

미약한 방출선의 세기 계산 INTENSITY ESTIMATION OF WEAK EMISSION LINES

선광일, 이대희
한국천문연구원

KWANG-IL SEON AND DAE-HEE LEE

Korea Astronomy and Space Science Institute, 61-1 Hwaam, Yusong, Taejon 305-348, Korea

E-mail: kiseon@kasi.re.kr

(Received November 15, 2005; Accepted: November 28, 2005)

ABSTRACT

We are often faced with the task of having to estimate the amplitude of a source signal in the presence of a background. In the simplest case, the background can be taken as being flat, and of unknown magnitude B , and the source signal of interest assumed to be the amplitude A of a peak of known shape and position. We present a robust method to find the most probable values of A and B by applying the one-dimensional Newton-Raphson method. In the derivation of the formula, we adopted the Bayesian statistics and assumed Poisson distribution so that the results could be applied to the analysis of very weak signals, as observed in FIMS (Far-ultraviolet IMaging Spectrograph).

key words: methods: data analysis — methods: statistical

1. 서론

천문학 분광관측에 있어서 방출선의 세기를 계산하는 일은 천체의 이온화 및 에너지 상태를 이해하는 데 있어서 매우 중요하다. 방출선의 선운곽(line spread function)을 실험실 측정등을 통하여 미리 잘 알고 있고 방출선의 세기가 강한 경우는 일반적으로 가우스(Gaussian) 통계를 만족한다고 가정할 수 있으므로 χ^2 최소화 방법을 통해 방출선의 세기를 어렵지 않게 계산할 수 있다. 그러나, 채널(channel)당 검출광자의 수가 5개 미만인 경우에는 가우스 통계를 사용할 수 없으며 프와송(Poisson) 통계를 사용해야만 한다 (Eadie et al. 1971).

X-선 또는 원자외선 분광관측의 경우 매우 미약한 방출선의 세기를 측정해야하는 경우가 자주 발생한다. 특히, FIMS/SPEAR (Far ultraviolet IMaging Spectrograph/Spectroscopy of Plasma Evolution from Astrophysical Radiation) 관측자료는 2차원 영상의 각 픽셀별로 분광정보를 담고 있어서 픽셀별로 방출선의 세기를 계산하기 위해서는 iraf 등의 잘 알려진 패키지를 사용하기에는 어려움이 따른다. 더군다나, iraf에 포함되어 있는 분석 패키지는 가우스 통계를 가정하고 있으므로 미약한 시그날을 처리하기에는 부적당하다. FIMS는 우리나라 최초의 천문관측위성인 과학기술위성 1호(STSAT-1)의 주 탑재체로서 2003년 9월 27일 발

사되었으며 약 2/3의 하늘을 관측하였다. FIMS 관측자료를 분석하기 위해 미약한 방출선의 세기 계산 방법을 이미 개발하여 사용하고 있지만 (Seon et al. 2005), 이 방법으로는 방출선의 세기를 정확하게 계산하지 못하는 경우가 종종 발생하였다.

이 논문에서는 방출선의 선운곽을 잘 알고 있을 때, 방출선의 세기 계산을 위해 베이지안(Bayesian) 방법을 적용하여 방출선의 세기가 매우 약할 때에도 적용 가능한 신뢰성 높은 알고리즘을 제시한다. 이 논문에서 기술한 방법은 Seon et al. (2005)에서 적용한 2차원 Newton-Raphson 방법을 좀 더 개선한 방법으로서 1차원 Newton-Raphson 방법을 사용하기 때문에 좀 더 신뢰성을 높일 수 있다.

2. 베이지안 분석

우리는 배경 신호(background signal)을 포함하고 있는 관측자료로부터 방출선의 세기를 계산하는 일에 자주 부딪힌다. 이 논문에서는 배경 신호의 세기는 알지 못 하지만 상수값을 갖고 있고 방출선의 선운곽을 실험실 관측등을 통해 이미 알고 있는 경우만을 다루기로 한다. 이러한 가정은 FIMS 자료로 부터 방출선 영상을 구할 때와 같이 많은 자료를 한꺼번에 처리할 때 매우 유용하게 사용할 수 있다. 미약한 방출선의 세기를 계산함에 있어 선운곽의 폭(width) 등을 알지 못한다고 가정

하고 선운과의 폭과 함께 방출선의 세기를 동시에 계산하는 것은 실제적으로 불가능한 경우가 많다. 따라서 미약한 방출선을 분석하고자 할 때는 선운과의 함수형태를 잘 알고 있다고 가정하는 것을 피할 수 없다.

이 연구에서는 일반적인 많이 알려져 있는 빈도주의(frequentist) 확률이론을 사용하지 않고 Sivia (1996)에서와 같이 베이지안 방법을 적용하여 방출선의 세기를 구하도록 한다. 베이지안 확률이론의 장점은 미약한 신호를 처리할 때, 배경값을 빼주고 나면 신호값이 음수로 나오는 현상을 막을 수 있다는 장점이 있다. 따라서 미약한 신호를 처리하기 위해서는 베이지안 방법을 이용하는 것이 필수적이라고 할 수 있다(베이지안 확률이론에 대한 자세한 설명은 Sivia 1996을 참조).

k 번째 채널에서 예상되는 광자개수는 방출원(emission source)의 신호와 배경 신호의 합에 비례할 것이다. 따라서 선운과 함수가 채널에 따라 f_k 로 주어지면, k 번째 채널에서 예상되는 이상적인 데이터값 D_k 은

$$D_k = Af_k + B \quad (k = 1, \dots, M) \quad (1)$$

이 된다. 이때, A 와 B 의 단위는 각각 counts와 counts/channel이다. 또한, 채널의 개수는 M 이며 선운과를 모든 채널에 대해 더한 값은 $F \equiv \sum_k f_k$ 이다. 일반적으로 $F = 1$ 이지만, 분석을 위해 사용하는 채널의 폭이 선운과의 폭에 비해 그리 넓지 않는 일반적인 경우를 대비하여 F 값을 1로 놓지 않고 공식을 유도하였다.

실제 관측된 측정치 $\{N_k\}$ 가 주어졌을 때, 실험치를 가장 잘 설명하는 방출선의 세기 A 와 배경의 세기 B 를 계산하기 위한 사후 확률분포함수 (posterior probability distribution function) $P(A, B | \{N_k\})$ 는 베이즈 정리(Bayes' Theorem)로 부터

$$P(A, B | \{N_k\}) = \frac{P(\{N_k\} | A, B) P(A, B)}{P(\{N_k\})} \quad (2)$$

로 주어진다. 이때, $P(\{N_k\} | A, B)$ 는 A, B 가 주어졌을 때 관측 자료의 확률을 의미하는 가능성(log-likelihood) 함수, $P(A, B)$ 는 사전 확률분포함수 (prior probability distribution function), $P(\{N_k\})$ 는 증거(evidence)이다. 이 연구에서는 일반적으로 받아들이는 균등 사전 분포(uniform prior)를 가정하였다. 즉, $P(A, B)$ 는 $A \geq 0$ 와 $B \geq 0$ 일 때만 1의 값을 갖고 그 외의 경우는 0이다. 따라서, A, B 값은 0 이상인 값만 가질 수 있다.

만약 관측자료가 서로 독립적이어서 임의의 채널에서 관측되는 광자의 개수가 다른 채널에서 관측되는 광자의 개수와 무관하고 프와송 분포를 따른다면 가능도 함수는 다음과 같이 주어진다.

$$P(\{N_k\} | A, B) = \prod_k \frac{D_k^{N_k} e^{-D_k}}{N_k!} \quad (3)$$

로그 가능성(log-likelihood) 함수는

$$L = \sum_k [N_k \ln(D_k) - D_k] + \text{constant} \quad (4)$$

이며, 사후 확률분포함수 $P(A, B)$ 는 다음과 같이 나타낼 수 있다.

$$P(A, B) \propto e^{L(A, B)} = \prod_k (Af_k + B)^{N_k} e^{-(AF + BM)} \quad (5)$$

즉, $P(A, B)$ 는 지수함수(exponential function)와 다항식의 곱이 된다. 또한 주변 분포함수(marginal distribution function) $P(A)$ 와 $P(B)$ 는 아래와 같이 정의된다.

$$P(A) = \int_0^\infty dB P(A, B) \quad (6)$$

$$P(B) = \int_0^\infty dA P(A, B) \quad (7)$$

물론, $P(A)$ 와 $P(B)$ 도 역시 지수함수와 다항식의 곱이 된다.

3. 방출선의 세기

사후 분포함수 $P(A, B)$ 가 최대값이 되도록 하는 가장 찾은 값 (most probable value, 또는 mode) A_m 및 B_m 이 우리가 찾고자 하는 방출선 및 배경의 세기가 된다. 즉, 다음의 연립방정식으로 부터 방출선 및 배경의 세기가 결정된다.

$$\begin{aligned} \frac{\partial L}{\partial A} &= \sum_k \left(\frac{N_k}{D_k} - 1 \right) \frac{\partial D_k}{\partial A} \\ &= \sum_k \frac{N_k f_k}{D_k} - F \\ &= 0 \end{aligned} \quad (8)$$

$$\begin{aligned} \frac{\partial L}{\partial B} &= \sum_k \left(\frac{N_k}{D_k} - 1 \right) \frac{\partial D_k}{\partial B} \\ &= \sum_k \frac{N_k}{D_k} - M \\ &= 0 \end{aligned} \quad (9)$$

다시 정리하면 다음과 같다.

$$\sum_k \frac{N_k f_k}{Af_k + B} = F \quad (10)$$

$$\sum_k \frac{N_k}{Af_k + B} = M \quad (11)$$

위의 두 연립방정식은 Newton-Raphson 방법을 이용하여 풀 수 있다(Press et al. 1992). 즉, A 와 B 의 2차원 평

면상에서 해를 구하는 것이다. 그러나, 2차원 Newton-Raphson 방법이 수렴하지 않는 경우가 종종 발생하였다.

이러한 문제점을 해결하고 연립방정식의 해를 어떤 경우든 구할 수 있도록 하기 위해 두 개의 연립방정식을 한개의 변수에 대한 방정식으로 표현하는 방법을 찾았다. B 가 0보다 크다고 가정하고, 식 (10)과 식 (11)의 양변에 B 를 곱하면 다음의 식들을 얻게 된다.

$$\begin{aligned} \sum_k \frac{N_k f_k}{(A/B)f_k + 1} &= BF \\ \sum_k \frac{N_k}{(A/B)f_k + 1} &= BM \end{aligned} \quad (12)$$

이때, 우변은 우변끼리, 좌변은 좌변끼리를 나누어주면 아래의 식을 얻을 수 있다.

$$g_1(A/B) \equiv \sum_k \frac{N_k(f_k - F/M)}{(A/B)f_k + 1} = 0 \quad (13)$$

마찬가지로 A 가 0보다 크다고 가정하고, 식 (10)과 식 (11)의 양변에 A 를 곱한 후 얻어지는 식은

$$\begin{aligned} \sum_k \frac{N_k f_k}{f_k + (B/A)} &= AF \\ \sum_k \frac{N_k}{f_k + (B/A)} &= AM \end{aligned} \quad (14)$$

이며, 우변과 좌변을 나누어주면 아래의 식을 얻는다.

$$g_2(B/A) \equiv \sum_k \frac{N_k(f_k - F/M)}{f_k + (B/A)} = 0 \quad (15)$$

따라서, 식 (10)과 식 (11)를 연립하여 푸는 것보다는 식 (13) 또는 식 (15)를 이용하여 A/B 또는 B/A 를 각각 구한 후, 식 (12) 또는 식 (14)의 좌변에 구해진 해를 대입하여 B 와 A 를 구할 수 있다. 두 개의 변수를 갖는 연립방정식이 아니라, 한개의 방정식을 풀므로써 좀 더 알고리즘을 간편하게 할 수 있어보인다. 그러나, 1차원 함수라고 해서 항상 Newton Raphson 방법을 이용했을 때 수렴하는 것은 아니므로 신뢰성 있는 해를 구할 수 있는지 확인하기 위해 함수 $g_1(x)$ 와 함수 $g_2(x)$ 의 수학적 특성을 분석할 필요가 있다.

두 함수의 모양으로부터 $A \neq 0$ 과 $B \neq 0$ 인 해가 존재하기 위해서는 $f_k > F/M$ 인 채널구간에 포함된 광자의 수와 $f_k < F/M$ 인 채널구간에 포함된 광자의 수가 적당히 균형을 이루어야 만 가능함을 알 수 있다. 두 함수의 모양은 $x = 0$ 인 경우와 아닌 경우를 비교함으로써 자세하게 분석할 수 있다. $g_1(x)$ 함수의 경우, $x = 0$ 일 때 k 번째 항은 $h_k \equiv N_k(f_k - F/M)$ 이며, $f_k >$

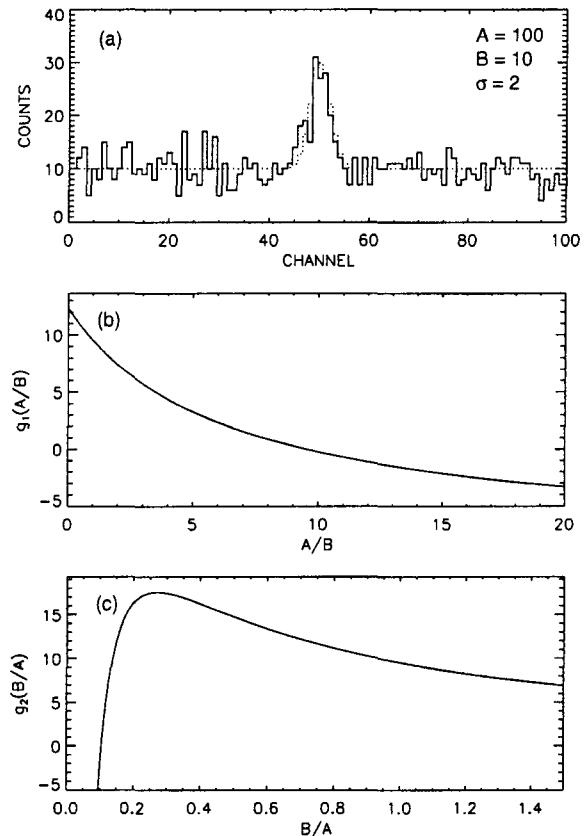


그림 1.— (a) $A = 100$, $B = 10$ 와 $\sigma = 2$ 인 가우스 함수 모양의 선운곽함수를 가정했을 때 얻어진 모의 스펙트럼 (점선은 가정한 모델, 실선은 모의 스펙트럼), (b) $g_1(A/B)$ 함수의 모양, (c) $g_2(B/A)$ 함수의 모양.

F/M 인 채널구간에 포함된 광자의 수와 $f_k < F/M$ 인 채널구간에 포함된 광자의 수에 의해 $g_1(x = 0)$ 값이 음수인지 양수인지가 결정된다. $x \neq 0$ 일 때, $g_1(x)$ 함수의 k 번째 항 h_k 의 공헌되는 정도를 결정짓는 인수는 $1/(x f_k + 1)$ 라고 생각할 수 있으며, 이 값은 x 가 증가할 때, $f_k > F/M$ 구간에서 그 밖의 구간에서보다 더 빨리 감소한다. 따라서, x 가 증가할 때, $f_k > F/M$ 구간에 포함된 값들에 의해 공헌되는 값들이 그 밖의 구간의 값들에 의해 공헌되는 값들보다 빠르게 감소한다. 즉, x 가 증가하면서 $f_k < F/M$ 인 항들이 더 우세하게 되어 점점 감소하게 된다. 다시 말하면, $g_1(x)$ 는 단조감소 함수이다. 만약 $g_1(x = 0) > 0$ 이었다면 x 가 증가하면서 $g_1(x)$ 값이 계속 감소하기 때문에 $g_1(x) = 0$ 인 해가 존재한다. 그러나, $g_1(x = 0) < 0$ 이었다면 $g_1(x) = 0$ 인 해는 존재하지 않는다.

반면에 $g_2(x)$ 함수의 경우, $x = 0$ 일 때 k 번째 항은 $h_k \equiv N_k(f_k - F/M)/f_k$ 이며, $x \neq 0$ 일 때, $g_2(x)$ 함수의 k 번째 항 h_k 의 공헌정도를 결정짓는 인수는 $1/(1 +$

x/f_k)이다. 이 값의 경우는 x 가 증가할 때, $f_k > F/M$ 인 구간에서 그렇지 않은 구간에서 보다 더 느리게 감소한다. 따라서, x 의 증가에 따라 $g_2(x)$ 값도 증가한다. 어느 정도 x 값이 커지게 되면 $1/(1+x/f_k) \approx f_k/x$ 가 되므로 $g_2(x) \approx g_1(x=0)/x$ 이 된다. $g_2(x)$ 를 유도하기 위한 기본가정이 $A \neq 0$ 인 해가 존재한다는 것이었으므로 $g_1(x=0) > 0$ 이고 따라서, x 가 어느 정도 커지기 시작하면 $g_2(x)$ 는 양수이면서 x 에 반비례하며 감소하는 함수가 된다. 따라서, $g_2(x)=0$ 인 해가 존재하려면 $g_2(x=0) < 0$ 이어야 가능하다. 또한, 증가하다가 다시 감소하므로 $g'_2(x)=0$ 인 지점이 반드시 존재한다. $g_2(x)=0$ 의 해가 존재한다면, $g'_2(x=0)$ 의 k 번째 항은 $-h_k/f_k$ 이므로 $g'_2(x=0) > 0$ 이고 $1/f_k$ 인수에 의해 $g_2(x=0)$ 인 경우보다 $f_k < F/M$ 인 항들이 더 우세하게 되어 절대값이 매우 커지게 된다. 다시 말하면 $x=0$ 근방에서 매우 급격하게 증가하는 함수모양을 갖게 된다.

그림 1은 $A = 100$, $B = 10$ 이고, 선운과 함수가 $\sigma = 2$ 인 가우스 함수 일때, 프와송 랜덤함수를 이용해서 만들어진 모의 스펙트럼과 두 함수의 모양을 보여준다. 그림으로부터 앞에서 기술한 두 함수 $g_1(x)$ 와 $g_2(x)$ 의 특성을 확인 할 수 있다. Newton-Raphson 방식을 사용했을 때, 항상 수렴하기 위해서는 함수의 모양이 급격하게 변해서는 안되고 변곡점 ($g'(x) = 0$ 인 지점)이 존재하지 않아야 한다 (Press et al. 1992). 또한, $g_2(x)$ 함수의 경우는 x 가 매우 큰 지점에서 출발하면 Newton-Raphson 방법의 특성상 $g_2(x)=0$ 의 해는 구하지 못하고 $x \rightarrow \infty$ 로 발산하게 된다. 따라서, B/A 에 대한 방정식 $g_2(B/A) = 0$ 보다는 A/B 에 대한 방정식 $g_1(A/B) = 0$ 이 풀기 수월하며 어떤 상황에서도 Newton-Raphson 방법을 통해 항상 수렴하는 값을 구할 수 있다.

앞의 논의로 부터 $A_m = 0$ 이거나 $B_m = 0$ 인 조건을 손쉽게 구할 수 있지만, 주변 분포함수의 특성으로부터 두 가지 조건을 구하여 앞에서 논의한 결과와 일치하는지 살펴보도록 하자. $A_m > 0$ 이 존재하기 위해서는 주변 분포함수 $P(A)$ 의 미분이 0이 되는 지점이 존재해야 하며, $P(A)$ 함수가 $A = 0$ 일 때, 최대가 되기 위해서는 $P(A)$ 의 미분값이 $A = 0$ 에서 음수가 되어야 한다. 즉, $P'(A=0) < 0$ 이어야 한다. $B_m = 0$ 인 조건도 동일하게 $P'(B=0) < 0$ 이 된다. 주변 분포함수의 정의로 부터 $P'(A=0)$ 와 $P'(B=0)$ 는 각각

$$\begin{aligned} \frac{dP(A)}{dA} \Big|_{A=0} &\propto \int_0^\infty dB e^{L(0,B)} \frac{\partial L}{\partial A} \Big|_{A=0} \\ &= \int_0^\infty dB e^{-BM} B^N \left(\sum_k \frac{N_k f_k}{B} - F \right) \\ &= \frac{(N-1)!}{M^N} \sum_k N_k f_k - \frac{N!F}{M^{N+1}} \end{aligned}$$

$$\propto \sum_k N_k f_k - N \frac{F}{M} \quad (16)$$

$$\begin{aligned} \frac{dP(B)}{dB} \Big|_{B=0} &\propto \int_0^\infty dA e^{L(A,0)} \frac{\partial L}{\partial B} \Big|_{B=0} \\ &= \int_0^\infty dA e^{-AF} A^N \left(\sum_k \frac{N_k}{Af_k} - M \right) \\ &= \frac{(N-1)!}{F^N} \sum_k \frac{N_k}{f_k} - \frac{N!M}{F^{N+1}} \\ &\propto \sum_k \frac{N_k}{f_k} - \frac{N}{F/M} \quad (17) \end{aligned}$$

가 된다. 여기서 N 은 $N \equiv \sum_k N_k$ 로 정의되며 모든 채널에 걸쳐 검출된 총 광자의 수를 나타낸다. 따라서,

$$A_m = 0 \text{ if } \sum_k N_k f_k \leq N \frac{F}{M} \quad (18)$$

$$B_m = 0 \text{ if } \sum_k \frac{N_k}{f_k} \leq \frac{N}{F/M} \quad (19)$$

임을 알 수 있다. 위의 두 조건은 앞에서 논의한 $g_1(x) = 0$ 과 $g_2(x) = 0$ 의 해가 존재할 조건 $g_1(x=0) > 0$ 과 $g_2(x=0) < 0$ 으로부터 구해지는 조건과 완전히 일치함을 확인 할 수 있다.

여기서 한가지 주목할 점은 식 (10)과 식 (11)과 함께 $Af_k + B \geq Af_k$ 또는 $Af_k + B \geq B$ 이라는 점을 이용하면 A_m 와 B_m 의 상한값을 구할 수 있다는 것이다. 즉,

$$A_m \leq \min \left(\frac{N}{F}, \frac{\sum_k N_k / f_k}{M} \right) = \frac{N}{F} \quad (20)$$

$$B_m \leq \min \left(\frac{\sum_k N_k f_k}{F}, \frac{N}{M} \right) = \frac{N}{M} \quad (21)$$

위의 식에서 $\min(a, b)$ 는 a 와 b 중 작은 수를 나타내며, 팔호안의 첫번째 값은 식 (10)을 이용하여 구해졌고, 두 번째 값은 식 (11)을 이용하여 구해졌다. 이때, $\min(a, b)$ 함수뒤의 등호는 바로 앞에서 구해진 $A_m > 0$ 와 $B_m > 0$ 이기 위한 조건으로부터 쉽게 알 수 있다. 위의 결과는 직관적으로 쉽게 이해될 수 있다. 즉, 방출선의 세기 A_m 은 관측된 모든 광자가 방출선에 의한 것이라고 가정했을 때 얻어지는 값보다 작거나 같아야 하고, 배경의 세기 B_m 은 관측된 모든 광자가 전부 배경 신호에 의한 것이라고 가정했을 때의 값보다 작거나 같아야 한다.

$B_m = 0$ 일 때 A_m 의 값과 $A_m = 0$ 일 때 B_m 의 값은 각각 식 (10)과 식 (11)로 부터

$$A_m = N/F \text{ if } B_m = 0 \quad (22)$$

$$B_m = N/M \text{ if } A_m = 0 \quad (23)$$

임을 알 수 있다. $g_1(x) = 0$ 을 Netwon-Raphson 방법으로 풀 때, 처음 시작점으로 $x_0 = 0$ 에서 출발하면 충분하다. 또는 좀 더 적당한 시작점을 잡는다면

$$A_0 = \sum_{f_k > F/M} N_k / F \quad (24)$$

$$B_0 = \sum_{f_k < F/M} N_k / M \quad (25)$$

으로 정의하고 $x_0 = A_0 / B_0$ 을 시작점으로 할 수도 있다.

4. 결론 및 요약

베이지안 통계이론을 이용하면 방출선의 세기와 배경의 세기를 나타내는 두 개의 변수에 대한 연립방정식을 구할 수 있다. Seon et al. (2005)은 백조자리 초신성 잔해의 분광영상을 얻기 위해 2차원 Newton-Raphson 방법을 이용하여 연립방정식을 풀었다. 그러나, FIMS 전천 서베이 자료등을 분석하는 과정에서 구해진 연립방정식을 2차원 Newton-Raphson 방법을 이용하여 풀 때, 종종 수렴하지 않는 경우를 접할 수 있었다. 따라서, 두 개의 방정식을 연립하여 푸는 문제를 한개의 방정식으로 풀 수 있도록 하였으며, 자세한 분석을 통해서 방정식 $g_2(x) = 0$ 에 Newton-Raphson 방법을 적용하면 언제나 수렴하는 해 A_m / B_m 를 찾을 수 있다는 것을 알았다. 이와 함께 식 (12)를 이용하면 A_m 과 B_m 을 얻을 수 있다.

참고문헌

- Eadie, W. T., Dryard, D., James, F. E., Roos, M., & Sadoulet, B., 1971, Statistical Methods in Experimental Physics, (Noth-Holland Publishing Company: Amsterdam)
- Seon, K.-I., Han, W., Nam, U.-W., Park, J.-H., Edelstein, J., Korpela, E., Sankrit, R., Min, K.-W., Ryu, K.-S., & Kim, I.-J., 2005, ApJL, accepted
- Sivia, D. S., 1996, Data Analysis, A Bayesian Tutorial, (Clarendon Press: Oxford)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P., 1992, Numerical Recipes in Fortran, 2nd Ed., (Cambridge University Press: New York)