történhetik. A kérdésre még visszatérünk; előbb azonban rámutatunk egy másik lehetőségre.

Ismeretes, hogy egy $\psi(x)$ függvény sorbafejthető Hermite-függvények szerint [15]:

$$\psi(x) = \sum_{\nu=0}^{\infty} \frac{c_{\nu}}{\nu!} \cdot \varphi^{(\nu)}(x)$$

$$\text{ahol } \varphi^{(\nu)}(x) = (-1)^{\nu} \cdot H_{\nu}(x) \cdot \varphi(x), \quad \varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

$H_{\nu}(x)$ pedig a $\frac{d^{\nu}}{dx^{\nu}} \left(e^{-\frac{x^2}{2}} \right) = (-1)^{\nu} H_{\nu}(x) e^{-\frac{x^2}{2}}$ -vel definiált Hermite-polinom.

Ekkor $c_{\nu} = (-1)^{\nu} \int_{-\infty}^{\infty} H_{\nu}(x) \psi(x) dx$.

Legyenek tehát a fenti $f^*(x, \mu) = f^*(x, \lambda^2)$ és $f(x)$ Hermite-sorfejtései:

$$f^*(x, \lambda^2) = f^*(x) = \sum_{\varrho=0}^{\infty} \frac{c_{\varrho}^*}{\varrho!} \varphi^{(\varrho)}(x)$$

és

$$f(x) = \sum_{h=0}^{\infty} \frac{c_h}{h!} \varphi^{(h)}(x);$$

(bebizonyítható, hogy $\sigma_k^2 < 1$ -nél — s ez alkalmas léptékkal mindig elérhető — ezek konvergensek). Mindezeket beírva $f^*(x)$ fenti Taylor-sorába (ennek jogosultsága bebizonyítható):

$$f^*(x) = \sum_{\varrho=0}^{\infty} \frac{c_{\varrho}^*}{\varrho!} \varphi^{(\varrho)}(x) = \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu} \lambda^{2\nu}}{2^{\nu} \nu!} \sum_{h=0}^{\infty} \frac{c_h}{h!} \varphi^{(h+2\nu)}(x)$$

Koefficiens-összehasonlítással kapjuk:

$$c_{\varrho}^* = \sum_{n=0}^{\lfloor \frac{\varrho}{2} \rfloor} \frac{c_{\varrho-2n} \cdot \varrho! \cdot (-1)^n \lambda^{2n}}{(\varrho-2n)! n! 2^n}$$

azaz $f(x)$ Hermite-sorának együtthatóiból $f^*(x)$ Hermite-együtthatói egyszerűen meghatározhatók.

Ezzel a módszerrel $f^*(x)$ tehát úgy nyerhető, hogy $f(x)$ -et Hermite-sorba fejtjük, azután együtthatóival az előbbi képlet szerint képezzük $f^*(x)$ Hermite-együtthatóit, végül az ezekkel szorzott Hermite-függvényeket összegezzük (»Hermite-szintézis«). Az egyes eljárásokban szereplő függvények összegezésének elvégzésére szerző matematikai gépet is tervezett; ennek leírása Kiadványunk más helyén található. Magától értetődően a sorfejtések és összegeзések mindig csak bizonyos tagszámig haladnak; a hiba esetenként megbecsülhető.

5. §.

Befejezésül foglalkozzunk azzal, hogyan használhatjuk fel a gyakorlatban $f^*(x, \lambda^2)$ -nek a 4. §-ban közölt

$$f^*(x, \lambda^2) = \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu} \lambda^{2\nu}}{2^{\nu} \cdot \nu!} \frac{\partial^{2\nu} f}{\partial x^{2\nu}}$$

előállítását.

A differenciászámítás módszereit fogjuk igénybe venni. Legyen a az $f(x)$ értelmezési tartományának egy pontja és $h > 0$ adott szám. Ismeretes hogy az $f(a)$, $f(a \pm h)$, $f(a \pm 2h)$, ... függvényértékek ismeretében $f(x)$ a helyen vett deriváltjai közelítőleg kifejezhetők a függvény a helyre vonatkozó magasabb differenciáival.

Állapodjunk meg abban, hogy $\nu = 3$ -nál megállunk $f^*(x, \lambda^2)$ sorfejtésében. Ekkor $f(x)$ deriváltjait az a helyen a következő képletek adják meg [16]:

$$f''(a) = \frac{2\sigma_3 - 27\sigma_2 + 270\sigma_1 - 490f_0}{180h^2} + \frac{47}{8480} h^6 \varepsilon M_8$$

$$f^{(4)}(a) = \frac{-\sigma_3 + 12\sigma_2 - 39\sigma_1 + 56f_0}{6h^4} + \frac{403}{5040} h^4 \varepsilon M_8$$

$$f^{(6)}(a) = \frac{\sigma_3 - 6\sigma_2 + 15\sigma_1 - 20f_0}{h^6} + \frac{169}{420} h^2 \varepsilon M_8$$

ahol

$$f_0 = f(a), \quad \sigma_\nu = f(a + \nu h) + f(a - \nu h), \quad (\nu = 1, 2, 3)$$

$$|\varepsilon| \leq 1, \quad M_8 = \max_{a-3h \leq x \leq a+3h} |f^{(8)}(x)|.$$

(Hasonló kifejezések találhatók a magasabb deriváltakra is, csak ott magasabb differenciák szerepelnek.)

Adott esetben legegyszerűbben úgy járhatunk el, hogy alkalmasan választott h értékkel a feldolgozandó görbe adataiból megállapítjuk az $f(\pm k \cdot h)$ ($k = 0, 1, 2, \dots$) értékeket, majd a fenti képletek segítségével közelítőleg kiszámítjuk a deriváltak értékeit a $0, \pm h, \pm 2h, \dots$ pontokban. Előre megválasztott λ^2 érték mellett sorfejtésünk alapján ekkor már megadhatjuk $f^*(x, \lambda^2)$ közelítő értékeit a $0, \pm h, \pm 2h, \dots$ pontokban és megrajzolhatjuk a további analízis alapjául szolgáló függvénygörbét.

Foglalkozunk még az eljárás hibájával az a helyen. Ez két forrásból ered: Egyrészt a Taylor-sor $\nu > 3$ indexű tagjainak elhanyagolásából, másrészt a differenciálhányadosok közelítő megállapításából. Jelöljük a két hibafajtát δ_1 , ill. δ_2 -vel.

Fenti képleteink alapján

$$|\delta_2| \leq \left(\frac{47}{8480} h^6 \frac{\lambda^2}{2} + \frac{403}{5040} h^4 \frac{\lambda^4}{8} + \frac{169}{420} h^2 \frac{\lambda^6}{48} \right) \varepsilon M_8.$$

M_8 -ban $f^{(8)}(x)$ maximuma szerepel bizonyos intervallumban. Ennek becs-

lése $f(x)$ alakja folytán lényegében véve egy $\varphi(x) = \frac{e^{-\frac{x^2}{2\tau^2}}}{\sqrt{2\pi\tau}}$ ($\tau > 0$, konstans)

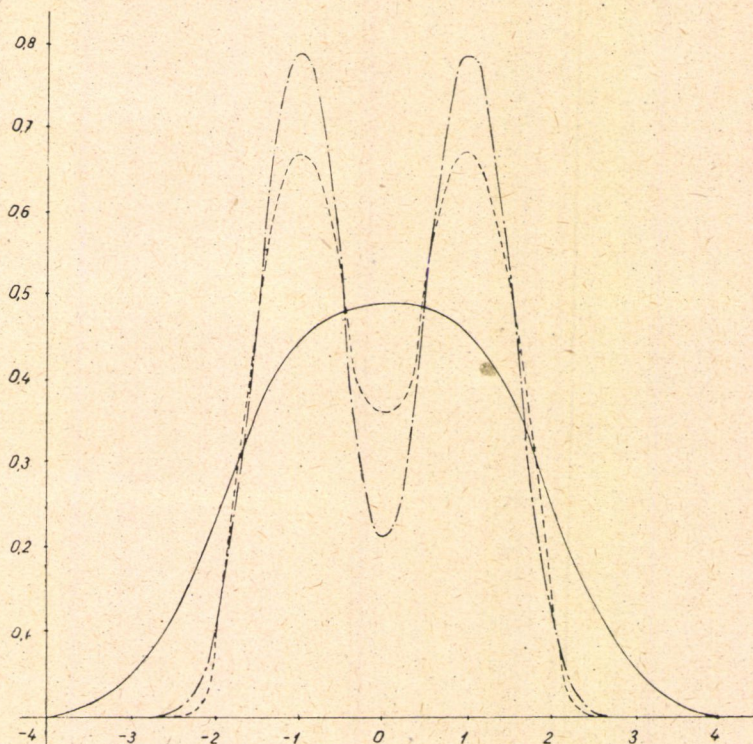
típusú függvény deriváltjainak, vagyis Hermite-függvényeknek a becslésére vezet. Ismeretes, hogy [17]:

$$\begin{aligned} \frac{d^{2n}}{dx^{2n}} (e^{-x^2}) = \\ = e^{-\frac{x^2}{2}} \left[(-1)^n 2^n \cdot 1 \cdot 3 \dots (2n-1) \cos \sqrt{4n+1} x + \frac{2^n \sqrt{(2n)!} \sqrt{\pi}}{\sqrt{5} \sqrt{4n+1}} |x|^5 \Theta_n(x) \right] \\ |\Theta_n(x)| < 1. \end{aligned}$$

Ennek alapján a jobboldal mindenik tagját maximumával becsülve könnyen kimutatható, hogy

$$\left| \frac{d^{2n}}{dx^{2n}} \left(\frac{e^{-\frac{x^2}{2\tau^2}}}{\sqrt{2\pi} \tau} \right) \right| \leq \frac{1}{\sqrt{2\pi} \tau^{2n+1}} \left[\frac{(2n)!}{2^n \cdot n!} + \frac{0,56 \sqrt{(2n)!}}{\sqrt{4n+1}} \right]$$

Ennek segítségével adható meg végül becslés δ_2 -re. δ_1 becslésére szintén felhasználhatjuk, minthogy δ_1 a Taylor-sor maradéktagja, amelyben ugyancsak Hermite-függvények szerepelnek különböző szórásokkal.



2. ábra

Mindezeket itt nem részletezhetjük, csak megjegyezzük, hogy Taylor-sorunk maradéktagja úgy tart zérushoz, mint $\frac{1}{\sqrt{n}}$. Az itt alkalmazott (elég durva) becslésekkel a hiba (főleg δ_1) elég lassan csökken. A becslést érthetően megnehezíti az, hogy $f(x)$ komponenseinek paramétereit tág határok közt állapíthatjuk csak meg (v. ö. a grafikus eljárásnál mondottakkal).

Illusztrálásul bemutatjuk az

$$f(x) = \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}} \right)$$

függvény analízisét $\lambda^2 = \frac{3}{4}$ -del, a sorfejtést a 4. tagnál hagyva abba (2. ábra). A vastag vonal mutatja $f(x)$ -et, amelynek egyetlen maximuma van, A szagatott vonal $f^*(x, \lambda^2)$ a mondott közelítésben. Összehasonlításként pontvonallal berajzoltuk a pontos eredményt;

$$f^*(x) = \frac{2}{\sqrt{2\pi}} (e^{-2(x-1)^2} + e^{-2(x+1)^2})$$

görbáját.

Amint látjuk, nagyobb eltérés csak az extremumok környékén van. Ez a hiba képletéből várható is. A hiba jóval kisebb, mint a becslésekből számított. A maximumok helyét (ami általában a legfontosabb) már ez a közelítés is jól megadja.

Ha nagyobb pontosság nem szükséges, ez az eljárás kényelmes, mint-hogy elvégzése semmi segédeszközt nem igényel.

IRODALOM

- [1] *Rényi Alfréd*: Újabb eredmények a valószínűségszámítás terén. MTA. Mat. Fiz. Oszt. Közl. II. 1. Budapest, 1952. 125—139. o. (főleg a 132. o.-tól.)
- [2] *G. Doetsch*: Die Elimination des Dopplereffekts bei spektroskopischen Feinstrukturen und exakte Bestimmung der Komponenten. Zs. Phys. 49 (1928). 705—730. o.
- [3] L. pl. *B. Sticker*: Über die Farbenhäufigkeitsfunktion in Sternhaufen. Zs. Astrophys. 1 (1930). 174—191. o.; *A. Kiss—C. Sándorfy*: Sur les méthodes d'analyse des courbes d'adsorption. Acta chem. et phys. Szeged. Tom. II. Fasc. 3 (1948). 71—76. o.; *Szigeti György*: Lumineszkáló anyagok. Elektrotechnika 1947. 4—5. sz.
- [4] L. pl. *The Svedberg—K. O. Pedersen*: Die Ultrazentrifuge. Dresden u. Leipzig. 1940. 265. o.;
- E. Wiedemann*: Über die Auswertung von Elektrophorese-Diagrammen nach L. G. Longworth und Philpot—Svensson. Helv. Chim. Acta. XXX. Pars. I. Fasc. I—V. (1947). 892—900. o.
- [5] L. pl. *F. Wuhrmann—Ch. Wunderly*: Die Bluteiweisskörper des Menschen. Basel. 1947. 106. o.
- [6] L. pl. *N. Sen*: Über den Einfluss des Dopplereffekts auf spektroskopische Feinstrukturen und seine Elimination. Phys. Zs. 23 (1922). 397. o.; *K. Pearson*: Contributions to the mathematical theory of evolution. Phil. Trans. Roy. Soc. London, (A), 185 (1894). 71—110. o.
- [7] *G. Doetsch*: Zerlegung einer Funktion in Gauss'sche Fehlerkurven und zeitliche Zurückverfolgung eines Temperaturzustandes. Math. Zs. 41 (1936). 283—318. o.; továbbá a (2) alatt idézett mű.
- [8] L. pl. a [3] alatt idézett csillagászati cikket és a [2] alatti cikk példáit.
- [9] L. pl. *A. Wallner—R. Ulke*: Auswertung von Diagrammen aus Elektrophoreseversuchen. Hoppe—Seyler's Zs. Phys. Chem. Bd. 290. Heft 3—6. 81—91. o.; *H. Labhart*: Ein Auswertegerät für Elektrophoresediagramme. Experientia. 36 (1947). 3. és *A. Kiss—C. Sándorfy*, valamint *Szigeti György* [3] alatt idézett dolgozatait.
- [10] *L. G. Doetsch* idézett dolgozatait.
- [11] L. pl. *B. V. Gnyegyenko—A. N. Kolmogorov*: Független valószínűségi változók összegeinek határeloszlásai. Budapest, 1951. 52. o.
- [12] L. a (2) alatt idézett művét.
- [13] *Rényi A.* közlése.
- [14] L. a [2] alatt idézett dolgozatot.
- [15] L. pl. a [11] alatt idézett monográfiát. 198. o.
- [16] *L. L. V. Kantorovics—V. I. Kriülov*: A felsőbb analízis közelítő módszerei. Budapest, 1953. 189—199. o.
- [17] *L. B. И. Смирнов*: Курс высшей математики. Москва, 1949, Том. III. Часть 2. 587. o.

ОПРЕДЕЛЕНИЕ КОМПОНЕНТОВ СМЕСИ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

П. Меддешин

Резюме

Автор занимается следующей проблемой: если плотность вероятностей, заданную эмпирически, можно рассматривать как смесь с постоянными весами нормальных функций плотности разных параметров, как определить компоненты? — Практическая важность проблемы иллюстрируется спектроскопическими и биологическими примерами, потом автор занимается графическими и аналитическими методами решения, следуя Г. Деч. Г. Деч получает решение с помощью преобразования Фурье. Автор-следуя идее А. Реньи-представляет решение в форме функционального ряда, построенного с помощью производных высших порядков смеси.

Автор иллюстрирует исследования двумя примерами. Дальше он описывает свое исследование о возможности обобщения и механического решения.

DÉTERMINATION DES COMPONENTS D'UN MÉLANGE DES FONCTIONS DE DISTRIBUTION

P. MEDGYESSY

Résumé

L'auteur traite du problème suivant :

Si, dans le calcul des probabilités, une fonction de fréquence (donnée de façon empirique) peut être comprise comme un mélange des fonctions normales de fréquence de paramètres différentes, pris avec des poids constants, comment peut-on déterminer les composants? — L'auteur expose l'importance du problème par des exemples spectroscopiques et biologiques, ensuite il traite des méthodes graphiques et analytiques de la solution, en se joignant aux recherches de G. Doetsch. Pour G. Doetsch la solution est donnée avec l'aide de la transformée de Fourier du mélange, tandis que l'auteur — suivant une idée de A. Rényi — l'obtient par une série des fonctions formée des dérivées du mélange. — La discussion est illustrée de deux exemples. L'auteur mentionne aussi quelques-unes de ses recherches se rapportant à la généralisation et à la solution par appareil.