

HANKA LÁSZLÓ–VINCZE ÁRPÁD

**GAMMA-SPEKTRUMOK KIÉRTÉKELÉSÉNEK  
MATEMATIKAI MÓDSZEREI IV.  
A MAXIMUM LIKELIHOOD MÓDSZER ÉS  
A VÁRHATÓ ÉRTÉK MAXIMALIZÁLÁSÁNAK ELVE**

**MATHEMATICAL METHODS OF  
GAMMA SPECTRUM'S EVALUATION IV.  
THE MAXIMUM LIKELIHOOD ESTIMATION AND  
THE PRINCIPLE OF EXPECTATION MAXIMIZATION**

---

A maximum likelihood becslés egy nagyon hatékony statisztikai módszer, amelyet akkor alkalmazunk, ha a rendelkezésre álló adatokhoz legjobban illeszkedő matematikai modellt szeretnénk meghatározni. A módszer lehetőséget ad arra, hogy a matematikai modell szabad paramétereit úgy hangoljuk, hogy az illeszkedés optimális legyen. A várható érték maximalizálásának elve olyan valószínűségi modellek paramétereinek maximum likelihood becslésére ad lehetőséget, amelyek olyan „rejtett” változóktól illetve paramétereiktől függenek, amelyeket nem lehet közvetlenül megfigyelni. Kulcsszavak: gamma spektrum, valószínűségeloszlás, paraméterbecslés, likelihood függvény, maximum likelihood becslés, Poisson eloszlás, Gauss eloszlás,  $\chi^2$  statisztika, a várható érték maximalizálásának elve.

---

Maximum likelihood estimation (MLE) is a very effective statistical method used to calculate the best way of fitting a mathematical model to some data. Modeling real data by estimating maximum likelihood offers a way of tuning the free parameters of the model to provide an optimum fit. The expectation maximization algorithm (EM) is used in statistics for finding maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved variables. Keywords: Gamma-ray spectra, probability distribution function, estimation of parameters, likelihood function, maximum likelihood estimation, Poisson distribution, Gauss distribution,  $\chi^2$  statistics, expectation maximization algorithm.

---

## Bevezetés

Az alábbiakban egy szcintillációs gamma-spektrométerrel felvett spektrum meghatározásának maximum likelihood becslésével foglalkozunk. A probléma vázlatosan a következő. A mérendő energiatartományt oszszuk 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 fotonok száma  $x_j$  ( $j = 1, 2, \dots, n$ ). A spektrométer csatornáinak a száma legyen  $m$ . Oszszuk 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, vagyis az  $[E_{i-1}, E_i]$  energiaintervallumban mért beütésszám  $y_i$  ( $i = 1, 2, \dots, m$ ). Feltesszük, hogy a detektor minden gamma fotont 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)  $y = Rx + \varepsilon$ , ahol az  $R = [R_{ij}] \in \mathbb{R}^{m \times n}$  mátrix a detektor válaszfüggvénye,  $\varepsilon \in \mathbb{R}^m$  pedig a zaj. Az  $R$  mátrix  $R_{ij}$  eleme annak az eseménynek a valószínűsége, hogy a  $j$ -edik energiatartományba tartozó gamma-foton az  $i$ -edik csatornában van detektálva. A probléma abban áll, hogy a válaszfüggvény és a mérési adatok birtokában meg kell határoznunk a valós  $x$  spektrumot.

A maximum likelihood becslés alkalmazásánál a  $j$ -edik intervallumba tartozó fotonok számát és a detektor által regisztrált beütésszámot is valószínűségi változónak tekintjük. Legyen a  $j$ -edik energiaintervallumba tartozó gamma fotonok száma a  $\xi_j$  ( $j = 1, 2, \dots, n$ ) valószínűségi változó, az  $i$ -edik csatornában regisztrált beütésszám pedig az  $\eta_i$  ( $i = 1, 2, \dots, m$ ) valószínűségi változó.

Nyilván mindkettő diszkrét eloszlású változó, amely nem negatív értékeket vesz fel. A radioaktív bomlás természetének ismeretében illetve a számláló berendezések által mért adatok tulajdonságainak ismeretében világos, hogy mind  $\xi_j$  mind  $\eta_i$  Poisson eloszlású változók [1]. Ha  $\xi_j$  átlagértéke  $\lambda_j$ , akkor a mondottak szerint annak valószínűsége, hogy a  $j$ -edik tartományban a fotonok száma éppen  $y_j$  az alábbi formulával adható meg:

$$(1.2) \quad P(\xi_j = x_j) = \frac{(\lambda_j)^{x_j}}{(x_j)!} \cdot \exp(-\lambda_j)$$

Továbbá ha  $\eta_i$  átlagértéke  $\lambda_i$ , akkor hasonlóan írható fel annak valószínűsége, hogy az  $i$ -edik csatornában a regisztrált beütésszám  $y_i$ :

$$(1.3) \quad P(\eta_i = y_i) = \frac{(\lambda_i)^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i)$$

Az alábbiakban elsősorban a Poisson eloszlásra, mint modellre támaszkodunk a számításaink során, a fontosabb eredményeket ennek az eloszlásnak az alkalmazásával vezetjük le, bár a 4. pontban kitérünk a normális eloszlással történő közelítés lehetőségére.

## 1. A maximum likelihood becslés

A maximum likelihood becslést (maximum likelihood estimation, MLE) — amelyet magyarul a „legnagyobb valószínűség elvének” lehetne fordítani —, a statisztikában a paraméterbecslések témakörében alkalmazzák. Tegyük fel, hogy adott egy szabad paraméterekkel leírt modell, amellyel feltevéseink szerint leírható az általunk vizsgált rendszer. A paramétereket mérés útján határozzuk meg, de a mérési adatok zajjal terhelték. Ez azt jelenti, hogy magukat a paramétereket nem tudjuk megmérni csak egy-egy valószínűségi változót, amely összefüggésben áll a modell paramétereivel. A mérés szempontjából a modell paramétere is valószínűségi változóknak tekintendők. A kérdés az, hogy a mért értékekből hogyan határozzuk meg a modell paramétereit, hogy a lehető legpontosabb becslést kapjuk.

A maximum likelihood módszert akkor alkalmazhatjuk az ismeretlen paraméterek becslésére, ha ezen paraméterek sűrűségfüggvényei ismeretlenek, de a mérést terhelő zaj eloszlása ismert. Abban az esetben, ha a paraméterek eloszlásáról nem tudunk semmit, a legegyszerűbb egyenletesnek feltételezni azt. A maximum likelihood becslés ezek alapján a következőt jelenti: maximalizálni kell a  $P$  (ez a mérési eredmény | a paraméterek értéke ennyi és ennyi) feltételes valószínűséget. A maximalizálandó függvény analitikus alakjának meghatározásához a Bayes-féle formulát használhatjuk. Ha feltesszük, hogy  $\alpha \in \mathbb{R}^k$  jelöli az ismeretlen paraméterek vektorát (esetünkben az ismeretlen spektrumot leíró paramétereket),  $y \in \mathbb{R}^m$  pedig a méréssel kapott adatok vektora (a detektorral mért spektrum), akkor a Bayes-formula az alábbi alakban írható:

$$(2.1) \quad P(\alpha|y) = \frac{P(\alpha)P(y|\alpha)}{P(y)}$$

Itt a  $P(\alpha)$  valószínűséget „a priori” valószínűségnek, a  $P(\alpha|y)$  valószínűséget „a posteriori valószínűségnek”,  $P(y|\alpha)$ -et pedig „likelihood valószínűségnek” nevezzük. Legyen most  $y_0$  egy konkrét mérési adatokat tartalmazó vektor. Ekkor az

$$(2.2) \quad L(\alpha) = P(y = y_0 | \alpha)$$

függvényt – amely az  $\alpha$  változó, tehát a paraméterek függvénye –, likelihood függvénynek nevezzük. Ez a függvény a mérési zaj eloszlásának ismeretében meghatározható. Az  $\alpha$  paramétervektor maximum likelihood becslése a likelihood függvény  $\alpha$ -paraméter szerinti maximalizálásával adódik. Számítástechnikai szempontból gyakran célszerűbb az  $L(\alpha)$  likelihood függvény helyett ennek természetes logaritmusát, az  $\ln L(\alpha)$  ún. log-likelihood függvényt maximalizálni. A logaritmus függvény monotonitása miatt ennek a két függvénynek a szélsőértékhelyei megegyeznek. Az  $y_i$  ( $i = 1, 2, \dots, m$ ) statisztikai szempontból véletlen mérési adatokat tehát valószínűségi változónak tekintjük, amely változók az  $\alpha_s$  ( $s = 1, 2, \dots, k$ ) paraméterektől függenek. Ha megváltoznak a paraméterek, akkor megváltozik az  $y = (y_1, y_2, \dots, y_m)$  valószínűségi vektorváltozóhoz tartozó együttes valószínűségeloszlás. A problémára tehát úgy tekinthetünk, mint valószínűség eloszlások egy olyan családjára, melyek az  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$  paramétervektorok által vannak indexelve. Kissé általánosabban, jelölje  $f(y | \alpha)$  a mérési adatok  $y = (y_1, y_2, \dots, y_m)$  vektorához tartozó valószínűségeloszlás együttes sűrűségfüggvényét, rögzített  $\alpha \in \mathbb{R}^k$  paramétervektor esetén:

$$(2.3) \quad f(y | \alpha) = f(y_1, y_2, \dots, y_m | \alpha_1, \alpha_2, \dots, \alpha_k)$$

Abban az esetben, ha az egyes mérési adatok egymástól függetlenek, az együttes sűrűségfüggvény az egyes mérési adatokhoz tartozó egyedi sűrűségfüggvények szorzataként írható fel:

$$(2.4) \quad f(y | \alpha) = f(y_1, y_2, \dots, y_m | \alpha_1, \alpha_2, \dots, \alpha_k) = f_1(y_1 | \alpha_1, \alpha_2, \dots, \alpha_k) \cdot f_2(y_2 | \alpha_1, \alpha_2, \dots, \alpha_k) \cdot \dots \cdot f_m(y_m | \alpha_1, \alpha_2, \dots, \alpha_k) = \prod_{i=1}^m f_i(y_i | \alpha)$$

A maximum likelihood módszer lényege ezek után úgy fogalmazható, hogy egy feltételezett modell és adott mérési adatok esetén a modellnek eleget tevő sűrűségfüggvények között meghatározandó az a sűrűségfüggvény, amely a legnagyobb valószínűséggel illeszkedik a mérési

adatokhoz. Mivel a mérési adatok  $y \in \mathbb{R}^m$  vektorát adottnak tekintjük, a sűrűségfüggvény az  $\alpha \in \mathbb{R}^k$  paraméterek függvénye. Úgy is fogalmazhatunk tehát, hogy meghatározandó a paraméterek azon értéke, amely leginkább megfelel a mérési adatokhoz tartozó valószínűség-eloszlásnak. A (2.3) jelölés alkalmazásával a likelihood függvényt az (2.5)  $L(\alpha) = f(y | \alpha)$  módon értelmezzük. Az  $L(\alpha)$  függvény — ismét hangsúlyozva, hogy az  $\alpha$  paraméter függvénye —, lényegében annak az eseménynek a valószínűségét reprezentálja, hogy az  $\alpha$  paramétervektor az  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$  értéket veszi fel, rögzített  $y = (y_1, y_2, \dots, y_m)$  mérési adatok esetén. A ML módszer alkalmazása során tehát az  $\alpha \mapsto L(\alpha)$  likelihood függvény maximumhelyét keressük. Mint említettük gyakran célszerűbb, egyszerűbb az  $\alpha \mapsto \ln L(\alpha)$  log-likelihood függvény maximumhelyét meghatározni, különösen akkor, ha a likelihood függvényt szorzat alakban írhatjuk. Ha ugyanis a mérési adatok függetlenek, akkor (2.4) szerint

$$(2.6) \quad L(\alpha) = \prod_{i=1}^m f_i(y_i | \alpha) = \prod_{i=1}^m f_i(y_i | \alpha_1, \alpha_2, \dots, \alpha_k)$$

és ekkor a log-likelihood függvény alakja:

$$(2.7) \quad \ln L(\alpha) = \sum_{i=1}^m \ln f_i(y_i | \alpha) = \sum_{i=1}^m \ln f_i(y_i | \alpha_1, \alpha_2, \dots, \alpha_k)$$

A szélsőérték létezésének szükséges feltétele, hogy az  $\alpha \in \mathbb{R}^k$  paraméter komponensei kielégítsék az ún. „likelihood egyenleteket”:

$$(2.8) \quad \frac{\partial \ln L(\alpha_1, \alpha_2, \dots, \alpha_k)}{\partial \alpha_s} = 0; \quad (s = 1, 2, \dots, k)$$

A gyakorlatban, amikor a modell nagyon sok paramétert tartalmaz, a maximum-feladat megoldása nem állítható elő zárt analitikus formában. A szélsőérték hely ekkor numerikus módszerekkel állítandó elő, nemlineáris optimalizálási eljárások alkalmazásával, iterációs módszerekkel. Célszerű ezen algoritmusok során az elvileg legbővebb  $\mathbb{R}^k$  paramétertartományt leszűkíteni és az optimumot az  $\mathbb{R}^k$  egy részhalmazán megkeresni, mert ez gyorsítja az iterációt. Ebben az esetben azonban fennáll a veszélye annak, hogy csak lokális maximumot állít elő az algoritmus és nem a likelihood függvény abszolút maximumát, mégpedig akkor, ha nem megfelelő paramétertartományt választunk. Célszerű ezért több különböző módszer és paramétertartomány esetén is meghatározni a maximumot, és ha minden esetben ugyanazt kapjuk optimális megoldásnak.

dásként, akkor fogadhatjuk el a kapott vektort a maximum likelihood módszer megoldásaként.

## 2. A maximum likelihood elv alkalmazása Poisson-eloszlás paramétereinek becslésére

Térjünk rá a gamma-sugárzás spektrumának vizsgálatára. Alkalmazzuk az 1. pontban bevezetett jelöléseket. Tegyük fel, hogy az  $i$ -edik  $[E_{i-1}, E_i]$  csatornában a regisztrált beütésszám  $y_i$ , míg az alkalmazott modell szerint a fotonok számának átlagértéke  $\lambda_i$  ( $i = 1, 2, \dots, m$ ). Teljesen általánosan, a modell szerinti  $\lambda_i$  várható értékek leírására szolgáló paraméterek legyenek  $\alpha_1, \alpha_2, \dots, \alpha_k$ . Tegyük fel tehát, hogy mindegyik  $\lambda_i$  az  $\alpha_s$  ( $s = 1, 2, \dots, k$ ) paraméterek függvénye:  $\lambda_i = \lambda_i(\alpha_1, \alpha_2, \dots, \alpha_k)$ . Jelölje  $P(y_i, \lambda_i)$  annak az eseménynek a valószínűségét, hogy az  $i$ -edik csatornában a regisztrált fotonszám  $y_i$ , feltéve, hogy a modell szerinti beütésszám várható értéke  $\lambda_i$  ( $i = 1, 2, \dots, m$ ). Ez utóbbi nagy számú mérés eredményeinek átlagolásával számítható. Mivel az egyes csatornában a beütések száma egymástól nyilvánvalóan független, a likelihood függvény az alábbi alakban írható:

$$(3.1) \quad L(\alpha_1, \alpha_2, \dots, \alpha_k) = \prod_{i=1}^m P(y_i, \lambda_i)$$

A tapasztalatok szerint, ha radioaktív bomlásról van szó, a számlálóberendezéssel regisztrált események száma, tehát a beütésszám Poisson-eloszlást követ, melynek várható értéke a fent bevezetett jelölés szerint éppen  $\lambda_i$ . Ebből következik, hogy

$$P(y_i, \lambda_i) = \frac{\lambda_i^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i)$$

ezért a (3.1) likelihood függvény a következő konkrét alakot ölti:

$$(3.2) \quad L(\alpha_1, \alpha_2, \dots, \alpha_k) = \prod_{i=1}^m \frac{\lambda_i^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i)$$

Képezzük a kapott függvény logaritmusát. Így kapjuk a log-likelihood függvényt:

$$\begin{aligned} \ln L(\alpha_1, \alpha_2, \dots, \alpha_k) &= \sum_{i=1}^m \ln \left( \frac{\lambda_i^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i) \right) = \sum_{i=1}^m (\ln(\lambda_i)^{y_i} + \ln(\exp(-\lambda_i)) - \ln(y_i)!) = \\ &= \sum_{i=1}^m (y_i \ln(\lambda_i) - \lambda_i - \ln(y_i)!) = \sum_{i=1}^m (y_i \ln(\lambda_i) - \lambda_i) + \text{állandó} \end{aligned}$$

A maximum likelihood becslés az  $\ln L(\alpha_1, \dots, \alpha_k)$  függvény  $\alpha_s$  paramétereinek szerinti deriválásával állítható elő:

$$\frac{\partial \ln L(\alpha_1, \alpha_2, \dots, \alpha_k)}{\partial \alpha_s} = \frac{\partial}{\partial \alpha_s} \left\{ \sum_{i=1}^m (y_i \ln(\lambda_i(\alpha_1, \dots, \alpha_k)) - \lambda_i(\alpha_1, \dots, \alpha_k)) \right\} = 0$$

( $s = 1, 2, \dots, k$ ). A konkrét számítások akkor végezhetőek el, ha az  $\lambda_i = \lambda_i(\alpha_1, \alpha_2, \dots, \alpha_k)$  várható értékekre konkrét matematikai modell van a birtokunkban. Első lépésként tegyük fel, hogy egyetlen radioaktív izotóp gamma sugárzását vizsgáljuk. Élünk azzal a nyilvánvaló feltevessel, hogy az  $i$ -edik csatornában a beütésszám átlagértéke,  $\lambda_i$ , arányos az izotóp  $A$  aktivitásával [1, 2]. Ebben az esetben tehát egyetlen  $\alpha$  paraméterünk van és ez az ismeretlen  $A$  aktivitás. Ha  $f_i$  jelöli az egységnyi időre és egységnyi aktivitásra vonatkozó spektrumot ( $i = 1, 2, \dots, m$ ), akkor az átlagértékre a  $\lambda_i = T \cdot A \cdot f_i$  összefüggés adódik, ahol  $T$  a spektrum felvételének teljes időtartama. Ekkor a (3.2) likelihood függvény az alábbi formában írható:

$$(3.3) \quad L(A) = \prod_{i=1}^m \frac{(T \cdot A \cdot f_i)^{y_i}}{(y_i)!} \cdot \exp(-T \cdot A \cdot f_i)$$

Így a log-likelihood függvény:

$$\ln L(A) = \sum_{i=1}^m (y_i \ln(T \cdot A \cdot f_i) - \ln(y_i)!) - T \cdot A \cdot f_i$$

A szélsőérték pedig a

$$\frac{\partial \ln L(A)}{\partial A} = \sum_{i=1}^m \left( y_i \frac{1}{A} - T \cdot f_i \right) = \frac{1}{A} \sum_{i=1}^m y_i - T \cdot \sum_{i=1}^m f_i = 0$$

egyenlet megoldásaként adódik:

$$(3.4) \quad A = \frac{\sum_{i=1}^m y_i}{T \cdot \sum_{i=1}^m f_i}$$

A (3.4) formula szolgáltatja tehát az ismeretlen aktivitás maximum likelihood becslését egyetlen izotóp esetén. Képezve a másodrendű deriváltat, az adódik, hogy:

$$\frac{\partial^2 \ln L(A)}{\partial A^2} = -\frac{1}{A^2} \sum_{i=1}^m y_i$$

Ez nyilvánvalóan negatív, ami pontosan azt jelenti, hogy a (3.4) megoldás a likelihood függvény maximumát szolgáltatja.

Általánosítsuk a fent megoldott problémát arra az esetre, amikor  $k$  db radioaktív izotóp keverékének spektrumát vesszük fel, továbbá vegyük tekintetbe  $(k+1)$ -edik komponensként a háttérrel is. Tegyük fel, hogy az  $s$  indexű izotóp aktivitása  $A_s$ . Ekkor az  $i$ -edik csatornában a beütésszám várható értéke:

$$\lambda_i = T \cdot \sum_{s=1}^{k+1} A_s f_{si}$$

ahol  $f_{si}$  az  $s$ -indexű izotóp egységnyi időre és egységnyi aktivitásra vonatkozó spektruma. A likelihood függvény most  $k + 1$  db paramétert tartalmaz:

$$(3.5) \quad L(A_1, \dots, A_k, A_{k+1}) = \prod_{i=1}^m \frac{\left( T \cdot \sum_{s=1}^{k+1} A_s \cdot f_{si} \right)^{y_i}}{(y_i)!} \cdot \exp\left( -T \cdot \sum_{s=1}^{k+1} A_s \cdot f_{si} \right)$$

Ennek logaritmus:

$$\ln L(A_1, \dots, A_k, A_{k+1}) = \sum_{i=1}^m \left( y_i \left( \ln \left[ T \cdot \sum_{s=1}^{k+1} A_s \cdot f_{si} \right] \right) - T \cdot \sum_{s=1}^{k+1} A_s \cdot f_{si} - \ln(y_i)! \right)$$



Szélsőérték ott lehet, ahol teljesülnek a

$$\frac{\partial \ln L(A_1, \dots, A_k, A_{k+1})}{\partial A_s} = 0; (s = 1, 2, \dots, k, k+1)$$

egyenletek. Elvégezve a deriválást, azt kapjuk, hogy

$$\frac{\partial \ln L(A_1, \dots, A_k, A_{k+1})}{\partial A_r} = \sum_{i=1}^m \left( \frac{y_i \cdot f_{ri}}{\sum_{s=1}^{k+1} A_s \cdot f_{si}} - T \cdot f_{ri} \right) = 0;$$

(r = 1, 2, \dots, k, k+1)

A kapott egyenletek átrendezésével kapjuk az egyes izotópok  $A_s$  aktivitásának maximum likelihood becslésére vonatkozó lineáris egyenletrendszert:

$$(3.6) \quad \sum_{s=1}^{k+1} T \cdot A_s \left( \sum_{i=1}^m f_{ri} \cdot f_{si} \right) = \sum_{i=1}^m y_i \cdot f_{ri}; (r = 1, 2, \dots, k, k+1)$$

Ez tehát egy  $k+1$  db egyenletből álló egyenletrendszer az ismeretlen  $A_s$  aktivitásokra vonatkozólag. A rendszer numerikus módszerekkel megoldható [2].

### 3. Közelítés normális eloszlással

Vizsgáljuk most azt az általánosabb esetet, amikor a mérési adatok nem írhatók le pozitív egész számokkal. Ekkor nem alkalmazható a Poisson-eloszlással történő leírás, azonban közelíthetünk Gauss-eloszlással. Ha a szokásos módon  $m_i$  jelöli az átlagértéket és  $\sigma_i$  a szórást, akkor a Gauss-eloszlás sűrűségfüggvénye:

$$(4.1) \quad f(y_i, m_i, \sigma_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(y_i - m_i)^2}{2\sigma_i^2}\right); (i = 1, 2, \dots, m)$$

Bár a normális eloszlás nem alkalmazható közvetlenül egy számláló berendezés által mért értékek, mint valószínűségeloszlás leírására, mégis azt mondhatjuk, hogy a Gauss-eloszlás a Poisson-eloszlás elfogadható közelítésének számít, ha a beütésszám csatornánként eléri legalább az 5 értéket. Ekkor a Poisson-eloszlás a normális eloszlás egész helyeken

felvett értékeivel közelíthető. A normálissal való közelítés kerülendő, ha az alacsony beütésszámok gyakoriak [3].

A gyakorlatban és az elméleti megfontolások során is feltételezik, hogy a  $\sigma_i$  szórások ismertek és függetlenek a modell paramétereitől [1]. Ez azt jelenti, hogy csak a várható értékek függenek a paramétereiktől:  $m_i = m_i(\alpha_1, \dots, \alpha_k)$ . Ha tehát a normális eloszlással történő közelítés feltételei fennállnak, akkor a (3.1) likelihood függvény alakja a (4.1) sűrűségfüggvény és a paraméterekkel kapcsolatos megállapodás figyelembevételével a következő:

$$(4.2) \quad L(\alpha_1, \alpha_2, \dots, \alpha_k) = \frac{1}{(\sqrt{2\pi})^m} \prod_{i=1}^m \frac{1}{\sigma_i} \cdot \exp\left(-\frac{(y_i - m_i(\alpha_1, \alpha_2, \dots, \alpha_k))^2}{2\sigma_i^2}\right)$$

Képezve (4.2) logaritmusát, kapjuk a log-likelihood függvényt:

$$\ln L(\alpha_1, \alpha_2, \dots, \alpha_k) = \sum_{i=1}^m \left( -\frac{(y_i - m_i(\alpha_1, \alpha_2, \dots, \alpha_k))^2}{2\sigma_i^2} - \ln \sigma_i \right) - m \cdot \ln \sqrt{2\pi}$$

A jelölések egyszerűsítése végett célszerű a kapott egyenletet megszorozni (-2)-vel:

$$\begin{aligned} -2 \ln L(\alpha_1, \alpha_2, \dots, \alpha_k) &= \sum_{i=1}^m \left( \frac{(y_i - m_i(\alpha_1, \alpha_2, \dots, \alpha_k))^2}{\sigma_i^2} + \ln \sigma_i^2 \right) + m \cdot \ln 2\pi = \\ &= \sum_{i=1}^m \frac{(y_i - m_i(\alpha_1, \alpha_2, \dots, \alpha_k))^2}{\sigma_i^2} + \left( \sum_{i=1}^m \ln \sigma_i^2 + m \cdot \ln 2\pi \right) \end{aligned}$$

Vegyük észre, hogy az összeg első szummája éppen a mért spektrum és a modellspektrum eltérését megadó  $\chi^2$  statisztika, a második, zárójelben levő összeg pedig konstans, független a paramétereiktől. Azt kaptuk, tehát, hogy ha normális eloszlással közelítünk és a  $\sigma_i$  szórásokat a paramétereiktől független ismert konstansoknak tekintjük, akkor:

$$-2 \ln L(\alpha_1, \alpha_2, \dots, \alpha_k) = \chi^2 + \text{állandó}$$

Az említett feltevések mellett igaz tehát, hogy a maximum likelihood becslés, és a legkisebb négyzetek módszere – azaz a  $\chi^2$  statisztika –, ugyanazt az eredményt szolgáltatják. Számláló berendezésekkel végzett mérések esetében azonban ezek a feltevések nem teljesülnek, de a fenti eredmény módosítható a vizsgált esetre. Ekkor ugyanis — a Poisson-eloszlásra való tekintettel —, pontosan teljesülnie kell annak, hogy a szórásnégyzet egyenlő a várható értékkel, azaz  $\sigma_i^2 = m_i$ . Ekkor tehát a

Gauss-féle log-likelihood függvény konstanstól eltekintve nem egyenlő a  $\chi^2$  statisztikával, hanem tartalmaz még egy additív, paramétereiktől függő tagot:

(4.4)

$$\begin{aligned} -2\ln L(\alpha_1, \alpha_2, \dots, \alpha_k) &= \sum_{i=1}^m \frac{(y_i - m_i(\alpha_1, \alpha_2, \dots, \alpha_k))^2}{m_i(\alpha_1, \alpha_2, \dots, \alpha_k)} + \sum_{i=1}^m \ln m_i(\alpha_1, \alpha_2, \dots, \alpha_k) + \text{állandó} \\ &= \chi_P^2 + \sum_{i=1}^m \ln m_i(\alpha_1, \alpha_2, \dots, \alpha_k) + \text{állandó} \end{aligned}$$

A  $\chi^2$  statisztika ezen módosítása Pearson-tól származik [3], erre utal a jelölésben alkalmazott „P” index. A paraméterek ML becslése a (2.8) feltételi egyenletek alkalmazásával elvileg ennek a formulának a felhasználásával is adódik, ha rendelkezünk konkrét modellfeltevésekkel a várható értékre vonatkozólag.

Vizsgáljuk most meg a ML hatékonyságának problémáját. Tisztán matematikai szempontból a  $\chi^2$  statisztika a  $v$  szabadságfokú khinégyzet eloszlást követi, ha az összeg tagjai páronként független standard normális eloszlású valószínűségi változók. A sűrűségfüggvény ekkor

$$(4.5) \quad f_v(x) = \frac{\left(\frac{1}{2}\right)^v}{\Gamma\left(\frac{v}{2}\right)} \cdot x^{\frac{v}{2}-1} \cdot \exp\left(-\frac{v}{2}\right)$$

ha  $x > 0$  és  $f_v(x) = 0$  ha  $x \leq 0$ . Itt  $\Gamma(x)$  az Euler-féle gammafüggvényt jelöli:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} \exp(-t) dt$$

(A gammafüggvény legismertebb tulajdonsága, hogy  $\Gamma(x) = (x-1) \cdot \Gamma(x-1)$ , tehát  $\Gamma(n) = (n-1)!$  ha  $n$  egész szám!) Ha a mérési adatok száma  $m$ , és az illesztett paraméterek száma  $k$ , akkor az eloszlás szabadságfoka:  $v = m - k$ . Mint azt már korábban említettük, számláló berendezésekkel végzett mérésekre vonatkozólag a felsorolt feltételek csak közelítőleg teljesülnek. Ezért nem árt megvizsgálni, hogy a megfigyelésekből számított  $\chi^2$  statisztika értéke milyen konfidencia intervallumba esik. Hibát követhetünk el ugyanis azáltal, hogy bizonyos paramétereket állandónak tekintünk, de lehet probléma eleve a modell feltevéseivel. Legyen  $\chi_{\min}^2$

a mérési adatok és modellparaméterek alapján számított aktuális értéke a  $\chi^2$  statisztikának. A konfidencia intervallum kijelöléséhez definiáljuk a

$$P(\chi_{\min}^2, v) = \int_{\chi_{\min}^2}^{\infty} f_v(x) dx$$

valószínűséget, ahol  $f_v(x)$  a  $v$  szabadságfokú khínégyzet eloszlás (4.5) sűrűségfüggvénye. Ha a számítások alapján az adódik, hogy  $P > 0,1$  akkor a modellt elfogadjuk, ha azonban  $P < 0,001$ , akkor a modellt elvetjük, mert nagy valószínűséggel feltevései nem helytállóak. A  $P$  értéket nevezzük az illeszkedés megbízhatóságának.

A megbízhatóság kritériumát Gauss-eloszlás feltételezése mellett a  $\chi^2$  statisztika alapján vezettük le. Felmerül a kérdés, adható-e valamilyen kritérium az illeszkedés megbízhatóságára vonatkozólag, közvetlenül a likelihood függvénnyel megfogalmazva? A válasz pozitív. Definiáljuk ennek érdekében a „likelihood hányadost” az alábbi módon:

$$(4.6) \quad \lambda = \max_{\alpha_i} \lambda' = \frac{\max L(m(\alpha_1, \dots, \alpha_k))}{\max L(m)}$$

A tört számlálójában az adott  $y \in \mathbb{R}^m$  mérési adatokhoz tartozó likelihood függvény — modell szerinti paramétertartományra vonatkozó —, paraméterek szerinti maximuma áll. A nevezőben pedig likelihood függvény abszolút maximuma szerepel, amit úgy kapunk, hogy a paraméterek tekintetében, a modellre vonatkozólag semmiféle megszorító feltevessel nem élünk. Világos, hogy  $\lambda$  és  $\lambda'$  egyaránt a  $[0, 1]$  intervallumban vannak. A modellben alkalmazott paraméterek legmegfelelőbb értékei a  $\lambda'$  hányados paraméterek szerinti maximalizálásával kaphatók. A nevezőbeli abszolút maximum pedig a

$$(4.7) \quad \frac{\partial}{\partial m_i} (-2 \ln L(m)) = 0$$

feltételekből adódnak, ahol az  $m_i$  modellváltozókat ebben a lépésben függetlennek tekintjük a paraméterektől. Definiáljuk most a „maximum likelihood  $\chi^2$ ” statisztikát a következő módon:

$$(4.8) \quad \chi_{\lambda}^2 = -2 \cdot \ln \lambda' = -2 \cdot \ln L(m(\alpha_1, \dots, \alpha_k)) + 2 \cdot \ln L(m)$$

Igazolható, hogy a  $\min(\chi_{\lambda}^2)$  aszimptotikusan megegyezik a klasszikus  $m - k$  szabadságfokú  $\chi^2$  eloszlással [3]. Eszerint a megbízhatóságra vonatkozó kritérium közvetlenül ebből a formulából is nyerhető. Mivel  $\chi_{\lambda}^2$

második tagja független a paramétereiktől, ezért a  $\chi^2$  paraméterek szerinti minimalizálása ekvivalens az  $L(m(\alpha_1, \dots, \alpha_k))$  likelihood függvény maximalizálásával.

Tanulságos a  $\chi^2$  alkalmazása a korábbiakban már vizsgált Poisson-eloszlás esetére. A becült  $m_i$  modellparaméter most a  $\lambda_i$  várható érték. Ha elsőként azt feltételezzük, hogy ez nem függ paramétereiktől, akkor a (4.7) feltétel a következő eredményt adja:

$$\begin{aligned} \frac{\partial}{\partial \lambda_i} (-2 \cdot \ln L(\lambda_1, \dots, \lambda_m)) &= \frac{\partial}{\partial \lambda_i} \left( -2 \cdot \ln \prod_{i=1}^m \frac{\lambda_i^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i) \right) = \\ &= \frac{\partial}{\partial \lambda_i} \left( -2 \cdot \sum_{i=1}^m (y_i \ln(\lambda_i) - \lambda_i - \ln(y_i)!) \right) = -2 \cdot \left( \frac{y_i}{\lambda_i} - 1 \right) = 0 \end{aligned}$$

Ami azt jelenti, hogy a várható értékekre a  $\lambda_i = y_i$  becslés adódik, ez szolgáltatja a likelihood függvény abszolút maximumát. Ennek alapján Poisson-eloszlás esetén a  $\chi^2$  statisztikára az alábbi kifejezést kapjuk:  $\chi^2 = 2 \cdot \ln L(m) + 2 \cdot \ln L(m(\alpha_1, \dots, \alpha_k)) =$

$$\begin{aligned} &= 2 \cdot \ln \prod_{i=1}^m \frac{y_i^{y_i}}{(y_i)!} \cdot \exp(-y_i) - 2 \cdot \ln \prod_{i=1}^m \frac{\lambda_i^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i) = \\ (4.9) \quad &= 2 \left( \sum_{i=1}^m (y_i \ln(y_i) - y_i - \ln(y_i)!) - \sum_{i=1}^m (y_i \ln(\lambda_i) - \lambda_i - \ln(y_i)!) \right) = \\ &= 2 \cdot \sum_{i=1}^m \left( (\lambda_i - y_i) - y_i \ln \frac{\lambda_i}{y_i} \right) \end{aligned}$$

A maximum likelihood módszer tehát Poisson-eloszlás esetén arra vezetett, hogy a likelihood függvény maximumának meghatározása egyenértékű a  $\chi^2$  statisztika (4.9) alakjának minimalizálásával.

Felhívjuk a figyelmet arra, hogy ha (4.9)-ben eltekintünk a konstans szorzótól és a kapott kifejezés ellentettjét vesszük, akkor éppen az entropia

$$S(y_i, \lambda_i) = \sum_{i=1}^m \left( (y_i - \lambda_i) - y_i \ln \frac{y_i}{\lambda_i} \right)$$

általános alakját kapjuk. Ennek maximalizálása egyenértékű  $\chi^2$  minimalizálásával. Megmutattuk tehát, hogy a gamma sugárzás spektrumának

Poisson-eloszlással történő modellezése esetén a maximum likelihood elv és a maximum entrópia módszer egyenértékűek [4, 5, 6].

#### 4. A várható érték maximalizálásának elve

A ML becslés alkalmazása esetén a probléma optimális megoldását közvetlenül a

$$\frac{\partial \ln L(\alpha_1, \alpha_2, \dots, \alpha_k)}{\partial \alpha_s} = 0; (s = 1, 2, \dots, k)$$

egyenletek által meghatározott rendszer megoldása szolgáltatja. Hogy a kapott stacionárius pont valóban abszolút maximum és nem csak lokális maximum, vagy esetleg minimum, külön megfontolások tárgyát képezi. Lineáris és nemlineáris egyenletrendszerek megoldására több különböző eljárás is ismert, tehát a probléma elvileg megoldható. Ilyennel találkoztunk (3.6)-ban.

Létezik azonban egy olyan iteratív eljárás is, amely a likelihood függvény maximumát közvetlenül a Bayes-formula alkalmazásával, tehát valószínűségelméleti alapon állítja elő. Ennek során közvetlenül a  $P(y | \alpha)$  feltételes valószínűség logaritmusát, tehát a log-likelihood függvényt maximalizáljuk.

Az eljárás, a „várható érték maximalizálásának elve” nevet viseli (Expectation Maximization Algorithm, EM). Mivel azonban ez a módszer szintén a likelihood függvény maximumát állítja elő, nem független a ML elvtől, hanem ellenkezőleg, arra épül. Ezért összefoglalóan úgy nevezhetjük, hogy „maximum likelihood becslés a várható érték maximalizálásának módszerével” (Maximum Likelihood estimation using Expectation Maximization Algorithm, ML-EM) [7].

Az alábbiakban ismertető eljárás számítások szempontjából legnagyobb előnye, hogy alkalmazása akkor is lehetséges, ha hiányzó vagy közvetlenül nem megfigyelhető paraméterek is bonyolítják a helyzetet.

A módszer megvilágítása érdekében tegyük fel, mint korábban, hogy  $y \in \mathbb{R}^m$  a mérési adatok véletlentől függő vektora, valószínűségi változó. Legyen továbbá  $\alpha \in \mathbb{R}^k$  az alkalmazott modellnek megfelelő paraméterek vektora. Feladatunk az  $\alpha_s$  ( $s = 1, 2, \dots, k$ ) paramétereket úgy meghatározni, hogy a  $P(y | \alpha)$  valószínűség, illetve ennek logaritmusa az  $L(\alpha) = \ln P(y | \alpha)$  log-likelihood függvény maximális legyen. Az EM

algorithmus iterációs úton állítja elő az  $L(\alpha)$  függvény maximumát. Tegyük fel, hogy a  $k$ -adik lépésben a paraméterek optimális becslése:  $\alpha^{(k)}$ . Mivel a feladat  $L(\alpha)$  maximalizálása, teljesülnie kell, hogy  $L(\alpha) > L(\alpha^{(k)})$ . A kitűzött problémával nyilván egyenértékű az

$$(5.1) \quad L(\alpha) - L(\alpha^{(k)}) = \ln P(y | \alpha) - \ln P(y | \alpha^{(k)})$$

különbség maximalizálása.

Mindezidáig nem vettünk figyelembe hiányzó, illetve közvetlenül nem megfigyelhető, vagy nem megfigyelt adatokat, paramétereket. Ha léteznek ilyenek, az EM algoritmus egy természetes keretet szolgáltat ezen paraméterek figyelembe vételére.

Megfordítva a logikát, a maximum likelihood becslés során mesterségesen is bevezethetünk „rejtett paramétereket” annak érdekében, hogy a számítás egyszerűbben kivitelezhető legyen.

Jelölje ezen paraméterek vektorát  $z = (z_1, z_2, \dots, z_R)$ , amelyet ugyancsak valószínűségi változónak tekintünk. A  $P(y | \alpha)$  valószínűség — a  $z$  paraméterváltozó figyelembevételével —, a teljes valószínűség tétele alapján számítható ki:

$$P(y|\alpha) = \sum_r P(y|z_r, \alpha) \cdot P(z_r|\alpha)$$

Az (5.1) egyenlet így az

$$L(\alpha) - L(\alpha^{(k)}) = \ln \sum_r P(y|z_r, \alpha) \cdot P(z_r|\alpha) - \ln P(y | \alpha^{(k)})$$

alakot ölti. Alkalmazzuk most a konkáv függvényekre vonatkozó

$$\sum_{r=1}^R \beta_r f(x_r) \leq f\left(\sum_{r=1}^R \beta_r x_r\right); \quad \beta_r \geq 0, \quad \sum_{r=1}^R \beta_r = 1$$

Jensen-egyenlőtlenséget a logaritmusfüggvényre. Ha alkalmazzuk a  $\beta_r = P(z_r | y, \alpha^{(k)}) \geq 0$  jelölést, és a nyilvánvalóan igaz

$$\sum_r P(z_r | y, \alpha^{(k)}) = 1$$

összefüggést, akkor a Jensen-egyenlőtlenség alkalmazása és azonos átalakítások után a következőt kapjuk:

$$\begin{aligned}
 L(\alpha) - L(\alpha^{(k)}) &= \ln \left( \sum_r P(y|z_r, \alpha) \cdot P(z_r|\alpha) \right) - \ln P(y|\alpha^{(k)}) = \\
 &= \ln \left( \sum_r P(y|z_r, \alpha) \cdot P(z_r|\alpha) \cdot \frac{P(z_r|y, \alpha^{(k)})}{P(z_r|y, \alpha^{(k)})} \right) - \ln P(y|\alpha^{(k)}) = \\
 &= \ln \left( \sum_r P(z_r|y, \alpha^{(k)}) \cdot \frac{P(y|z_r, \alpha) \cdot P(z_r|\alpha)}{P(z_r|y, \alpha^{(k)})} \right) - \ln P(y|\alpha^{(k)}) \geq \\
 &\geq \sum_r P(z_r|y, \alpha^{(k)}) \cdot \ln \frac{P(y|z_r, \alpha) \cdot P(z_r|\alpha)}{P(z_r|y, \alpha^{(k)})} - \ln P(y|\alpha^{(k)}) = \\
 &= \sum_r P(z_r|y, \alpha^{(k)}) \cdot \ln \frac{P(y|z_r, \alpha) \cdot P(z_r|\alpha)}{P(z_r|y, \alpha^{(k)}) \cdot P(y|\alpha^{(k)})} = \Delta(\alpha|\alpha^{(k)})
 \end{aligned}$$

Az egyenlőtlenséglánc végén bevezetett jelöléssel azt kaptuk tehát, hogy

$$(5.2) \quad L(\alpha) \geq L(\alpha^{(k)}) + \Delta(\alpha|\alpha^{(k)}) = f(\alpha|\alpha^{(k)})$$

Az  $f(\alpha|\alpha^{(k)})$  függvény tehát nem haladhatja meg a likelihood függvényt. Egyenlőség pedig éppen az  $\alpha^{(k)}$  helyen teljesül, ugyanis:

$$\begin{aligned}
 f(\alpha^{(k)}|\alpha^{(k)}) &= L(\alpha^{(k)}) + \Delta(\alpha^{(k)}|\alpha^{(k)}) = \\
 &= L(\alpha^{(k)}) + \sum_r P(z_r|y, \alpha^{(k)}) \cdot \ln \frac{P(y|z_r, \alpha^{(k)}) \cdot P(z_r|\alpha^{(k)})}{P(z_r|y, \alpha^{(k)}) \cdot P(y|\alpha^{(k)})} = \\
 (5.3) \quad &= L(\alpha^{(k)}) + \sum_r P(z_r|y, \alpha^{(k)}) \cdot \ln \frac{P(y, z_r|\alpha^{(k)})}{P(z_r, y|\alpha^{(k)})} = \\
 &= L(\alpha^{(k)}) + \sum_r P(z_r|y, \alpha^{(k)}) \cdot \ln 1 = L(\alpha^{(k)})
 \end{aligned}$$



Tehát a  $k$ -edik iterációs lépés eredményeként adódó  $\alpha^{(k)}$  helyen teljesül, hogy  $L(\alpha) = f(\alpha | \alpha^{(k)})$ . Ennek alapján nyilvánvaló, hogy bármely olyan  $\alpha$  választása esetén, amely megnöveli az  $f(\alpha | \alpha^{(k)})$  függvényt, növekszik az  $L(\alpha)$  függvény értéke is. Annak érdekében, hogy  $L(\alpha)$  értékét tekintve a lehető legnagyobb növekedést érjük el az EM algoritmus alkalmazása keretében, világos, hogy az  $f(\alpha | \alpha^{(k)})$  függvény maximumát kell meghatározni. Legyen a  $(k+1)$ -edik iterációs lépésben  $\alpha$  becslése pontosan az  $f(\alpha | \alpha^{(k)})$  függvény maximumhelye:  $\alpha^{(k+1)}$ . Ha a számítások során eltekintünk a konstans tagoktól és alkalmazzuk a feltételes valószínűség definícióját, akkor  $\alpha$   $(k+1)$ -edik közelítésére a következő formulát kapjuk:

(5.4)

$$\begin{aligned} \alpha^{(k+1)} &= \max_{\alpha} \left\{ L(\alpha^{(k)}) + \sum_r P(z_r | y, \alpha^{(k)}) \cdot \ln \frac{P(y|z_r, \alpha) \cdot P(z_r | \alpha)}{P(z_r | y, \alpha^{(k)}) \cdot P(y | \alpha^{(k)})} \right\} = \\ &= \max_{\alpha} \left\{ \sum_r P(z_r | y, \alpha^{(k)}) \cdot \ln P(y|z_r, \alpha) \cdot P(z_r | \alpha) \right\} = \\ &= \max_{\alpha} \left\{ \sum_r P(z_r | y, \alpha^{(k)}) \cdot \ln \frac{P(y, z_r, \alpha)}{P(z_r, \alpha)} \cdot \frac{P(z_r, \alpha)}{P(\alpha)} \right\} = \\ &= \max_{\alpha} \left\{ \sum_r P(z_r | y, \alpha^{(k)}) \cdot \ln P(y, z_r | \alpha) \right\} = \\ &= \max_{\alpha} E[\ln P(y, z_r | \alpha)] \end{aligned}$$

Az adódik tehát, hogy az  $\alpha^{(k+1)}$  közelítést az  $E[\ln P(y, z | \alpha)]$  feltételes várható érték  $\alpha$  paraméter szerinti maximuma szolgáltatja.

A fentiek szerint az ME algoritmus két lépésből áll:

1. lépés: („E-step”, expectation-step) Az  $\alpha^{(k)}$   $k$ -edik közelítés birtokában meg kell határoznunk az

$$(5.5) \quad E[\ln P(y, z_r | \alpha)] = \sum_r P(z_r | y, \alpha^{(k)}) \cdot \ln P(y, z_r | \alpha)$$

feltételes várható értéket.

2. lépés: („M-step”, maximization-step) Maximalizálni kell a várható értéket az  $\alpha$  paraméterek szerint. A maximumhely szolgáltatja a  $(k+1)$ -edik  $\alpha^{(k+1)}$  közelítést.

Ezek után felváltva alkalmazzuk az 1. és 2. lépéseket. A kifejtett algoritmus természetesen lehetőséget ad minden egyes iterációs lépésben a „rejtett paraméterek”,  $z$  becslésére is. Mivel a  $z$  értékeit közvetlenül megmérni, megfigyelni nem tudjuk, természetesnek mondható, hogy a  $(k+1)$ -edik lépésben  $z$  értékét azonosítjuk a

$$(5.6) \quad z^{(k+1)} = E[z \mid y, \alpha^{(k)}]$$

feltételes várható értékkel. A leírtak fényében úgy dolgozunk, hogy az iteráció során felvesszük kezdőértékként az  $\alpha^{(0)}$  paramétert, majd az (5.6) és (5.4) formulák ismételt alkalmazásával rendre meghatározzuk  $z$  és  $\alpha$  aktuális értékét:

$$\alpha^{(0)} \rightarrow z^{(1)} \rightarrow \alpha^{(1)} \rightarrow z^{(2)} \rightarrow \alpha^{(2)} \rightarrow z^{(3)} \rightarrow \alpha^{(3)} \rightarrow z^{(4)} \rightarrow \dots$$

Ismételten kiemeljük, hogy ha az  $L(\alpha)$  likelihood függvény közvetlen maximalizálása helyett az  $f(\alpha \mid \alpha^{(k)})$  függvényt maximalizáljuk egy iterációs eljárás minden lépésében, akkor az azzal az előnnyel jár, hogy figyelembe tudjuk venni a hiányzó paramétereket, a nem megfigyelt adatokat, amelyeket a  $z$  vektorban foglaltunk össze. Abban az esetben, ha becslést kívánunk adni ezekre a paraméterekre, a fenti algoritmus lehetőséget ad rá. A 6. pontban részletesen megvizsgálunk egy ilyen esetet. Mint azt a bevezetőben már említettük, ilyen paraméterek mesterséges bevezetése gyakran haszonnal is jár, mert megkönnyíti a számítások elvégzését.

Az algoritmus abban az esetben is működik, ha az  $f(\alpha \mid \alpha^{(k)})$  függvény egzakt maximumhelyének megállapítása nehézségekbe ütközik, vagy akár kivitelezhetetlen. Ebben az esetben az  $f(\alpha \mid \alpha^{(k)})$  maximumának megállapítása, tehát a függvényérték legnagyobb mértékű növelése helyett elegendő egy olyan  $\alpha^{(k+1)}$  becslés meghatározása, amely az  $f(\alpha \mid \alpha^{(k)})$  függvény értékét „egyszerűen csak növeli”, de nem a legnagyobb mértékben. Ebben az esetben is teljesül tehát, hogy

$$(5.7) \quad f(\alpha^{(k)} \mid \alpha^{(k)}) \leq f(\alpha^{(k+1)} \mid \alpha^{(k)})$$

(Természetesen lehetőség szerint megköveteljük a szigorú egyenlőtlenséget.) Ez az eljárás az EM algoritmus általánosításának tekintendő (Generalized EM Algorithm, GEM). Felmerül az iteráció konvergenciájának kérdése. Megmutattuk (5.3)-ban, hogy  $\Delta(\alpha^{(k)} \mid \alpha^{(k)}) = 0$ . Vizsgál-

jük ennek felhasználásával a  $(k+1)$ -edik iterációs lépésben  $L(\alpha)$  értékét. (5.2) és (5.7) szerint:

$$L(\alpha^{(k+1)}) \geq L(\alpha^{(k)}) + \Delta(\alpha^{(k+1)} | \alpha^{(k)}) = f(\alpha^{(k+1)} | \alpha^{(k)}) \geq f(\alpha^{(k)} | \alpha^{(k)}) = L(\alpha^{(k)}) + \Delta(\alpha^{(k)} | \alpha^{(k)}) = L(\alpha^{(k)})$$

Ami azt jelenti, hogy az  $L(\alpha^{(k)})$  ( $k \in \mathbb{N}$ ) sorozat monoton növekedő. Világos, hogy a sorozatnak felső korlátja  $\ln P(y | \alpha)$ . Ez azt jelenti, hogy az  $L(\alpha^{(k)})$  sorozat monoton és korlátos, tehát létezik határértéke. Más szavakkal ez annyit jelent, hogy az iteráció minden esetben konvergens. Probléma a lokális maximumokkal illetve a nyeregpontokkal lehet, mert elképzelhető, hogy az iteráció határértéke nem az abszolút maximum. Ennek ellenőrzéseképpen célszerű az iterációt több különböző  $\alpha^{(0)}$  kezdőértékkel lefuttatni. Ha minden esetben ugyanaz adódik optimális megoldásként, akkor fogadjuk el az iteráció eredményét optimális becslésként.

## 5. Az EM algoritmus alkalmazása gammaszpektrumok kiértékelésére

A bevezetőben alkalmazott jelölésekkel élve tegyük fel, hogy a  $j$ -edik energiaintervallumban a fotonok száma  $x_j$ , ( $j = 1, 2, \dots, n$ ). A fotonok száma Poisson-eloszlást követ  $\lambda_j$  várható értékkel ( $j = 1, 2, \dots, n$ ). Feladatunk, hogy a detektorral mért spektrum, tehát az  $y_i$  ( $i = 1, 2, \dots, m$ ) adatok alapján meghatározzuk az  $x_j$  fotonszámokat, vagy ami ezzel gyakorlatilag egyenértékű, ezek  $\lambda_j$  átlagértékét. Tegyük fel, hogy a  $j$ -edik intervallumba tartozó fotont a detektor az  $i$ -edik csatornában regisztrálja. Ennek valószínűségét jelöli  $R_{ij}$ . Feltehető, hogy minden egyes emittált fotont a detektor regisztrál valamelyik csatornában, tehát teljesül, hogy

$$(6.1) \sum_{i=1}^m R_{ij} = 1 \quad ; (j = 1, 2, \dots, n).$$

A detektor válaszfüggvényének  $R_{ij}$  komponensei mérések útján meghatározható, detektorra jellemző állandók. Ezeket az alábbiakban ismertnek tételezzük fel.

A detektor által az  $i$ -edik csatornában regisztrált  $y_i$  beütésszám szintén Poisson eloszlású valószínűségi változó, melynek átlagértéke  $\lambda_i$  ( $i = 1, 2, \dots, m$ ).

Ezek szerint  $\frac{(\lambda_i)^{y_i}}{(y_i)!} \cdot \exp(-\lambda_i)$  annak valószínűsége, hogy az  $i$ -edik csatornában a beütésszám  $y_i$ . Mivel  $\lambda_i$  ennek a várható értéke, teljesül a

$$(6.2) \quad \lambda_i = E[y_i] = \sum_{j=1}^n y_i R_{ij}$$

összefüggés is. Legyen most  $z_{ij}$  ( $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ ) azon fotonok száma, amelyek a  $j$ -edik energiaintervallumba tartoznak és az  $i$ -edik csatornában vannak regisztrálva. Az 5. pont jelöléseivel és gondolataival összhangban, ezek azok a „rejtett paraméterek” amelyeket közvetlenül megfigyelni nem tudunk. Az EM algoritmus különleges sajátossága, hogy segítségével ezekre a paraméterekre is adhatunk becslést. A  $z_{ij}$  „paraméterek” is Poisson eloszlást követnek, jelölje ezek átlagértékét  $\lambda_{ij}$ . Nyilván teljesül, hogy (6.3)  $\lambda_{ij} = R_{ij} \cdot \lambda_j$ .

Feltehetjük, hogy az egyes energiatarományokba tartozó fotonok száma egymástól független, és ugyanezt elmondhatjuk a detektor egyes csatornáiban regisztrált beütésszámokról is. Ez azt jelenti, hogy a likelihood függvény szorzatalakban írható. Az eddigiekben általánosan  $\alpha$ -val jelöltük azokat a paramétereket amelyeket úgy szeretnénk hangolni, hogy a modell a legjobban illeszkedjen a mérési adatokhoz. Ezt tekintjük a likelihood függvény változójának. Az alábbi vizsgálatokban ezek a paraméterek a  $\lambda_{ij}$  átlagértékek, illetve (6.3) szerint a  $\lambda_j$  várható értékek. Ezért a likelihood függvény a

$$(6.4) \quad L(\lambda_{ij}) = \prod_{i=1}^m \prod_{j=1}^n \frac{(\lambda_{ij})^{z_{ij}}}{(z_{ij})!} \cdot \exp(-\lambda_{ij})$$

alakban írható. Ennek logaritmus, a log-likelihood függvény így a következő:

$$(6.5) \quad \ln L(\lambda_{ij}) = \sum_{i=1}^m \sum_{j=1}^n (-\lambda_{ij} + z_{ij} \ln(\lambda_{ij}) - \ln(z_{ij})!)$$

Ha itt ténylegesen figyelembe vesszük a (6.3) összefüggést, akkor a  $\lambda_j$  paraméterektől függő alakot kapjuk:

$$(6.6) \quad \ln L(\lambda_j) = \sum_{i=1}^m \sum_{j=1}^n (-R_{ij} \lambda_j + z_{ij} \ln(\lambda_j) + z_{ij} \ln(R_{ij}) - \ln(z_{ij})!)$$

Alkalmazzuk  $\ln L(\lambda_j)$  maximumának meghatározására az EM algoritmust. Az algoritmus két lépésből áll:

1. lépés („E-step”): Meg kell határozni az (5.5)-nek megfelelő

$$E[\ln P(y, z | \lambda)] = \sum_z P(z | y, \lambda^{(k)}) \cdot \ln P(y, z | \lambda)$$

feltételes várható értéket. Felhívjuk a figyelmet, hogy itt  $\ln P(y, z | \lambda)$  éppen az  $\ln L(\lambda)$  log-likelihood függvény. A (6.6) z-szerinti feltételes várható értékének kiszámítását végezhetjük tagonként. Az első tag konstans, így átlaga saját maga. A többi három tagban is elég  $z_{ij}$  feltételes várható értékét kiszámítani, majd a második és harmadik tagban konstanssal szorozni. A k-adik iterációs lépésben kapott  $\lambda_{ij}^{(k)}$  közelítés birtokában, az (5.6) összefüggés szerint, ez az átlagérték éppen a  $z_{ij}$  közelítő értéke a (k+1)-edik iterációs lépésben: (6.7)  $z_{ij}^{(k+1)} = E[z_{ij} | y_i, \lambda_{ij}^{(k)}]$

Ennek meghatározása érdekében először számítsuk ki a  $P[z_{ij} | y_i, \lambda_{ij}^{(k)}]$  feltételes valószínűségeket. Ha felhasználjuk a nyilvánvaló

$\lambda_i = \sum_{j=1}^m \lambda_{ij}$  és az  $y_i = \sum_{j=1}^m z_{ij}$  összefüggéseket, a feltételes valószínűség

fogalmát, továbbá azt, hogy  $z_{ij}$  és  $y_i$  Poisson eloszlásúak, akkor azt kapjuk, hogy:

$$\begin{aligned} P(z_{ij} | y_i, \lambda_{ij}) &= \frac{P(z_{ij}, y_i | \lambda_{ij})}{P(y_i | \lambda_{ij})} = \frac{\frac{(\lambda_{ij})^{z_{ij}} \cdot \exp(-\lambda_{ij}) \cdot \left( \sum_{\alpha \neq i} \lambda_{\alpha j} \right)^{y_i - z_{ij}}}{(z_{ij})! \cdot (y_i - z_{ij})!} \cdot \exp\left(-\sum_{\alpha \neq i} \lambda_{\alpha j}\right)}{\frac{(\lambda_i)^{y_i} \cdot \exp(-\lambda_i)}{(y_i)!}} = \\ &= \frac{(y_i)!}{(z_{ij})! (y_i - z_{ij})!} \cdot \left( \frac{\lambda_{ij}}{\lambda_i} \right)^{z_{ij}} \cdot \left( 1 - \frac{\lambda_{ij}}{\lambda_i} \right)^{y_i - z_{ij}} \end{aligned}$$

A számítás eredménye azt mutatja, hogy a  $P[z_{ij} | y_i, \lambda_{ij}^{(k)}]$  feltételes valószínűségeloszlás éppen Binomiális eloszlás. Ennek pedig közismert a várható értéke, ami egyben a  $z_{ij}^{(k+1)}$  közelítő érték:

$$(6.8) \quad z_{ij}^{(k+1)} = E[z_{ij} | y_i, \lambda_{ij}^{(k)}] = y_i \cdot \frac{\lambda_{ij}^{(k)}}{\lambda_i} = y_i \cdot \frac{\lambda_{ij}^{(k)}}{\sum_{j=1}^n \lambda_{ij}^{(k)}}$$

2. lépés ("M-step"): Maximalizálnunk kell a (6.6) függvény feltételes várható értékét. Ehhez először formálisan behelyettesítjük a (6.8) összefüggést a (6.6) függvénybe, ezzel képeztük a várható értéket:

$$(6.9) \quad \ln L(\lambda_j) = \sum_{i=1}^m \sum_{j=1}^n \left( -R_{ij} \lambda_j + z_{ij}^{(k+1)} \ln(\lambda_j) + z_{ij}^{(k+1)} \ln(R_{ij}) - \ln(z_{ij}^{(k+1)})! \right)$$

Majd kiszámítjuk a  $\frac{\partial \ln L(\lambda_j)}{\partial \lambda_j}$  deriváltakat és megoldjuk a  $\frac{\partial \ln L(\lambda_j)}{\partial \lambda_j} = 0$ ;

( $j = 1, 2, \dots, m$ ) egyenleteket. Elvégezve a deriválást azt kapjuk, hogy:

$$\sum_{i=1}^m \left( -R_{ij} + z_{ij}^{(k+1)} \cdot \frac{1}{\lambda_j} \right) = 0$$

Ha felhasználjuk a (6.1) összefüggést, akkor kapjuk ezeknek az egyenleteknek a megoldását:

$$\lambda_j^{(k+1)} = \sum_{i=1}^m z_{ij}^{(k+1)}$$

Helyettesítsük most ebbe az összefüggésbe a (6.8) eredményt:

$$\lambda_j^{(k+1)} = \sum_{i=1}^m y_i \cdot \frac{\lambda_{ij}^{(k)}}{\sum_{j=1}^n \lambda_{ij}^{(k)}}$$

Mivel (6.3) szerint még az is teljesül, hogy  $\lambda_{ij} = R_{ij} \cdot \lambda_j$ , ezért a  $\lambda_j$  paraméterekre vonatkozó iterációs formula végül a következő alakot ölti:

$$(6.10) \quad \lambda_j^{(k+1)} = \sum_{i=1}^m y_i \cdot \frac{R_{ij} \lambda_j^{(k)}}{\sum_{j=1}^n R_{ij} \lambda_j^{(k)}} = \lambda_j^{(k)} \sum_{i=1}^m \frac{y_i R_{ij}}{\sum_{j=1}^n R_{ij} \lambda_j^{(k)}}$$

A (6.10) formula tehát a valós spektrumkomponensek átlagértékére vonatkozó iterációs formula. Mivel  $\lambda_j$  az  $x_j$  átlaga, gyakorlatilag ugyanez a formula szolgál az  $x_j$  fotonszámok iterációs becslésére is [5, 6]:

$$(6.11) \quad \lambda_j^{(k+1)} = x_j^{(k)} \sum_{i=1}^m \frac{y_i R_{ij}}{\sum_{j=1}^n R_{ij} x_j^{(k)}}$$

Az ML-EM algoritmus — bár eredetileg pozitron emissziós tomográfiai vizsgálatok kiértékeléséhez lett kifejlesztve —, nagyon hatékonyan alkalmazható módszer a spektrumok kiértékelésének problémakörében is. Hátránya, hogy az iteráció viszonylag lassan konvergál, és bizonytalanság van abban, hogy a megoldás abszolút vagy lokális maximumot ad-e, de előnye, hogy a dekonvolúciós módszerek között a legjobb felbontást biztosítja, a legélesebb csúcsokat és a legmélyebb völgyeket állítja elő a spektrumban, és akkor is eredményesen alkalmazható, ha viszonylag alacsony a jel/zaj arány.

## Felhasznált irodalom

1. Jánossy Lajos: Mérési eredmények kiértékelésének elmélete és gyakorlata. Akadémiai Kiadó. 1968
2. V.A. Muravsky–S.A. Tolstov–A.L. Kholmetskii: Comparison of the least squares and the maximum likelihood estimators for gamma-spectrometry Nuclear Instruments and Methods in Physics Research B 145, (1998) 573–577
3. T. Hauschild–M. Jentschel: Comparison of maximum likelihood estimation and chi-square statistics applied to counting experiments. Nuclear Instruments and Methods in Physics Research A 457, (2001) 384–401
4. E.T. Jaynes: Information Theory and Statistical Mechanics. Physical Review 106, (1957) 620–630.
5. L. Bouchet: A Comparative study of deconvolution methods for gamma-ray spectra. Astronomy & Astrophysics Supplement Series. 113, (1995) 167–183.
6. L.J. Meng–D. Ramsden: An inter comparison of Three Spectral-deconvolution Algorithms for Gamma-ray Spectroscopy 47, (2000) No.4.
7. Todd K. Moon: The Expectation Maximization Algorithm: IEEE Signal Processing Magazine, 1996 November