

HANKA LÁSZLÓ– VINCZE ÁRPÁD

**GAMMA-SPEKTRUMOK KIÉRTÉKELÉSÉNEK
MATEMATIKAI MÓDSZEREI III.
A MAXIMUM ENTRÓPIA MÓDSZER**

**MATHEMATICAL METHODS
OF GAMMA SPECTRUM'S EVALUATION III.
THE MAXIMUM ENTROPY METHOD**

A maximum entrópia módszer (MEM) egy olyan valószínűségelméleti eljárás, amely eredményesen alkalmazható gamma-spektrumok dekonvolúciójára. A maximum entrópia elvnek számos előnye van a hagyományos dekonvolúciós módszerekhez képest. A módszer alkalmazásával nagyobb felbontású spektrumot kapunk, mint a lineáris regularizációs technikákkal, a megoldásként kapott spektrum szükségképpen nemnegatív és a módszer lehetőséget ad arra is, hogy egyszerűen figyelembe vegyünk a χ^2 statisztikát, amellyel kompenzálni lehet a fluktuációk hatását. Ebben a dolgozatban az elmélet matematikai alapjait ismertetjük. Kulcsszavak: Gamma-spektrum, információ, valószínűségi változó, entrópia, maximum entrópia módszer, Lagrange-módszer, χ^2 -statisztika, kereszt-érvényesítési eljárás, relatív entrópia, Poisson-eloszlás.

Maximum entropy method (MEM) is a probabilistic method, which can be successfully applied to deconvolution of gamma-ray spectra. The maximum entropy model has several advantages over conventional methods. It provides a better resolution than linear regularisation methods, the solution is positively constrained, and it also allows one to include the additional χ^2 statistic to compensate for the fluctuations in real spectra. In this paper we present the mathematical basis of this theory. Keywords: Gamma-ray spectra, information, probability distribution, entropy, maximum-entropy method, Lagrange-method, χ^2 -statistics, cross-validation method, relative entropy, Poisson distribution.

1. Bevezetés

Az alábbiakban egy szcintillációs gamma-spektrométerrel felvett spektrum meghatározásának valószínűségi módszereivel foglalkozunk. A probléma vázlatosan a következő. A mérendő energiatartományt osszuk

fel n db intervallumra az E'_0, E'_1, \dots, E'_n osztópontokkal. Tegyük fel, hogy a j -edik $[E'_{j-1}, E'_j]$ energiatartományban a beütésszám x_j , a gamma-fotonok száma összesen pedig $x_1 + x_2 + \dots + x_n = N$. A spektrométer csatornáinak a száma legyen m . Osszuk fel tehát a mérőeszköz által vizsgált energiatartományt m részre, ahol az osztópontokat, $E_0, E_1, E_2, \dots, E_m$ jelöli. Tegyük fel, hogy az i -edik csatornában, tehát az $[E_{i-1}, E_i]$ energiaintervallumban mért beütésszám y_i ($i = 1, 2, \dots, m$). Feltesszük, hogy teljesül az $y_1 + y_2 + \dots + y_m = N$ feltétel, tehát a detektor minden gamma fótont regisztrál. Az $y = (y_1, y_2, \dots, y_m)$ mérési adatok, és a tényleges $x = (x_1, x_2, \dots, x_n)$ spektrum közötti matematikai kapcsolat a következő módon írható fel:

$$(1.1) \quad y = Rx + \varepsilon$$

ahol az $R = [R_{ij}] \in R^{m \times n}$ mátrix a detektor válaszfüggvénye, $\varepsilon \in R^m$ pedig a zaj. Az R mátrix R_{ij} eleme annak az eseménynek a valószínűsége, hogy az i -edik energiatartományba tartozó gamma-foton a j -edik csatornában van detektálva. Mivel ez egy lineáris egyenletrendszer, kézenfekvő a probléma megoldását a lineáris algebra módszereivel meghatározni. Az alkalmazható regularizációs módszerekről részletesen olvashatunk [1]-ben. Az egyenletrendszer megoldására azonban eredményesen, sőt nagyobb hatékonysággal alkalmazhatók a valószínűségelmélet és a matematikai statisztika módszerei. Ezen módszerek megvilágítása érdekében rámutatunk arra, hogy ha a fotonok száma összesen N , akkor a

$$(1.2) \quad p_j = \frac{x_j}{N}; (j = 1, 2, \dots, n)$$

nem negatív értékek valószínűségeloszlást alkotnak. Hasonlót állíthatunk a detektorral mért y spektrumról is, ha az y_i mérési adatokat szintén osztjuk N -el. A spektrumot tehát – az N konstanstól eltekintve – úgy tekinthetjük, mint egy ismeretlen valószínűségeloszlást. Az alábbiakban kifejtendő maximum-entrópia módszer ennek az eloszlásnak a meghatározására ad lehetőséget.

2. Az entrópia

Az entrópia fogalma megjelenik az információelméletben, a statisztikus fizikában [2], és számos alkalmazásra talál a fizika egyéb területein, a csillagászatban és az orvostudományban. A konkrét fizikai tartalomtól

eltekintve, az információtól, mint matematikai mennyiségtől az alábbi tulajdonságokat várjuk el:

Minél nagyobb egy esemény valószínűsége, bekövetkezésével annál kevesebb információhoz jutunk, hiszen az információt úgy is tekinthetjük, mint egy adott eseménnyel kapcsolatban a bizonytalanságunk mértékét. Minimális az információtartalma a biztos eseménynek, mivel ebben az esetben a bizonytalanságunk zérus. Minél kisebb tehát az esemény valószínűsége, az általa hordozott információ annál több, tehát az információ az esemény valószínűségének növekedésével csökken. Az I információ eszerint az esemény p valószínűségének a függvénye: $I = I(p)$. A biztos esemény bekövetkezésével minimális információhoz jutunk, legyen ez definíció szerint zérus, tehát $I(1) = 0$. A 0 valószínűségű eseményhez pedig a legnagyobb információmennyiség tartozik.

Ha két független esemény egyszerre következik be, elvárjuk, hogy az általuk hordozott információ összeadódjék. Mivel ekkor a valószínűségek szorzódnak, teljesülnie kell az

$$(2.1) \quad I(p_1 \cdot p_2) = I(p_1) + I(p_2)$$

függvényegyenletnek. A mondott követelményeknek megfelel az

$$(2.2) \quad I(p) = -k \cdot \ln p$$

függvény, ahol k egy pozitív konstans. Definíció szerint tehát $I(p) = -k \cdot \ln p$ egy p valószínűségű esemény bekövetkezésével nyert információ. Legyen most ξ , egy olyan diszkrét eloszlású valószínűségi változó,

melynek eloszlása: $P = \{p_1, p_2, \dots, p_n\}$, $p_j \geq 0$, $\sum_{j=1}^n p_j = 1$. Azonban nem

tudjuk biztosan, hogy a p_j ($j = 1, 2, \dots, n$) valószínűségű események közül melyik esemény következik be. Ezért a ξ valószínűségi változóval kapcsolatosan csak a $-k \cdot \ln p_j$ információk átlagértékét, az alábbiakban E -vel jelölt várható értéket („Expected value”) kaphatjuk információként. Legyen

$$(2.3) \quad S(P) = E[-k \cdot \ln p_j] = \sum_{j=1}^n p_j (-k \cdot \ln p_j) = -k \sum_{j=1}^n p_j \cdot \ln p_j$$

a P valószínűség-eloszláshoz tartozó információ átlagos mennyisége. Ezt a mennyiséget a P valószínűségeloszlás entrópiájának nevezzük. Ha a k konstans értékét 1-nek választjuk, és a természetes logaritmus he-

lyett 2-alapú logaritmust alkalmazunk, akkor kapjuk az információelméletben használatos „információentrópia” fogalmát:

$$(2.4) \quad S(P) = -k \sum_{j=1}^n p_j \cdot \log_2 p_j$$

melynek mértékegysége ebben az esetben 1 bit. Ha a (2.3) képletben a k konstans éppen a Boltzmann-féle állandó, akkor S a statisztikus fizikából ismert entrópia. Ebben az esetben p_j egy adott makroállapot megvalósulásának a valószínűségét jelenti. Az alábbiakban – mivel elsősorban fizikai alkalmazásokról lesz szó –, a k arányossági tényezőt 1-nek választjuk, és általánosan entrópiának nevezzük az

$$(2.3) \quad S(P) = - \sum_{j=1}^n p_j \cdot \ln p_j$$

mennyiséget.

Keressünk most kapcsolatot az entrópia fogalma és a gamma sugárzás detektálásával kapcsolatos probléma mennyiségei között. Mélyebb fizikai megfontolások nélkül, egyszerű kombinatorikai megfontolással kapjuk azon lehetőségek W számát, amikor – az 1. pont jelöléseivel élve –, az N db gamma foton közül éppen x_j db foton tartozik a j -edik, $[E_{j-1}, E_j]$ energiatartományba:

$$(2.4) \quad W = \frac{N!}{x_1! \cdot x_2! \cdot \dots \cdot x_n!}$$

Ezt a W mennyiséget a statisztikus fizikában az adott makroállapot termodinamikai valószínűségének nevezik. Ez nem más, mint az adott makroállapotot megvalósító mikroállapotok száma. Esetünkben W jelentése: egy adott $x \in \mathbb{R}^n$ spektrum ennyiféle módon valósítható meg N db gamma-fotonnal. A számítások egyszerűbben végrehajthatók, ha a W mennyiség logaritmusát képezzük:

$$(2.5) \quad \ln W = \ln \frac{N!}{x_1! \cdot x_2! \cdot \dots \cdot x_n!} = \ln \frac{N!}{(Np_1)! \cdot (Np_2)! \cdot \dots \cdot (Np_n)!} = \ln(N!) - \sum_{j=1}^n \ln(Np_j)!$$

Alkalmazzuk most a faktoriálisok aszimptotikus közelítésére vonatkozó Stirling-féle formulát, amely szerint

$$n \rightarrow \infty \text{ esetén } n! \approx \left(\frac{n}{e} \right)^n,$$

amely kifejezés logaritmikus alakban az

$$(2.6) \quad \ln(n!) \approx n \cdot \ln(n) - n$$

formát ölti. Ennek felhasználásával (2.5) a következő alakban írható:

(2.7)

$$\begin{aligned} \ln W &\approx N \cdot \ln N - N - \sum_{j=1}^n (Np_j \cdot \ln(Np_j) - Np_j) = N \cdot \ln N - N - N \sum_{j=1}^n (p_j \cdot \ln(Np_j)) + N \cdot \sum_{j=1}^n p_j = \\ &= N \cdot \ln N - N \sum_{j=1}^n (p_j \cdot \ln N) - N \cdot \sum_{j=1}^n p_j \ln p_j = N \cdot \ln N - N \cdot \ln N \sum_{j=1}^n p_j - N \cdot \sum_{j=1}^n p_j \ln p_j = \\ &= -N \cdot \sum_{j=1}^n p_j \ln p_j \end{aligned}$$

Ha tehát a gamma fotonok száma elég nagy, akkor az adott $x \in \mathbb{R}^n$ spektrumot megvalósító lehetőségek számának logaritmus – az N arányossági tényezőtől eltekintve – éppen az

$$S = - \sum_{j=1}^n p_j \cdot \ln p_j$$

entrópiával egyezik meg.

Visszatérve az entrópia információelméleti vonatkozására, a következőt mondhatjuk. Ha $p_j = 1$, és $p_k = 0$ ha $k \neq j$, tehát a biztos eseménnyel állunk szemben, akkor $S = 0$, hiszen $\ln 1 = 0$, továbbá közismert tény, hogy $\lim_{x \rightarrow 0} x \ln x = 0$. Ekkor tehát az entrópia minimális. Innen az is világos,

hogy annál kisebb az entrópia értéke, minél több olyan p_j valószínűség van a P eloszlásban, amely relatíve nagy, közel van 1-hez. De felmerül a kérdés, hogy mely esetben maximális az entrópia értéke, és mennyi a maximuma. Ha semmiféle megszorító feltételezéssel nem

élünk, csak a $\sum_{j=1}^n p_j = 1$ normálási feltétellel, akkor az $S(p) = S(p_1, p_2, \dots$

, $p_n)$ függvény a maximumát a

$$\frac{\partial S(p_1, p_2, \dots, p_n)}{\partial p_j} = 0; \quad (j = 1, 2, \dots, n)$$

egyenletek által meghatározott helyen veheti fel. Az egyenletrendszer megoldása helyett, a logaritmus függvény tulajdonságainak ismeretében elemi úton is előállíthatjuk az entrópia maximumát. Vizsgáljuk ennek érdekében az $S - \ln(n)$ különbséget:

$$S - \ln(n) = -\sum_{j=1}^n p_j \ln p_j - \ln n = -\sum_{j=1}^n p_j \ln p_j - \sum_{j=1}^n p_j \ln n = -\sum_{j=1}^n p_j \ln(p_j n) = \sum_{j=1}^n p_j \ln\left(\frac{1}{p_j n}\right)$$

Használjuk most fel a közismert $\ln x \leq x - 1$ egyenlőtlenséget. Ekkor kapjuk, hogy

$$S - \ln(n) \leq \sum_{j=1}^n p_j \left(\frac{1}{p_j n} - 1\right) = \sum_{j=1}^n \left(\frac{1}{n} - p_j\right) = 1 - 1 = 0$$

Az S entrópia maximuma tehát $\ln n$, és mivel az $\ln x = x - 1$ egyenlőség az $x = 1$ helyen teljesül, az adódik, hogy az entrópia a $p_j = \frac{1}{n}$, ($j = 1, 2, \dots, n$) egyenletes eloszlás esetén maximális. Megmutattuk tehát, hogy ha a $\sum_{j=1}^n p_j = 1$ normálási feltételen kívül semmiféle feltételezéssel nem

élünk, akkor a maximális entrópiájú, tehát maximális bizonytalanságú, legtöbb információt hordozó eloszlás a

$$p_1 = p_2 = \dots = p_n = \frac{1}{n}$$

egyenletes eloszlás, és ebben az esetben az entrópia maximális értéke $\ln(n)$. A következőkben azt vizsgáljuk, hogyan általánosíthatók az említett információelméleti megfontolások arra az esetre, amikor a normálási feltétel mellett, az adott probléma kapcsán felmerülő egyéb megszorításokkal is élünk. Ez a gondolat vezet el a maximum-entrópia módszeréhez.

3. A maximum entrópia elv

A maximum entrópia módszer (MEM) egy olyan eljárás, amellyel a rendelkezésre álló információk birtokában egyértelműen meghatározható az a valószínűségeloszlás, amely egy adott probléma leírásához a legjobban megfelelő, leginkább ésszerű [2]. A 2. pontban kiderült, hogy ha ismeretek hiányában a normálási feltétel mellett semmiféle egyéb információval nem rendelkezünk a keresett eloszlással kapcsolatban, akkor az entrópia az egyenletes eloszlás esetén maximális. Tegyük most fel, hogy rendelkezünk a probléma leírását elősegítő információval is.

Azt állítjuk, hogy ha olyan eloszlást választunk, amelyhez a rendelkezésre álló információkkal összhangban maximális entrópia tartozik, akkor ez a legnagyobb bizonytalanságot jelentő és egyben a leginkább racionális valószínűségeloszlás. Tegyük fel ugyanis, hogy egy ettől eltérő olyan eloszlást választunk, amelynek entrópiája kisebb, tehát kisebb a bizonytalanságunk, kevesebb információt várunk a megoldástól. Ez azt jelenti, hogy olyan információt is feltételeztünk, amellyel nem rendelkezünk. Ha pedig olyan eloszlást választunk, amelyhez az adott feltételekhez tartozó maximumnál nagyobb entrópia tartozik, tehát nagyobb a bizonytalanságunk, akkor az azt jelenti, hogy nem veszünk figyelembe, elhanyagolunk olyan információkat, amelyekkel viszont rendelkezünk. A maximális entrópiájú eloszlás tehát a legésszerűbb megoldás.

Kvantitatíve feladatunk tehát az

$$(3.1) \quad S = - \sum_{j=1}^n p_j \cdot \ln p_j$$

entrópia maximalizálása abban az esetben, ha a $\sum_{j=1}^n p_j = 1$ feltétel mellett

rendelkezünk még az eloszlásra vonatkozó egyéb információval. Első lépésként egyetlen feltétellel élve, ez az információ legyen – az eloszlás konkrét ismeretének hiányában – egy $f(x)$ függvény várható értéke:

$$(3.2) \quad E[f(x)] = \sum_{j=1}^n p_j \cdot f(x_j) = F$$

Keressük tehát a (3.1) entrópia maximumát a normálási feltétel és a (3.2) feltétel mellett. Ez tehát egy feltételes szélsőértékfeladat, amelyet a Lagrange-módszerrel oldhatunk meg. Alkalmazva ennek érdekében a λ_0 és λ Lagrange-féle multiplikátorokat, kapjuk, hogy maximalizálandó az

$$(3.3) \quad L(p_1, p_2, \dots, p_n, \lambda_0, \lambda) = - \sum_{j=1}^n p_j \ln p_j + (\lambda_0 - 1) \left(1 - \sum_{j=1}^n p_j \right) + \lambda \left(F - \sum_{j=1}^n p_j f(x_j) \right)$$

Lagrange-függvény. (A $\lambda_0 - 1$ együtthatót ebben a formában a számítások egyszerűsítése érdekében alkalmaztuk.) Szélsőérték ott lehet, ahol teljesülnek a

$$\frac{\partial L}{\partial p_j} = 0; \quad (j = 1, 2, \dots, n)$$

feltételek. Elvégezve a deriválást a következő egyenletek adódnak:

$$\frac{\partial L}{\partial p_j} = -(\ln p_j + 1) - (\lambda_0 - 1) - \lambda f(x_j) = 0; \quad (j = 1, 2, \dots, n)$$

A megoldás az alábbi:

$$(3.4) \quad p_j = \exp(-\lambda_0 - \lambda f(x_j)); \quad (j = 1, 2, \dots, n)$$

A λ_0 és λ Lagrange-multiplikátorok a $\sum_{j=1}^n p_j = 1$ és a (3.2) feltételekbe

törtéző helyettesítéssel adódnak.

$$\sum_{j=1}^n p_j = \sum_{j=1}^n \exp(-\lambda_0 - \lambda f(x_j)) = e^{-\lambda_0} \sum_{j=1}^n \exp(-\lambda f(x_j)) = 1$$

Ha itt bevezetjük az elméletben szokásos

$$(3.5) \quad Z(\lambda) = \sum_{j=1}^n \exp(-\lambda f(x_j))$$

jelölést, akkor a $\lambda_0 = \ln Z(\lambda)$ összefüggést kapjuk. Az $f(x)$ függvény $E[f(x)] = F$ átlagértékére pedig az

$$(3.6)$$

$$\begin{aligned} E[f(x)] &= \sum_{j=1}^n \exp(-\lambda_0 - \lambda f(x_j)) \cdot f(x_j) = e^{-\lambda_0} \sum_{j=1}^n \exp(-\lambda f(x_j)) \cdot f(x_j) = \\ &= \frac{1}{Z(\lambda)} \left(-\frac{\partial}{\partial \lambda} Z(\lambda) \right) = -\frac{\partial}{\partial \lambda} \ln Z(\lambda) = F \end{aligned}$$

összefüggés adódik. Ebből az egyenletből határozható meg a λ szorzó értéke. Ennek ismeretében pedig $\lambda_0 = \ln Z(\lambda)$ alapján adódik λ_0 . Ha a megoldások a birtokunkban vannak, már könnyen meghatározhatjuk az entrópia maximális értékét:

(3.7)

$$S_{\max} = -\sum_{j=1}^n p_j \ln p_j = -\sum_{j=1}^n p_j \cdot (-\lambda_0 - \lambda f(x_j)) = \lambda_0 \sum_{j=1}^n p_j + \lambda \sum_{j=1}^n p_j f(x_j) = \lambda_0 + \lambda F$$

Általánosítsuk most a fentiekben megoldott szélsőérték problémát arra az esetre, amikor r db feltételünk van, mindegyik egy $f_k(x)$ függvény

$E[f_k(x)] = F_k$ átlagértéke formájában ($k = 1, 2, \dots, r$). Vezessük most be a $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_r$ Lagrange-féle szorzókat. Ebben az esetben a (3.3) Lagrange függvény általános alakja:

(3.8)

$$L(p_1, \dots, p_n, \lambda_0, \dots, \lambda_r) = -\sum_{j=1}^n p_j \ln p_j + (\lambda_0 - 1) \left(1 - \sum_{j=1}^n p_j \right) + \sum_{k=1}^r \lambda_k \left(F_k - \sum_{j=1}^n p_j f_k(x_j) \right)$$

A szélsőérték az egy feltételre elvégzett számítások értelemszerű módosításával adódik. Ha bevezetjük (3.5) mintájára a

$$(3.9) \quad Z(\lambda_1, \dots, \lambda_r) = \sum_{j=1}^n \exp(-\lambda_1 f_1(x_j) - \lambda_2 f_2(x_j) - \dots - \lambda_r f_r(x_j))$$

függvényt, akkor a λ_0 konstans értéke változatlanul $\lambda_0 = \ln Z(\lambda)$ alakban adódik, a maximális entrópiájú valószínűségeloszlás pedig a következő:

$$(3.10) \quad p_j = \exp(-\lambda_0 - \lambda_1 f_1(x_j) - \dots - \lambda_r f_r(x_j)); \quad (j = 1, 2, \dots, n)$$

Ez jól láthatóan a (3.4) eloszlás megfelelője több feltétel esetére. Az $f_k(x)$ függvények átlagértékére vonatkozólag az

$$(3.11) \quad E[f_k(x)] = -\frac{\partial}{\partial \lambda_k} \ln Z(\lambda_1, \dots, \lambda_r) = F_k; \quad (k = 1, 2, \dots, r)$$

egyenletrendszer adódik. Ez az r db egyenlet szolgál a $\lambda_1, \dots, \lambda_r$ együtthatók meghatározására. A gyakorlatban a megoldásra numerikus módszereket alkalmaznak, mert a megoldás ritkán áll elő zárt, analitikus alakban. A megoldás ismeretében könnyen előállíthatjuk az entrópia maximumát:

$$(3.12) \quad S_{\max} = \lambda_0 + \lambda_1 F_1 + \dots + \lambda_r F_r$$

4. A maximum entrópia módszer alkalmazása gammaspektrumok meghatározására

A 3. pontban általánosan meghatároztuk a maximális entrópiájú valószínűségeloszlást. Most — az 1. pontban bevezetett jelölések használatával — alkalmazzuk az elméletet az $x \in R^n$ spektrum meghatározására. A 2. pontban megmutattuk, hogy ha összesen N db gamma fótont detektál a spektrométer, és az $x = (x_1, x_2, \dots, x_n)$ spektrum megvalósulási lehetőségeinek száma W , akkor W logaritmusá éppen a spektrumnak, mint valószínűség-eloszlásnak az entrópiájával egyezik meg. A N arányossági tényezőtől eltekintve (2.7) szerint:

$$\ln W \approx S = -\sum_{j=1}^n p_j \ln p_j$$

Ha alkalmazzuk az (1.2)-ben bevezetett $p_j = \frac{x_j}{N}$ jelölést, akkor az entrópia matematikai alakját a számítások szempontjából praktikusabb formára hozhatjuk:

$$\begin{aligned} S &= -\sum_{j=1}^n p_j \ln p_j = -\sum_{j=1}^n \frac{x_j}{N} \ln \frac{x_j}{N} = -\frac{1}{N} \sum_{j=1}^n x_j \ln \frac{x_j}{N} = -\frac{1}{N} \sum_{j=1}^n (x_j \ln x_j - x_j \ln N) = \\ &= -\frac{1}{N} \sum_{j=1}^n x_j \ln x_j + \frac{1}{N} \ln N \sum_{j=1}^n x_j = -\frac{1}{N} \sum_{j=1}^n x_j \ln x_j + \ln N \end{aligned}$$

Mivel a maximum-entrópia elv alkalmazása az S entrópia szélsőértékének meghatározását jelenti, az S fenti kifejezésében eltekinthetünk az $\frac{1}{N}$ szorzótól és az $\ln N$ additív állandótól, hiszen ezek nem befolyásolják a szélsőértéket. Az alábbiakban az $x \in R^n$ spektrumhoz rendelt entrópiának fogjuk nevezni az

$$(4.2) \quad S = -\sum_{j=1}^n x_j \ln x_j$$

mennyiséget [3]. A maximum-entrópia elv szerint ennek a függvénynek a maximuma szolgáltatja az ideális spektrumot. Ahogyan azt a 2. pontban láttuk, ha csak a nyilvánvaló $x_1 + x_2 + \dots + x_n = N$ feltételt írjuk elő, akkor az egyenletes eloszlás esetén veszi fel az entrópia a maximumát.

Ez nyilván nem egy valós gamma-spektrum, de világos az is, hogy a gyakorlatban ismerjük a mérésel kapott $y \in \mathbb{R}^m$ spektrumot, továbbá a detektor $R = [R_{ij}] \in \mathbb{R}^{m \times n}$ válaszfüggvényét. A mérési adatok pedig természetesen módon megszorító feltételeket jelentenek a keresendő spektrumra vonatkozólag. Ha $x \in \mathbb{R}^n$ a feltételezett spektrum, akkor a válaszfüggvény a

$$(4.3) \quad d_i = \sum_{j=1}^n R_{ij} x_j; \quad (i = 1, 2, \dots, m)$$

értékeket szolgáltatja. A mérésel kapott $y = (y_1, y_2, \dots, y_m)$ spektrum, és a számítás eredményeként adódó $d = (d_1, d_2, \dots, d_m)$ közötti eltérés, a zaj, a statisztikus ingadozásoknak, a detektor sajátosságainak tulajdonítható. Ennek a két vektornak az $\varepsilon = y - d$ különbsége a jól ismert χ^2 statisztikával vehető figyelembe. Ha feltételezzük, hogy σ_i az i -edik csatornában a mérési hiba szórása — amit a gyakorlatban ismertnek tételeznek fel [4] —, akkor az m -szabadságfokú χ^2 eloszlás a

$$(4.4) \quad \chi^2 = \sum_{i=1}^m \frac{(d_i - y_i)^2}{\sigma_i^2}$$

formulával adható meg. Ennek az eloszlásnak a várható értéke, $E[\chi^2]$, közismerten m , ami ebben az esetben éppen a detektor csatornáinak száma. Ez az információ a gamma spektrum meghatározásához egy járulékos feltétel. Vegyük észre, hogy ez pontosan megfelel a 2. pontban bevezetett $f(x)$ függvénynek. Az $f(x)$ várható értékére vonatkozó (3.2) feltétel így az $E[\chi^2] = m$ alakot ölti. Maximalizáljuk ezek után a (4.2) entrópiát ennek a feltételnek a figyelembe vételével. A Lagrange-függvény alakja, a (4.3) összefüggés figyelembe vételével a következő:

$$(4.5) \quad L(x_1, \dots, x_n, \lambda_0, \lambda) = -\sum_{j=1}^n x_j \ln x_j + \lambda_0 \left(N - \sum_{j=1}^n x_j \right) + \frac{\lambda}{2} \left(m - \sum_{i=1}^m \frac{(d_i - y_i)^2}{\sigma_i^2} \right) =$$

$$= -\sum_{j=1}^n x_j \ln x_j + \lambda_0 \left(N - \sum_{j=1}^n x_j \right) + \frac{\lambda}{2} \left(m - \sum_{i=1}^m \frac{\left(\sum_{j=1}^n R_{ij} x_j - y_i \right)^2}{\sigma_i^2} \right)$$

Ez a (3.3) Lagrange-függvény a vizsgált esetben. (A Lagrange-szorót ismét a számítások egyszerűsítése végett alkalmaztuk „felezett” alakban.) A szélsőérték a $\frac{\partial L}{\partial x_j} = 0$; ($j = 1, 2, \dots, n$) egyenletek megoldásával állítható elő. Elvégezve a deriválást, azt kapjuk, hogy:

$$\frac{\partial L}{\partial x_j} = -(\ln x_j + 1) - \lambda_0 - \frac{\lambda}{2} \sum_{i=1}^m \frac{2 \left(\sum_{j=1}^n R_{ij} x_j - y_i \right) R_{ij}}{\sigma_i^2} = -\ln x_j - 1 - \lambda_0 + \lambda \sum_{i=1}^m R_{ij} \frac{(y_i - d_i)}{\sigma_i^2} = 0$$

($j = 1, 2, \dots, n$). Ezekből az egyenletekből az ideális spektrumot az

(4.6)

$$x_j = \exp \left(-1 - \lambda_0 + \lambda \sum_{i=1}^m R_{ij} \frac{(y_i - d_i)}{\sigma_i^2} \right) = C \cdot \exp \left(-1 + \lambda \sum_{i=1}^m R_{ij} \frac{(y_i - d_i)}{\sigma_i^2} \right);$$

($j = 1, 2, \dots, n$)

formula szolgáltatja [3], ahol $C = \exp(-\lambda_0)$, állandó. Ez az általános alakú (3.4) megoldás konkrét megfelelője. A λ állandó ismeretében λ_0 értéke a (3.5) összefüggésre támaszkodva, a $\lambda_0 = \ln Z(\lambda)$ alapján számítható.

A maximum-entrópia módszer több előnnyel is rendelkezik a hagyományos lineáris algebrai módszerekhez, regularizációs technikákhoz képest. Mindenekelőtt a dekonvolvált spektrum a (4.6) összefüggés szerint nem negatív, ami alapvető követelmény egy spektrummal szemben. A regularizáció módszerek csak komoly nehézségek árán tudják biztosítani ennek a kívánalomnak a teljesülését. Másodsor, a normalizálási feltétel csak egy konstansként van jelen a megoldandó egyenletekben, és a megoldásban.

A (4.6) formula szolgáltatja tehát a mérési adatokkal és a fluktuációkkal összhangban a legvalószínűbb spektrumot. Mivel azonban a d_i

értékek a $d_i = \sum_{j=1}^n R_{ij} x_j$; ($i = 1, 2, \dots, m$) összefüggéseken keresztül

függenek az x_j ($j = 1, 2, \dots, n$) megoldásoktól, a (4.6) egyenletet legcélszerűbb iterációval megoldani. Az $x \in R^n$ spektrum nulladik közelítése, tehát az iteráció kezdőértéke legyen az entrópia abszolút maximumát jelentő egyenletes eloszlás:

$$x^{(0)} = \left(\frac{N}{n}, \frac{N}{n}, \dots, \frac{N}{n} \right)$$

Ha ismerjük a mérésrel kapott $y \in \mathbb{R}^m$ spektrumot, az y_i mérési adatok, mint valószínűségi változók σ_i szórását valamint a detektor $R \in \mathbb{R}^{m \times n}$ válaszfüggvényét, és első lépésben adunk egy kicsi értéket a λ szorzónak, akkor előállíthatjuk az x spektrumot az alábbi iterációs formulával:

$$(4.7) \quad x_j^{(k+1)} = C \cdot \exp \left(-1 + \lambda \sum_{i=1}^m R_{ij} \frac{\left(y_i - \sum_{j=1}^n R_{ij} x_j^{(k)} \right)}{\sigma_i^2} \right)$$

$k \geq 0, j = 1, 2, \dots, n$. Első kérdés, hogy az iteráció hányadik lépésben ér véget. Ehhez választunk egy kicsiny $\delta > 0$ pozitív számot. Ha adott L esetén minden $j = 1, 2, \dots, n$ indexre teljesül, hogy $\left| x_j^{(L+1)} - x_j^L \right| < \delta$, akkor az L -edik iterációs lépés eredményét fogadjuk el megoldásnak. Az iterációval kapcsolatban felmerülő második kérdés, hogy a λ együtthatót, amit szokás regularizációs paraméternek is nevezni, hogyan választjuk meg. Erre vonatkozólag két eljárást említünk.

Az első módszer alkalmazása során a λ értékét próbálgatással határozzuk meg. Ennek lényege, hogy elsőként annyira kicsire választjuk λ -t, hogy az iteráció konvergáljon. Ezek után a λ értékét növeljük addig – minden egyes esetben lefuttatva az iterációt –, amíg a χ^2 értéke el nem éri a várható értékét, m -et. Az ilyen módon kapott megoldást fogadjuk el ideális spektrumnak.

Ettől elegánsabb és hatékonyabb módszer a λ megválasztására, az ún. „keresztértékesítési eljárás” [5]. Ennek során a λ értékét szisztematikusan és egyértelműen határozhatjuk meg az alább ismertető módon. A módszer előnye, hogy alkalmazása nem igényel semmiféle előzetes hipotézist, és független a χ^2 statisztikától. Tegyük fel, hogy a mérési adatok a szokásos jelölésekkel a következők: y_1, y_2, \dots, y_m . A módszer lényege abban áll, hogy az y_i (i rögzített) adatot megbecsüljük a többi, $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_m$ mérési adat birtokában. Ezen $m - 1$ db adat felhasználásával előállítjuk az $x_{1(i)}, \dots, x_{n(i)}$, megoldásvektort – az (i) in-

de az arra utal, hogy az y_i érték becsléséről van szó –. Az $x_{j(i)}$ megoldások ismeretében az $y_{i(i)}(\lambda)$ becslült érték az

$$y_{i(i)}(\lambda) = \sum_{j=1}^n R_{ij} x_{j(i)}$$

formulával adódik. A λ regularizációs paraméter értékéül azt a valós számot választjuk melynek alkalmazásával a legjobb becslést kapjuk az y_i mérési adatokra vonatkozólag. Konkrétabban azt a λ értéket alkalmazzuk az iterációban, amelyre a becslült értékek és a mérési eredmények közötti négyzetes eltérés átlaga minimális. Definiáljuk ebből a célból a

$$C(\lambda) = \frac{1}{m} \sum_{i=1}^m \frac{(y_{i(i)}(\lambda) - y_i)^2}{\sigma_i^2}$$

függvényt. A λ optimális értékét a $C(\lambda)$ minimumhelye szolgáltatja.

5. A relatív entrópia fogalma és alkalmazása a spektrumok vizsgálatában

A gamma spektrumok vizsgálata során szokás az entrópia egy általánosabb alakjának alkalmazása is. A fogalom tisztázásának érdekében tegyük fel, hogy adott egy $P = \{p_1, p_2, \dots, p_n\}$ valószínűségeloszlás. Az egyes események bekövetkezésével kapcsolatban azonban modellfeltevéseink vannak, amely szerint az eloszlás a tényleges P helyett $Q = \{q_1, q_2, \dots, q_n\}$. Mielőtt megfigyeléseinket elvégezzük, egzaktul a P valószínűségeloszlás

$$S(P) = - \sum_{j=1}^n p_j \cdot \ln p_j$$

entrópiája méri a bizonytalanságunkat, információink hiányát. Ezzel szemben modellfeltevéseinkből ugyanezre a bizonytalanságra az

$$S(P, Q) = - \sum_{j=1}^n p_j \cdot \ln q_j$$

érték, az ún. keresztentrópia adódik.

A kettő különbségére a

$$D(P|Q) = S(P, Q) - S(P) = -\sum_{j=1}^n p_j \ln q_j - \left(-\sum_{j=1}^n p_j \ln p_j \right) =$$

$$= \sum_{j=1}^n p_j \ln p_j - \sum_{j=1}^n p_j \ln q_j = \sum_{j=1}^n p_j \ln \frac{p_j}{q_j}$$

Kifejezés adódik. A $D(P|Q)$ mennyiséget a P valószínűségeloszlás Q eloszlásra vonatkozó relatív entrópiájának, vagy a két eloszlás Kullback-Liebler távolságának nevezzük [6]. (Hangsúlyozzuk, hogy $D(P|Q)$ nem „metrika” a hagyományos értelemben, mert nem szimmetrikus: $D(P|Q) \neq D(Q|P)$.) Könnyen kimutatható, hogy $D(P|Q)$ nem negatív. Ehhez alkalmazzuk a konvex függvényekre vonatkozó

$$\sum_{j=1}^n \alpha_j f(x_j) \geq f\left(\sum_{j=1}^n \alpha_j x_j\right); \quad \alpha_j \geq 0, \quad \sum_{j=1}^n \alpha_j = 1$$

Jensen-egyenlőtlenséget a konvex $(-\ln x)$ függvényre és a $p_j \geq 0$, $\sum_{j=1}^n p_j = 1$ feltételt kielégítő együtthatókra:

$$D(P|Q) = \sum_{j=1}^n p_j \ln \frac{p_j}{q_j} = -\sum_{j=1}^n p_j \ln \frac{q_j}{p_j} \geq -\ln \sum_{j=1}^n p_j \frac{q_j}{p_j} = -\ln \sum_{j=1}^n q_j = -\ln 1 = 0$$

A Jensen-egyenlőtlenségből az is adódik, hogy egyenlőség pontosan akkor teljesül, ha minden j -re igaz, hogy $p_j = q_j$, tehát $P = Q$, azaz a két valószínűségeloszlás megegyezik. A relatív entrópia

$$D(P|Q) = S(P, Q) - S(P) \geq 0$$

tulajdonsága úgy is fogalmazható, hogy a hibás feltevésekből adódó bizonytalanság nem lehet kisebb mint a megfigyeléseket megelőző tényleges bizonytalanság értéke.

A spektrum vizsgálatára visszatérve, legyen a valós $x \in \mathbb{R}^n$ spektrum a P eloszlás megfelelője. A modellfeltevéseink szerinti spektrum pedig legyen az $m = (m_1, m_2, \dots, m_n) \in \mathbb{R}^n$ vektorral adva. (Ennek konkrét formájára még visszatérünk.) Ez utóbbi feleljen meg a Q eloszlásnak.

Ekkor az x valós spektrum m modellspektrumra vonatkozó relatív entrópiája:

$$D(x|m) = \sum_{j=1}^n x_j \ln \frac{x_j}{m_j}$$

A fentiek szerint $D(x|m) \geq 0$, és pontosan akkor zérus, ha $x = m$. A feladat most abban áll, hogy $D(x|m)$ értékét minimalizálni kell. Ha összhangot szeretnénk teremteni a maximum-entrópia elvvel, akkor $D(x|m)$ minimalizálása helyett maximalizáljuk D ellentettjét. Ismét a szokásos S jelölést alkalmazva, a feladat az

$$(5.1) \quad S(x, m) = - \sum_{j=1}^n x_j \ln \frac{x_j}{m_j}$$

relatív entrópia maximumának meghatározása. Ennek maximuma tehát zérus, és $x = m$ a maximumhely. Jaynes javaslatára az S entrópia (5.1) pontbeli kifejezését kiegészítették egy olyan taggal, ami x és m eltérését méri [2, 5, 7]. Mellőzve egyelőre minden fizikai indoklást, legyen az entrópia az

$$(5.2) \quad S(x, m) = \sum_{j=1}^n \left((x_j - m_j) - x_j \ln \frac{x_j}{m_j} \right)$$

összefüggéssel értelmezve. Megmutatjuk, hogy $S(x, m)$ nem vesz fel pozitív értéket. Egyben általánosítjuk is a problémát két tetszőleges valószínűség-eloszlásra. Legyen P és Q a bevezetőben említett két eloszlás, és legyen

$$S(P, Q) = \sum_{j=1}^n \left((p_j - q_j) - p_j \ln \frac{p_j}{q_j} \right) = \sum_{j=1}^n p_j \left(\left(1 - \frac{q_j}{p_j} \right) - \ln \frac{p_j}{q_j} \right)$$

Vizsgáljuk a problémához illeszkedő $f(x) = 1 - x - \ln \frac{1}{x} = 1 - x + \ln x$

függvényt. Mivel $f''(x) = -\frac{1}{x^2}$, ezért az $f(x)$ függvény konkáv. Alkalmazzuk most a konkáv függvényekre vonatkozó

$$\sum_{j=1}^n \alpha_j f(x_j) \leq f \left(\sum_{j=1}^n \alpha_j x_j \right); \quad \alpha_j \geq 0, \quad \sum_{j=1}^n \alpha_j = 1$$

Jensen-egyenlőtlenséget a $p_j \geq 0$, $\sum_{j=1}^n p_j = 1$, együtthatókra. Eszerint

$$\begin{aligned} S(P, Q) &= \sum_{j=1}^n p_j \left(\left(1 - \frac{q_j}{p_j} \right) - \ln \frac{p_j}{q_j} \right) \leq \sum_{j=1}^n p_j - \sum_{j=1}^n p_j \frac{q_j}{p_j} + \ln \sum_{j=1}^n p_j \frac{q_j}{p_j} = \\ &= \sum_{j=1}^n p_j - \sum_{j=1}^n q_j + \ln \sum_{j=1}^n q_j = 1 - 1 + \ln 1 = 0 \end{aligned}$$

Visszatérve a fizikai alkalmazásra, azt kaptuk, hogy a Jaynes-féle $S(x, m)$ entrópia értéke nem pozitív és maximumát, a nulla értéket $x = m$ esetén veszi fel.

Az entrópia Jaynes által javasolt általánosabb alakjának fizikai indoklásához modellezzük a vizsgált radioaktív bomlási folyamatot – számláló-berendezéssel regisztrált beütések számát – Poisson-eloszlással [4]. Tegyük fel tehát, a gamma bomlás fizikai tulajdonságaira támaszkodva, hogy a j -edik energiatartományba tartozó fotonok száma Poisson-eloszlást követ, melynek átlagértéke λ_j . Modellfeltevésünk tehát az, hogy az átlagos beütésszám $m_j = \lambda_j$ ($j = 1, 2, \dots, n$). (λ_j itt a Poisson-eloszlás paramétere.) Ebben az esetben annak valószínűsége, hogy a j -edik energia intervallumban a beütésszám éppen x_j , az alábbi formulával írható le:

$$(5.3) \quad P(x_j) = \frac{m_j^{x_j}}{x_j!} \exp(-m_j); j = 1, 2, \dots, n$$

A tapasztalatok szerint feltehető, hogy az egyes energiatartományokban a beütésszámok egymástól függetlenek. Ebben az esetben annak valószínűsége, hogy a spektrum éppen $x = (x_1, x_2, \dots, x_n)$, az (5.3) valószínűségek szorzata lesz:

$$(5.4) \quad P = \prod_{j=1}^n P(x_j) = \prod_{j=1}^n \frac{m_j^{x_j}}{x_j!} \exp(-m_j)$$

Keressük azt a spektrumot, amelyre ez a valószínűség maximális. A számítások egyszerűsítése érdekében vegyük (5.4) logaritmusát, majd alkalmazzuk a (2.6) Stirling-formulát. Ekkor a következőt kapjuk:

$$\begin{aligned} \ln P &= \ln \prod_{j=1}^n \frac{m_j^{x_j}}{x_j!} \exp(-m_j) = \sum_{j=1}^n \ln \left(\frac{m_j^{x_j}}{x_j!} \exp(-m_j) \right) = \sum_{j=1}^n (x_j \ln m_j - m_j - \ln x_j!) = \\ &= \sum_{j=1}^n (x_j \ln m_j - m_j - (x_j \ln x_j - x_j)) = \sum_{j=1}^n \left((x_j - m_j) - x_j \ln \frac{x_j}{m_j} \right) \end{aligned}$$

Ha tehát élünk azzal az ésszerű, tapasztalat által megerősített feltevés-
sel, hogy a radioaktív bomlás Poisson-eloszlással írható le, akkor a
spektrum leírásánál természetesen módon adódik az entrópia Jaynes által
javasolt általánosabb formája. A maximum entrópia elv alkalmazásánál
vegyük figyelembe az entrópia így definiált alakját. Ebben az esetben a
(3.3) illetve (4.5) Lagrange-függvény az alábbi formában írható:

$$L(x_1, \dots, x_n, \lambda_0, \lambda) = \sum_{j=1}^n \left((x_j - m_j) - x_j \ln \frac{x_j}{m_j} \right) + \lambda_0 \left(N - \sum_{j=1}^n x_j \right) + \frac{\lambda}{2} \left(m - \sum_{i=1}^m \frac{\left(\sum_{j=1}^n R_{ij} x_j - y_i \right)^2}{\sigma_i^2} \right)$$

A maximumhely ebben az esetben is a $\frac{\partial L}{\partial x_j} = 0$; ($j = 1, 2, \dots, n$) egyen-

letek megoldásával állítható elő. Elvégezve a deriválást, azt kapjuk,
hogy:

$$\frac{\partial L}{\partial x_j} = 1 - \left(\ln \frac{x_j}{m_j} + 1 \right) - \lambda_0 - \frac{\lambda}{2} \sum_{i=1}^m \frac{2 \left(\sum_{j=1}^n R_{ij} x_j - y_i \right) R_{ij}}{\sigma_i^2} = -\ln \frac{x_j}{m_j} - \lambda_0 + \lambda \sum_{i=1}^m R_{ij} \frac{(y_i - d_i)}{\sigma_i^2} = 0 \quad (j = 1, 2, \dots, n).$$

Ezen egyenletek megoldása az

(5.6)

$$x_j = m_j \exp \left(-\lambda_0 + \lambda \sum_{i=1}^m R_{ij} \frac{(y_i - d_i)}{\sigma_i^2} \right) = C \cdot m_j \exp \left(\lambda \sum_{i=1}^m R_{ij} \frac{(y_i - d_i)}{\sigma_i^2} \right);$$

($j = 1, 2, \dots, n$)

spektrumot szolgáltatja [5,7], ahol alkalmaztuk a $C = \exp(-\lambda_0)$ egysze-
rűsítő jelölést. A λ állandó értékét például a kereszt érvényesítési eljá-
rással határozhatjuk meg. Ennek ismeretében a λ_0 a (3.5) jelölés alkal-

mazásával a $\lambda_0 = \ln Z(\lambda)$ egyenlőség által van meghatározva. Mivel azonban a (4.6) megoldáshoz hasonlóan a d_i értékek a

$$d_i = \sum_{j=1}^n R_{ij} x_j \quad (i = 1, 2, \dots, m)$$

összefüggéseken keresztül függenek az x_j ($j = 1, 2, \dots, n$) megoldásoktól, a spektrumot ismét legcélszerűbb iterációval előállítani. A spektrum nulladik közelítésének tekintjük az $x^{(0)} = \left(\frac{N}{n}, \frac{N}{n}, \dots, \frac{N}{n} \right)$ egyenletes eloszlást, és $k \geq 0$ esetén alkalmazzuk (4.7) analógiájára az

$$(5.7) \quad x_j^{(k+1)} = C \cdot m_j \cdot \exp \left(\lambda \sum_{i=1}^m R_{ij} \frac{\left(y_i - \sum_{j=1}^n R_{ij} x_j^{(k)} \right)}{\sigma_i^2} \right)$$

iterációs formulát. Mint azt korábban is hangsúlyoztuk, a bevett gyakorlatnak megfelelően, a σ_i szórások értékét ismertnek tekintjük. Mivel az entrópia (5.2) alakjának levezetésénél pontosabb fizikai megközelítést alkalmaztunk, természetes módon elvárható, hogy az (5.7) formula alapján kapott megoldás pontosabb a mélyebb fizikai alapokat nélkülöző, egyszerűbb úton kapott (4.7) összefüggés alkalmazásával előállított spektrumnál. Valóban, a tapasztalat szerint a Jaynes-féle entrópia alkalmazásával hatékonyabb a dekonvolúciós eljárás, mint az entrópia egyszerűbb (4.2) formájára támaszkodva. A kapott spektrum részletesebb, nagyobb felbontású, élesebb csúcsokat és mélyebb völgyeket szolgáltat, hatékonyabban felbontja az egymást részben átfedő csúcsokat.

Összefoglalásként elmondhatjuk, hogy a maximum-entrópia elv egy hatékony, nagy felbontású spektrumot szolgáltató módszer a gamma-spektrumok dekonvolúciós technikái között. A lineáris algebra regularizációs módszereihez képest nagy előnye, hogy „pozitív szemidefinit”, tehát csak nem negatív megoldásokat szolgáltat és nagyobb a felbontása is. További előnye például a regularizációs eljárások közül ismert, nagyon hatékony SVD-felbontás alkalmazásával szemben, hogy kevésbé érzékeny a zajra. Hátránya viszont, hogy nagyon sok számítást igényel [3, 5, 7].

Felhasznált irodalom

1. E.T. Jaynes: Information Theory and Statistical Mechanics. Physical Review 106, (1957) 620-630.
2. Hanka László - Vincze Árpád: Gamma-spektrumok kiértékelésének matematikai módszerei II. Regularizációs módszerek. Bolyai Szemle. 3, (2008) 33-53.
3. Jánossy Lajos: Mérési eredmények kiértékelésének elmélete és gyakorlata. Akadémiai Kiadó. 1968
4. Jose M. Los Arcos: Gamma-ray spectra deconvolution by maximum-entropy methods. Nuclear Instruments and methods in Physics Research A (1996) 634-636
5. L. Bouchet: A Comparative study of deconvolution methods for gamma-ray spectra. Astronomy & Astrophysics Supplement Series. Ser. 113, (1995) 167-183.
6. L.J. Meng - D. Ramsden: An inter comparison of Three Spectral-deconvolution Algorithms for Gamma-ray Spectroscopy. IEEE Transactions on nuclear science. 47, (2000) No.4.
7. Thomas M. Cover - Joy A. Thomas: Elements of information theory. Wiley & sons. 2005