

バイアス項を含めた対数尤度の計算

梅原 大祐

初出: 平成 19 年 7 月 6 日

修正: 平成 19 年 8 月 24 日

修正: 平成 21 年 9 月 12 日

1 はじめに

確率分布間の距離を表すものとして KL 情報量がある．KL 情報量とは，真の確率密度関数を q とし，モデル化した確率密度関数を p としたときに，

$$D(q||p) = E_Q \left[\ln \frac{q(x)}{p(x)} \right] = \int_{-\infty}^{\infty} q(x) \ln \frac{q(x)}{p(x)} dx \quad (1)$$

で与えられる．ただし， Q は確率密度関数 q を持つ確率分布であり， $dQ(x) = q dx$ の関係がある．KL 情報量 $D(q||p)$ が小さいときに，確率密度関数 p は真の確率密度関数 q に近いものとして考えられる．すなわち，真の確率密度関数 q が与えられたとき KL 情報量 $D(q||p)$ を小さくする確率分布 p は精度良く真の確率分布 q をモデル化できていると考えることができる．我々は確率的な現象を生み出す現実の問題をモデル化したい場合，その真の分布 q に対して，KL 情報量 $D(q||p)$ が可能な限り小さくなる分布 p を探すことが有効である．このことは，平均対数尤度

$$E_Q[\ln p(x)] = \int_{-\infty}^{\infty} q(x) \ln p(x) dx \quad (2)$$

が可能な限り大きい確率密度関数 p を探すことと同義である．KL 情報量の式 (1) に比べて，平均対数尤度の式 (2) は表現が簡易であるため，以降は情報量を表す尺度として，平均対数尤度を用いることとする．

しかし，我々は真の確率分布は知ることはできない．得ることができるのは，真の確率分布に従って生成されるデータ系列 $x = (x_i)_{1 \leq i \leq n}$ である．エルゴード性を仮定した場合，系列 $y_i = \ln(x_i)$ の平均は，その期待値に確率収束し，

$$\frac{1}{n} \sum_{i=1}^n \ln p(x_i) \rightarrow E_Q[\ln p(x)], \quad n \rightarrow \infty \quad (3)$$

となる．式 (3) の左辺はデータ系列から計算可能であり，情報量規準として適切な量である．通常，系列長 n を掛け合わせた

$$\sum_{i=1}^n \ln p(x_i) \quad (4)$$

を対数尤度と呼ぶが，系列長 n の増加に従い増加していくので，本稿では

$$\frac{1}{n} \sum_{i=1}^n \ln p(x_i) \quad (5)$$

のことを対数尤度と呼ぶことにする．

2 バイアス項を含めた対数尤度

モデル化する分布 p が u 個のパラメータ $\theta = (\theta_i)_{1 \leq i \leq u}$ を用いて表現されるとし、モデル化する分布 p を $p(x|\theta)$ と記述する．ここで、 u はパラメータの次元と呼ぶ．このとき、平均対数尤度を

$$\eta(\theta) = E_Q[\ln p(x|\theta)] \quad (6)$$

と表記し、対数尤度を

$$\ell(\theta) = E_{\hat{Q}}[\ln p(x|\theta)] = \frac{1}{n} \sum_{i=1}^n \ln p(x_i|\theta) \quad (7)$$

と表記する．ここで、 \hat{Q} は系列 x による経験分布関数を表す．

パラメータ θ の推定は、対数尤度 $\ell(\theta)$ が最大になる最尤推定量 $\theta = \hat{\theta}$ を選択する．すなわち、 θ の極値を与える尤度方程式 $\partial \ell(\theta) / \partial \theta = 0$ を満足する $\theta = \hat{\theta}$ を求めることが望まれる．しかし、そのモデルのよさの評価として対数尤度 $\ell(\hat{\theta})$ の大小比較では公正なモデル比較にならないことが知られている．これは、 $\ell(\hat{\theta})$ が最尤モデルの平均対数尤度 $\eta(\hat{\theta})$ の推定量として図1に示すようなバイアス $b(Q)$ を持ち、そのバイアスの大きさがパラメータの次元により異なることに起因する．一般に、パラメータの次元数が増加するに伴い、最尤推定量の不確実性が増加し、バイアス項 $b(Q)$ が大きくなる．パラメータの次元を含めた平均対数尤度 $\eta(\hat{\theta})$ によるモデル比較が望まれる．

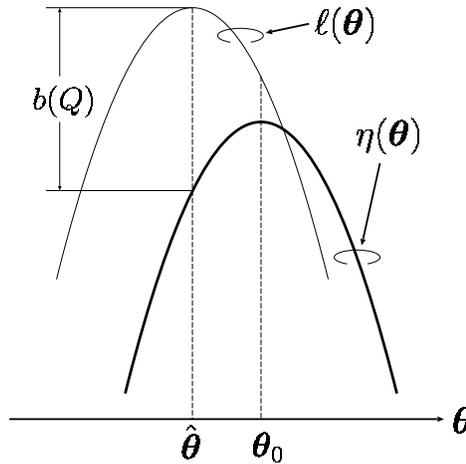


図 1: 平均対数尤度のバイアス

バイアス項 $b(Q)$ は

$$b(Q) = \frac{1}{n} \text{Tr}(I_Q(\theta_0) J_Q(\theta_0)^{-1}) \quad (8)$$

と展開され、

$$\eta(\hat{\theta}) = \ell(\hat{\theta}) - b(Q) \quad (9)$$

となる [1]．ここで、 $\theta = \theta_0$ は平均対数尤度 $\eta(\theta)$ を最大にする真の最尤推定量である．なお、バイアス項を含めた平均対数尤度 $\eta(\hat{\theta})$ は竹内情報量規準 (Takeuchi Information Criterion, TIC) に相当する．また、

Tr は行列のトレースを表し, $I_Q(\theta)$ と $J_Q(\theta)$ はそれぞれ

$$I_Q(\theta) = E_Q \left[\frac{\partial \ln p(x|\theta)}{\partial \theta} \frac{\partial \ln p(x|\theta)}{\partial \theta^T} \right] \quad (10)$$

$$= \int_{-\infty}^{\infty} q(x) \frac{\partial \ln p(x|\theta)}{\partial \theta} \frac{\partial \ln p(x|\theta)}{\partial \theta^T} dx \quad (11)$$

$$J_Q(\theta) = -E_Q \left[\frac{\partial^2 \ln p(x|\theta)}{\partial \theta \partial \theta^T} \right] \quad (12)$$

$$= - \int_{-\infty}^{\infty} q(x) \frac{\partial^2 \ln p(x|\theta)}{\partial \theta \partial \theta^T} dx \quad (13)$$

で定義される $u \times u$ 行列である. ただし, T は転置を表す. ちなみに, ある θ_0 により真の分布 $q(x) = p(x|\theta_0)$ と書き表される場合, $I_Q(\theta_0) = J_Q(\theta_0)$ が成立し, 式 (9) は

$$\eta(\hat{\theta}) = \ell(\hat{\theta}) - \frac{u}{n} \quad (14)$$

と展開され, この平均対数尤度 $\eta(\hat{\theta})$ は赤池情報量規準 (Akaike Information Criterion, AIC) に相当し, 正規化 AIC と呼ぶ.

しかし, 真の分布 Q と真の最尤推定量 θ_0 を得ることはできない. そこで, 観測データに基づく経験分布 \hat{Q} と経験分布より得られた最尤推定量 $\hat{\theta}$ を用いて次のように求められる.

$$I_{\hat{Q}}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln p(x_i|\theta)}{\partial \theta} \frac{\partial \ln p(x_i|\theta)}{\partial \theta^T} \Bigg|_{\theta=\hat{\theta}} \quad (15)$$

$$J_{\hat{Q}}(\hat{\theta}) = - \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln p(x_i|\theta)}{\partial \theta \partial \theta^T} \Bigg|_{\theta=\hat{\theta}} \quad (16)$$

この結果, バイアス項を加えた対数尤度 $\ell'(\hat{\theta})$ は,

$$\ell'(\hat{\theta}) = \ell(\hat{\theta}) - \frac{1}{n} Tr(I_{\hat{Q}}(\hat{\theta})J_{\hat{Q}}(\hat{\theta})^{-1}) \quad (17)$$

で与えられ, 正規化 TIC と呼ぶ.

正規化 AIC には, モデル化した確率密度関数 p が真の確率密度関数を含まなければ精度が得られないという欠点があるが, 真の最尤推定量 θ_0 におけるバイアス項 $b(Q) = \ell(\hat{\theta}) - \eta(\hat{\theta})$ を与える. 一方, 正規化 TIC には, 真の分布 Q と真の最尤推定量 θ_0 が得られないため, 経験分布 \hat{Q} と経験分布より得られた最尤推定量 $\hat{\theta}$ を用いなくてはならない. また, バイアス項の推定による変動も考慮しなくてはならない. 実際の適用の際には, 以上の点を考慮に含めて適切な情報量規範を選択しなければならない.

以降の例では, 代表的な確率密度関数のモデルパラメータの最尤推定法について述べる. さらに, 各確率密度関数の正規化 TIC の式を与える.

例 1 指数分布 $p_{\text{exp}}(x|\lambda) = p(x|\lambda)$ について考える. 確率変数 X の定義域は $0 \leq X < \infty$ である.

$$p(x|\lambda) = \lambda \exp(-\lambda x) \quad (18)$$

$$\ln p(x|\lambda) = \ln \lambda - \lambda x \quad (19)$$

平均は $E[X] = 1/\lambda$, 分散は $E[(X - E[X])^2] = 1/\lambda^2$ である. このとき,

$$\frac{\partial \ln p(x|\lambda)}{\partial \lambda} = \frac{1}{\lambda} - x \quad (20)$$

$$\frac{\partial^2 \ln p(x|\lambda)}{\partial \lambda^2} = -\frac{1}{\lambda^2} \quad (21)$$

であり, $\partial \ell(\lambda)/\partial \lambda = 0$ を満足する最尤推定量 $\hat{\lambda}$ は,

$$\hat{\lambda} = n \left(\sum_{i=1}^n x_i \right)^{-1} \quad (22)$$

で与えられる. さらに,

$$I_{\hat{Q}}(\hat{\lambda}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\hat{\lambda}} - x_i \right)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{\hat{\lambda}^2} \quad (23)$$

$$J_{\hat{Q}}(\hat{\lambda}) = -\frac{1}{n} \sum_{i=1}^n \left(-\frac{1}{\hat{\lambda}^2} \right) = \frac{1}{\hat{\lambda}^2} \quad (24)$$

が導かれる. 正規化 TIC $\ell'(\hat{\lambda})$ は

$$\ell'(\hat{\lambda}) = \ell(\hat{\lambda}) - \frac{1}{n} I_{\hat{Q}}(\hat{\lambda}) J_{\hat{Q}}(\hat{\lambda})^{-1} \quad (25)$$

$$= \frac{1}{n} \sum_{i=1}^n (\ln \hat{\lambda} - \hat{\lambda} x_i) - \frac{1}{n} \left(\frac{\hat{\lambda}^2}{n} \sum_{i=1}^n x_i^2 - 1 \right) \quad (26)$$

$$= \ln \hat{\lambda} - \frac{n-1}{n} - \frac{\hat{\lambda}^2}{n^2} \sum_{i=1}^n x_i^2 \quad (27)$$

と計算される.

例 2 レイリー分布 $p_{\text{Rayleigh}}(x|\sigma^2) = p(x|\sigma^2)$ について考える.

$$p(x|\sigma^2) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (28)$$

$$\ln p(x|\sigma^2) = \ln\left(\frac{x}{\sigma^2}\right) - \frac{x^2}{2\sigma^2} \quad (29)$$

平均は $E[X] = \sigma\sqrt{\pi/2}$, 分散は $E[(X - E[X])^2] = (2 - \pi/2)\sigma^2$ である. 互いに独立で, 同一な, 平均 0, 分散 σ^2 を持つ 2 つのガウス確率変数 V, W の大きさ $X = \sqrt{V^2 + W^2}$ を与える確率分布としてよく知られている. また, 無線通信の分野ではレイリーフェージングを与える確率密度関数として用いられる. このとき,

$$\frac{\partial \ln p(x|\sigma^2)}{\partial \sigma^2} = -\frac{1}{\sigma^2} + \frac{x^2}{2\sigma^4} \quad (30)$$

$$\frac{\partial^2 \ln p(x|\sigma^2)}{(\partial \sigma^2)^2} = \frac{1}{\sigma^4} - \frac{x^2}{\sigma^6} \quad (31)$$

であり, $\partial \ell(\sigma^2)/\partial \sigma^2 = 0$ を満足する最尤推定量 $\hat{\sigma}^2$ は,

$$\hat{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^n x_i^2 \quad (32)$$

で与えられる. さらに,

$$I_{\hat{Q}}(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \left(-\frac{1}{\hat{\sigma}^2} + \frac{x_i^2}{2\hat{\sigma}^4} \right)^2 = \frac{1}{4n\hat{\sigma}^8} \sum_{i=1}^n x_i^4 - \frac{1}{\hat{\sigma}^4} \quad (33)$$

$$J_{\hat{Q}}(\hat{\sigma}^2) = -\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\hat{\sigma}^4} - \frac{x_i^2}{\hat{\sigma}^6} \right) = \frac{1}{\hat{\sigma}^4} \quad (34)$$

が導かれる．正規化 TIC $\ell'(\hat{\sigma}^2)$ は

$$\ell'(\hat{\sigma}^2) = \ell(\hat{\sigma}^2) - \frac{1}{n} I_{\hat{Q}}(\hat{\sigma}^2) J_{\hat{Q}}(\hat{\sigma}^2)^{-1} \quad (35)$$

$$= \frac{1}{n} \sum_{i=1}^n \left(\ln \left(\frac{x_i}{\hat{\sigma}^2} \right) - \frac{x_i^2}{2\hat{\sigma}^2} \right) - \frac{1}{n} \left(\frac{1}{4n\hat{\sigma}^4} \sum_{i=1}^n x_i^4 - 1 \right) \quad (36)$$

$$= \frac{1}{n} \sum_{i=1}^n \ln \left(\frac{x_i}{\hat{\sigma}^2} \right) - \frac{1}{4n^2\hat{\sigma}^4} \sum_{i=1}^n x_i^4 - \frac{n-1}{n} \quad (37)$$

と計算される．

例 3 ワイブル分布 $p_{\text{Weibull}}(x|\beta, \Omega) = p(x|\beta, \Omega)$ について考える．

$$p(x|\beta, \Omega) = \frac{\beta x^{\beta-1}}{\Omega} \exp \left(-\frac{x^\beta}{\Omega} \right) \quad (38)$$

$$\ln p(x|\beta, \Omega) = \ln \left(\frac{\beta}{\Omega} \right) + (\beta - 1) \ln x - \frac{x^\beta}{\Omega} \quad (39)$$

ここで， β はワイブル係数 (形状パラメータ)， Ω は尺度パラメータと呼ばれる．なお， $\alpha = \Omega^{1/\beta}$ とおきワイブル分布を β, α で定義される場合もある (MATLAB Statistics Toolbox, Excel の Weibull 関数など)．これは単純に表現上の問題であり， α も尺度パラメータと呼ばれる．なお， $\beta = 1$ のときワイブル分布 $p_{\text{Weibull}}(x|1, \Omega)$ は指数分布 $p_{\text{exp}}(x|\Omega^{-1})$ に一致し， $\beta = 2$ のときワイブル分布 $p_{\text{Weibull}}(x|2, \Omega)$ はレイリー分布 $p_{\text{Rayleigh}}(x|(\Omega/2))$ に一致する．

平均 $E[X]$ と分散 $E[(X - E[X])^2]$ は，

$$E[X] = \Omega^{1/\beta} \Gamma \left(1 + \frac{1}{\beta} \right) \quad (40)$$

$$E[(X - E[X])^2] = \Omega^{2/\beta} \left\{ \Gamma \left(1 + \frac{2}{\beta} \right) - \Gamma^2 \left(1 + \frac{1}{\beta} \right) \right\} \quad (41)$$

と計算される．ただし， $\Gamma(z)$ はガンマ関数を表し，実部が正の複素数 z に対して，

$$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt \quad (42)$$

と定義される．

各パラメータによる $\ln p(x|\beta, \Omega)$ の 1 階偏微分は

$$\frac{\partial \ln p(x|\beta, \Omega)}{\partial \beta} = \frac{1}{\beta} + \ln x - \frac{x^\beta \ln x}{\Omega} \quad (43)$$

$$\frac{\partial \ln p(x|\beta, \Omega)}{\partial \Omega} = -\frac{1}{\Omega} + \frac{x^\beta}{\Omega^2} \quad (44)$$

で与えられ，尤度方程式 $\partial \ell(\beta, \Omega) / \partial(\beta, \Omega) = \mathbf{0}$ は，

$$\frac{1}{\beta} + \frac{1}{n} \sum_{i=1}^n \ln x_i - \frac{1}{n\Omega} \sum_{i=1}^n x_i^\beta \ln x_i = 0 \quad (45)$$

$$\Omega - \frac{1}{n} \sum_{i=1}^n x_i^\beta = 0 \quad (46)$$

であり，式 (45)，(46) の解が最尤推定量 $\hat{\beta}$ ， $\hat{\Omega}$ となるが，明示的に与えることはできない．そこで，数値計算法による最尤推定量の導出が必要となる．式 (45)，(46) から次の関係式が導かれる．

$$\Omega^{-1} = n \left[\sum_{i=1}^n x_i^{\beta} \right]^{-1} \quad (47)$$

$$\beta = n \left[\sum_{i=1}^n (\Omega^{-1} x_i^{\beta} - 1) \ln x_i \right]^{-1} \quad (48)$$

式 (47)，(48) に基づく Ω^{-1} と β の反復法で最尤推定量 $\hat{\beta}$ ， $\hat{\Omega}$ を求めることができる．ワイブル分布の反復法のアルゴリズム [2] は次の通りである．

Step.1 (初期化)

x_1, x_2, \dots, x_n を入力データとする．

反復番号 $k \leftarrow 0$ ，収束判定値 $\varepsilon \leftarrow 10^{-7}$ ，反復回数の上限 $K \leftarrow 500$ ，

$\hat{\Omega}^{-1}$ の初期値 $a_k \leftarrow 1$ ， $\hat{\beta}$ の初期値 $b_k \leftarrow 1$ ．

Step.2 (更新)

反復番号のインクリメント $k \leftarrow k + 1$ ．

$$a_k \leftarrow n \left[\sum_{i=1}^n x_i^{b_{k-1}} \right]^{-1}, \quad b_k \leftarrow n \left[\sum_{i=1}^n (a_{k-1} x_i^{b_{k-1}} - 1) \ln x_i \right]^{-1} \quad (49)$$

Step.3 (収束判定)

$|a_k - a_{k-1}| < \varepsilon$ かつ $|b_k - b_{k-1}| < \varepsilon$ であれば， $\hat{\Omega} \leftarrow a_k^{-1}$ ， $\hat{\beta} \leftarrow b_k$ として終了．そうでなければ，Step.4 へ進む．

Step.4 (反復回数判定)

$k < K$ であれば，Step.2 へ戻る． $k = K$ のとき，エラーとして終了．

一方，各パラメータによる $\ln p(x|\beta, \Omega)$ の 2 階偏微分は

$$\frac{\partial^2 \ln p(x|\beta, \Omega)}{\partial \beta^2} = -\frac{1}{\beta^2} - \frac{x^{\beta} (\ln x)^2}{\Omega} \quad (50)$$

$$\frac{\partial^2 \ln p(x|\beta, \Omega)}{\partial \Omega^2} = \frac{1}{\Omega^2} - \frac{2x^{\beta}}{\Omega^3} \quad (51)$$

$$\frac{\partial^2 \ln p(x|\beta, \Omega)}{\partial \beta \partial \Omega} = \frac{x^{\beta} \ln x}{\Omega^2} \quad (52)$$

となる．以上より，

$$I_{\hat{Q}}(\hat{\beta}, \hat{\Omega}) = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{x_i^{\hat{\beta}}}{\hat{\Omega}}\right)^2 (\ln x_i)^2 - \frac{1}{\hat{\beta}^2} & \frac{1}{n\hat{\Omega}^2} \sum_{i=1}^n \left(1 - \frac{x_i^{\hat{\beta}}}{\hat{\Omega}}\right) x_i^{\hat{\beta}} \ln x_i + \frac{1}{\hat{\beta}\hat{\Omega}} \\ \frac{1}{n\hat{\Omega}^2} \sum_{i=1}^n \left(1 - \frac{x_i^{\hat{\beta}}}{\hat{\Omega}}\right) x_i^{\hat{\beta}} \ln x_i + \frac{1}{\hat{\beta}\hat{\Omega}} & \frac{1}{n\hat{\Omega}^4} \sum_{i=1}^n x_i^{2\hat{\beta}} - \frac{1}{\hat{\Omega}^2} \end{bmatrix} \quad (53)$$

$$J_{\hat{Q}}(\hat{\beta}, \hat{\Omega}) = \begin{bmatrix} \frac{1}{\hat{\beta}^2} + \frac{1}{n\hat{\Omega}} \sum_{i=1}^n x_i^{\hat{\beta}} (\ln x_i)^2 & -\frac{1}{\hat{\Omega}^2} \sum_{i=1}^n x_i^{\hat{\beta}} \ln x_i \\ -\frac{1}{\hat{\Omega}^2} \sum_{i=1}^n x_i^{\hat{\beta}} \ln x_i & \frac{1}{\hat{\Omega}^2} \end{bmatrix} \quad (54)$$

と展開でき，正規化 TIC $\ell'(\hat{\beta}, \hat{\Omega})$ は

$$\ell'(\hat{\beta}, \hat{\Omega}) = \ell'(\hat{\beta}, \hat{\Omega}) - \frac{1}{n} \text{Tr}(I_{\hat{Q}}(\hat{\beta}, \hat{\Omega}) J_{\hat{Q}}(\hat{\beta}, \hat{\Omega})^{-1}) \quad (55)$$

$$= \frac{1}{n} \sum_{i=1}^n \ln p(x_i | \hat{\beta}, \hat{\Omega}) - \frac{1}{n} \text{Tr}(I_{\hat{Q}}(\hat{\beta}, \hat{\Omega}) J_{\hat{Q}}(\hat{\beta}, \hat{\Omega})^{-1}) \quad (56)$$

と計算される．

例 4 平均 0 のガウス分布 $p_{\text{Gauss},0}(x|\sigma^2) = p(x|\sigma^2)$ について考える．

$$p(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (57)$$

$$\ln p(x|\sigma^2) = -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2} \quad (58)$$

平均 0，分散 σ^2 のガウス分布は受信機において通信信号に加わる電力 σ^2 の熱雑音のモデル化としてよく用いられる．

平均は $E[X] = 0$ であり，分散は $E[(X - E[X])^2] = \sigma^2$ である．このとき，

$$\frac{\partial \ln p(x|\sigma^2)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4} \quad (59)$$

$$\frac{\partial^2 \ln p(x|\sigma^2)}{(\partial \sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{x^2}{\sigma^6} \quad (60)$$

であり， $\partial \ell(\sigma^2)/\partial \sigma^2 = 0$ を満足する最尤推定量 $\hat{\sigma}^2$ は，

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad (61)$$

で与えられる．さらに，

$$I_{\hat{Q}}(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \left(-\frac{1}{2\hat{\sigma}^2} + \frac{x_i^2}{2\hat{\sigma}^4}\right)^2 = \frac{1}{4n\hat{\sigma}^8} \sum_{i=1}^n x_i^4 - \frac{1}{4\hat{\sigma}^4} \quad (62)$$

$$J_{\hat{Q}}(\hat{\sigma}^2) = -\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{2\hat{\sigma}^4} - \frac{x_i^2}{\hat{\sigma}^6}\right) = \frac{1}{2\hat{\sigma}^4} \quad (63)$$

が導かれる．正規化 TIC $\ell'(\hat{\sigma}^2)$ は

$$\ell'(\hat{\sigma}^2) = \ell(\hat{\sigma}^2) - \frac{1}{n} I_{\hat{Q}}(\hat{\sigma}^2) J_{\hat{Q}}(\hat{\sigma}^2)^{-1} \quad (64)$$

$$= \frac{1}{n} \sum_{i=1}^n \left(-\frac{1}{2} \ln(2\pi\hat{\sigma}^2) - \frac{x_i^2}{2\hat{\sigma}^2}\right) - \frac{1}{n} \left(\frac{1}{2n\hat{\sigma}^4} \sum_{i=1}^n x_i^4 - \frac{1}{2}\right) \quad (65)$$

$$= -\frac{1}{2} \ln(2\pi\hat{\sigma}^2) - \frac{1}{2n^2\hat{\sigma}^4} \sum_{i=1}^n x_i^4 - \frac{n-1}{2n} \quad (66)$$

と計算される．

例 5 平均 0 の 2 次混合ガウス分布 $p_{\text{GM},0}(x|P_0, \sigma_0^2, \sigma_1^2) = p(x|P_0, \sigma_0^2, \sigma_1^2)$ について考える．

$$p(x|P_0, \sigma_0^2, \sigma_1^2) = \frac{P_0}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) + \frac{1-P_0}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{x^2}{2\sigma_1^2}\right) \quad (67)$$

$$\ln p(x|P_0, \sigma_0^2, \sigma_1^2) = -\frac{1}{2} \ln(2\pi) + \ln\left(\frac{P_0}{\sigma_0} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) + \frac{1-P_0}{\sigma_1} \exp\left(-\frac{x^2}{2\sigma_1^2}\right)\right) \quad (68)$$

ただし, $0 \leq P_0 \leq 1$ である. 平均 0 の 2 次混合ガウス分布はインパルス性雑音の表現として用いられる. P_0 はインパルスが発生していない確率を表し, インパルスが発生していないときは平均電力 σ_0^2 を持つ背景ガウス雑音が現れていることが表現される. 一方, $1 - P_0$ はインパルスが発生している確率を表し, インパルスが発生しているときは平均電力 σ_0^2 を持つ背景ガウス雑音に平均電力 $\sigma_1^2 - \sigma_0^2$ を持つインパルスが現れていることが表現される. このとき, インパルスの希少性から $P_0 > 1/2$ 及びインパルス成分の追加から $\sigma_1^2 > \sigma_0^2$ という条件が自然に賦与される.

導関数を導くため, 次の関数を定義する. ただし, $\theta = (P_0, \sigma_0^2, \sigma_1^2)$ である.

$$\alpha(x|\sigma^2) = \frac{1}{\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (69)$$

$$\beta(x|\sigma_0^2, \sigma_1^2) = \alpha(x|\sigma_0^2) - \alpha(x|\sigma_1^2) \quad (70)$$

$$\gamma(x|\theta) = P_0\alpha(x|\sigma_0^2) + (1 - P_0)\alpha(x|\sigma_1^2) \quad (71)$$

このとき, 関数 $\alpha(x|\sigma^2)$ の σ^2 による導関数は,

$$\frac{\partial \alpha(x|\sigma^2)}{\partial \sigma^2} = \alpha(x|\sigma^2) \frac{\partial \ln \alpha(x|\sigma^2)}{\partial \sigma^2} = \alpha(x|\sigma^2) \left(-\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4}\right) \quad (72)$$

$$\frac{\partial^2 \alpha(x|\sigma^2)}{(\partial \sigma^2)^2} = \frac{\partial}{\partial \sigma^2} \left(\alpha(x|\sigma^2) \frac{\partial \ln \alpha(x|\sigma^2)}{\partial \sigma^2} \right) \quad (73)$$

$$= \alpha(x|\sigma^2) \left(\left(\frac{\partial \ln \alpha(x|\sigma^2)}{\partial \sigma^2} \right)^2 + \frac{\partial^2 \ln \alpha(x|\sigma^2)}{(\partial \sigma^2)^2} \right) \quad (74)$$

$$= \alpha(x|\sigma^2) \left(\left(-\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4} \right)^2 + \frac{1}{2\sigma^4} - \frac{x^2}{\sigma^6} \right) \quad (75)$$

$$= \alpha(x|\sigma^2) \left(\frac{3}{4\sigma^4} - \frac{3x^2}{2\sigma^6} + \frac{x^4}{4\sigma^8} \right) \quad (76)$$

となる.

平均は $E[X] = 0$ であり, 分散は $E[(X - E(X))^2] = P_0\sigma_0^2 + (1 - P_0)\sigma_1^2$ である. 各パラメータによる $\ln p(x|\theta)$ の 1 階偏微分は

$$\frac{\partial \ln p(x|\theta)}{\partial P_0} = \frac{\beta(x|\sigma_0^2, \sigma_1^2)}{\gamma(x|\theta)} \quad (77)$$

$$\frac{\partial \ln p(x|\theta)}{\partial \sigma_0^2} = \frac{P_0}{\gamma(x|\theta)} \frac{\partial \alpha(x|\sigma_0^2)}{\partial \sigma_0^2} = \frac{P_0\alpha(x|\sigma_0^2)}{\gamma(x|\theta)} \left(-\frac{1}{2\sigma_0^2} + \frac{x^2}{2\sigma_0^4} \right) \quad (78)$$

$$\frac{\partial \ln p(x|\theta)}{\partial \sigma_1^2} = \frac{1 - P_0}{\gamma(x|\theta)} \frac{\partial \alpha(x|\sigma_1^2)}{\partial \sigma_1^2} = \frac{(1 - P_0)\alpha(x|\sigma_1^2)}{\gamma(x|\theta)} \left(-\frac{1}{2\sigma_1^2} + \frac{x^2}{2\sigma_1^4} \right) \quad (79)$$

と展開される. 最尤推定量 $\hat{\theta}$ を尤度方程式 $\partial \ell(\theta)/\partial \theta = 0$ から導く必要があるが, ここではモーメント法 [3] による最尤推定量 $\hat{\theta}$ の導出方法について述べる.

最初に, 次の 3 つのモーメント a, b, c を導出する.

$$a = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{n} \sum_{i=1}^n |x_i| \quad (80)$$

$$b = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad (81)$$

$$c = \frac{\sqrt{2\pi}}{4} \cdot \frac{1}{n} \sum_{i=1}^n |x_i|^3 \quad (82)$$

このとき，標準偏差の最尤推定量は，

$$\hat{\sigma}_0 = \frac{ab - c + \sqrt{(ab - c)^2 - 4(a^2 - b)(b^2 - ac)}}{2(a^2 - b)} \quad (83)$$

$$\hat{\sigma}_1 = \frac{ab - c - \sqrt{(ab - c)^2 - 4(a^2 - b)(b^2 - ac)}}{2(a^2 - b)} \quad (84)$$

で与えられ，発生確率の最尤推定量は，

$$\hat{P}_0 = \frac{1}{2} \left(1 + \frac{2a^3 - 3ab + c}{\sqrt{(ab - c)^2 - 4(a^2 - b)(b^2 - ac)}} \right) \quad (85)$$

で与えられる．

一方，各モデルパラメータによる $\ln p(x|\boldsymbol{\theta})$ の 2 階偏微分は，

$$\frac{\partial^2 \ln p(x|\boldsymbol{\theta})}{\partial P_0^2} = -\frac{\beta(x|\sigma_0^2, \sigma_1^2)^2}{\gamma(x|\boldsymbol{\theta})^2} \quad (86)$$

$$\frac{\partial^2 \ln p(x|\boldsymbol{\theta})}{(\partial \sigma_0^2)^2} = \frac{P_0 \alpha(x|\sigma_0^2)}{\gamma(x|\boldsymbol{\theta})} \left(\frac{3}{4\sigma_0^4} - \frac{3x^2}{2\sigma_0^6} + \frac{x^4}{4\sigma_0^8} - \frac{P_0 \alpha(x|\sigma_0^2)}{\gamma(x|\boldsymbol{\theta})} \left(-\frac{1}{2\sigma_0^2} + \frac{x^2}{2\sigma_0^4} \right)^2 \right) \quad (87)$$

$$\frac{\partial^2 \ln p(x|\boldsymbol{\theta})}{(\partial \sigma_1^2)^2} = \frac{(1 - P_0) \alpha(x|\sigma_1^2)}{\gamma(x|\boldsymbol{\theta})} \left(\frac{3}{4\sigma_1^4} - \frac{3x^2}{2\sigma_1^6} + \frac{x^4}{4\sigma_1^8} - \frac{(1 - P_0) \alpha(x|\sigma_1^2)}{\gamma(x|\boldsymbol{\theta})} \left(-\frac{1}{2\sigma_1^2} + \frac{x^2}{2\sigma_1^4} \right)^2 \right) \quad (88)$$

$$\frac{\partial^2 \ln p(x|\boldsymbol{\theta})}{\partial P_0 \partial \sigma_0^2} = \frac{\alpha(x|\sigma_0^2)}{\gamma(x|\boldsymbol{\theta})} \left(-\frac{1}{2\sigma_0^2} + \frac{x^2}{2\sigma_0^4} \right) \left(1 - \frac{P_0 \beta(x|\sigma_0^2, \sigma_1^2)}{\gamma(x|\boldsymbol{\theta})} \right) \quad (89)$$

$$\frac{\partial^2 \ln p(x|\boldsymbol{\theta})}{\partial P_0 \partial \sigma_1^2} = -\frac{\alpha(x|\sigma_1^2)}{\gamma(x|\boldsymbol{\theta})} \left(-\frac{1}{2\sigma_1^2} + \frac{x^2}{2\sigma_1^4} \right) \left(1 + \frac{(1 - P_0) \beta(x|\sigma_0^2, \sigma_1^2)}{\gamma(x|\boldsymbol{\theta})} \right) \quad (90)$$

$$\frac{\partial \ln p(x|\boldsymbol{\theta})}{\partial \sigma_0^2 \partial \sigma_1^2} = -\frac{P_0(1 - P_0) \alpha(x|\sigma_0^2) \alpha(x|\sigma_1^2)}{\gamma(x|\boldsymbol{\theta})^2} \left(-\frac{1}{2\sigma_0^2} + \frac{x^2}{2\sigma_0^4} \right) \left(-\frac{1}{2\sigma_1^2} + \frac{x^2}{2\sigma_1^4} \right) \quad (91)$$

となる．以上を用いて，正規化 TIC $\ell'(\hat{\boldsymbol{\theta}})$ が数値計算できる．

3 まとめ

文献 [1] をもとに，バイアス項を含めた対数尤度である，正規化 TIC の計算式を与えた．代表的な確率密度関数である，指数分布，レイリー分布，ワイブル分布，平均 0 のガウス分布，平均 0 の 2 次混合ガウス分布に関して，観測系列からの最尤推定量の導出方法と正規化 TIC の計算方法を示した．

参考文献

- [1] 小西貞則, 北川源四郎, シリーズ 予測と発見の科学 情報量規準, 朝倉書店, 2004 年.
- [2] R による統計処理, <http://aoki2.si.gunma-u.ac.jp/R/>
- [3] K. Fukunaga and T. E. Flick, “Estimation of the parameters of a Gaussian mixture using the method of moments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 4, pp. 410–416, July 1983.