論　文

バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化するエントロピー推定量

志賀　元紀† 　横田　康成††

An Entropy Estimator That Minimizes Mean Squared Error under Condition of Restricted Averaged Squared Bias Error

Motoki SHIGA† and Yasunari YOKOTA††

あらまし　エントロピー推定は，様々な分野で有用な解析法として用いられている．一般に，離散的な記号を出力する独立生起情報源に対するエントロピーの推定は，各記号の生起確率の最ゆう推定値をエントロピー関数に代入することにより行われる．しかし，こうした推定量は，平均二乗誤差の最小性や不偏性をもたないことが示されている．更に，Paninski は，エントロピー推定においては，バイアス誤差と推定分散がトレードオフに制約されることを指摘した．本論文では，バイアス誤差の二乗平均と平均二乗誤差のトレードオフを最適に制御可能なエントロピー推定量を提案する．具体的には，エントロピー推定量が 1 変数関数の和で表されることを仮定し，バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小にするように導出される．提案法の推定精度と従来法の推定精度を比較した結果，提案法は，これら二つの誤差のトレードオフを制御することにより，従来法よりも高い推定精度が得られることが示された．

キーワード　エントロピー，エントロピー推定量，バイアス誤差，平均二乗誤差，トレードオフ

1. まえがき

エントロピー推定は，様々な分野で有用な解析法の一つとして用いられている．脳神経科学においても，近年，神経細胞から発生するスパイク列の情報量が推定され，これにより，神経細胞は，従来の予想よりも多くの情報を伝達していることが示された [1], [2]．こうしたことから，近年，エントロピーをより精度良く推定する手法が渇望されるようになった．

一般に，離散的な記号を出力する独立生起情報源に対するエントロピー推定は，観測された記号列から各記号の生起確率を最ゆう推定し，その値をエントロピー関数に代入することにより行われる．しかし，こうした一般的なエントロピー推定量は，平均二乗誤差を最小化しないこと [3], [4] や，不偏性を示さないこと [5], [6] が知られている．

そこで，Wolpert らは，生起確率の事前分布を一様分布と仮定し，エントロピーのベイズ推定量を導出した [3]．これは，平均二乗誤差を最小化する推定量となっている．また，横田は，エントロピー関数のテイラー展開を用いて，一般的なエントロピー推定量に含まれるバイアス誤差を近似的に求めた [6]．更に，バイアス誤差を抑制するため，定式化したバイアス誤差の符号を反転して一般的なエントロピー推定量に加えた推定量を提案した [6]．

推定問題においては，複数回の推定をそれぞれ独立な観測結果を用いて行い，得られた推定値の平均値を最終的な推定値とすることがある．前述した神経細胞からのスパイク列のエントロピー推定においても，こうした手法がとられている [2]．こうした手法では，仮に推定を無限回試行でき，各推定値が独立ならば，各推定にバイアス誤差がゼロとなる推定量，つまり不偏推定量を選択すれば，最終的な推定値の平均二乗誤差をゼロにすることができる．一方，1 回だけの推定では，平均二乗誤差を最小にする推定量が平均二乗誤差

† 岐阜大学工学研究科電子情報システム工学専攻，岐阜市
Graduate School of Engineering, Gifu University, 1–1 Yanagido, Gifu-shi, 501–1193 Japan
†† 岐阜大学工学部応用情報学科，岐阜市
Department of Information Science, Faculty of Engineering, Gifu University, 1–1 Yanagido, Gifu-shi, 501–1193 Japan

を最小にする意味で最も望ましい．試行回数が1回から無限回の間では，平均二乗誤差とバイアス誤差を同時に最小化できない，つまりこれらにトレードオフ関係があるならば，個々の推定においては，最終的な推定値の平均二乗誤差を最小化するためのバイアス誤差と平均二乗誤差の最適な配分比が存在する．しかも，最終的な平均二乗誤差を最小化するためには，個々の推定には，バイアス誤差を一定値以下に抑える条件下で平均二乗誤差を最小にする推定量を用いる必要がある．

本論文では，まずはじめに，離散的独立生起情報源に対し，1変数関数の和で表現可能なエントロピー推定量のクラスを仮定し，このクラスに属するすべての推定量のバイアス誤差と平均二乗誤差の一般式を導出する．このクラスには，一般的な推定量，ベイズ推定量 [3], [4]，バイアス誤差を抑制する推定量 [6]，ミニマックス推定量 [7] などかなり広範なエントロピー推定量が含まれる．このクラスの仮定のもとで，バイアス誤差の二乗平均を任意の値に制約する条件下で，平均二乗誤差を最小化する推定量を導出し，新たなエントロピー推定量として提案する．

2. 各種エントロピー推定法

2.1 一般的なエントロピー推定量

M 種類の情報源記号を，生起確率 $\boldsymbol{r} = (r_1, r_2, \ldots, r_M)$ で出力する独立生起情報源を仮定する．この情報源のエントロピーは，

$$H = \mathrm{ent}(\boldsymbol{r}) \tag{1}$$

で与えられる．ただし，

$$\mathrm{ent}(\boldsymbol{r}) \equiv -\sum_{i=1}^{M} r_i \log r_i$$

は，M 元エントロピー関数を表す．

情報源の真の生起確率 $\boldsymbol{r}$ が未知である場合に，この情報源から出力された N 個の標本を観測し，この情報源のエントロピーを推定する問題を考える．標本中に含まれている各情報源記号の数（以下，生起個数）が $\boldsymbol{n} = (n_1, n_2, \ldots, n_M)$ であったとき，一般的には，情報源記号の生起確率を $\hat{\boldsymbol{r}} = (n_1/N, n_2/N, \ldots, n_M/N)$ と最ゆう推定し，生起確率の真値に置き換えて，式 (1) に代入することにより，この情報源のエントロピーを

$$\hat{H}_{\mathrm{conv}}(\boldsymbol{n}) = \mathrm{ent}(\boldsymbol{n}/N) \tag{2}$$

と推定することが多い．本論文では，$\hat{H}_{\mathrm{conv}}(\boldsymbol{n})$ を一般的推定量と呼ぶ．

2.2 平均二乗誤差を最小にするエントロピー推定量

情報源記号の真の生起確率 $\boldsymbol{r}$ に関する事前知識がない場合，$\boldsymbol{r}$ の事前確率 $p(\boldsymbol{r})$ は，$\boldsymbol{r}$ によらず一定，つまり，

$$p(\boldsymbol{r}) = \begin{cases} \Gamma(M), & \boldsymbol{r} \in R, \\ 0, & \text{others} \end{cases} \tag{3}$$

と仮定される．ただし，$\Gamma(\cdot)$ は，ガンマ関数を表し，また，R は，$0 \le r_i \le 1, i = 1, 2, \ldots, M$ かつ $\sum_{i=1}^{M} r_i = 1$ を満たすベクトル $\boldsymbol{r} = (r_1, r_2, \ldots, r_M)$ の集合を表す．このとき，任意のエントロピー推定量 $\hat{H}(\boldsymbol{n})$ の平均二乗誤差 J_s は，

$$\begin{aligned} J_s &= \mathop{E}_{\boldsymbol{n},\boldsymbol{r}}[\{\hat{H}(\boldsymbol{n}) - \mathrm{ent}(\boldsymbol{r})\}^2] \\ &= \int \sum_{\boldsymbol{n} \in S_n} \left\{\hat{H}(\boldsymbol{n}) - \mathrm{ent}(\boldsymbol{r})\right\}^2 p(\boldsymbol{n}|\boldsymbol{r}) p(\boldsymbol{r}) d\boldsymbol{r} \\ &= \int \sum_{\boldsymbol{n} \in S} \left\{\hat{H}(\boldsymbol{n}) - \mathrm{ent}(\boldsymbol{r})\right\}^2 P_{N,\boldsymbol{r}}(\boldsymbol{n}) \Gamma(M) d\boldsymbol{r} \end{aligned} \tag{4}$$

で与えられる．ただし，$\mathop{E}_{\boldsymbol{n},\boldsymbol{r}}[\cdot]$ は，生起個数 $\boldsymbol{n}$ と生起確率 $\boldsymbol{r}$ の両者に関する期待値，S は $n_i \in \{0, 1, \ldots, N\}, i = 1, 2, \ldots, M$ かつ $\sum_{i=1}^{M} n_i = N$ を満たすベクトル $\boldsymbol{n} = (n_1, n_2, \ldots, n_M)$ の集合，$P_{N,\boldsymbol{r}}(\boldsymbol{n})$ は，多項分布

$$P_{N,\boldsymbol{r}}(\boldsymbol{n}) \equiv \frac{N!}{\prod_{i=1}^{M} n_i!} \prod_{j=1}^{M} r_j^{n_j}$$

を表す．また，式 (4) 中の積分の範囲は，集合 R 全域である．以降においても，積分範囲の記述がない場合，積分範囲を集合 R 全域とする．さて，平均二乗誤差 J_s を最小にするエントロピー推定量は，ベイズ推定により与えられ，

$$\begin{aligned} \hat{H}_{\mathrm{LMS}}(\boldsymbol{n}) &= \frac{\int \mathrm{ent}(\boldsymbol{r}) P_{N,\boldsymbol{r}}(\boldsymbol{n}) d\boldsymbol{r}}{\int P_{N,\boldsymbol{r}}(\boldsymbol{n}) d\boldsymbol{r}} \\ &= \sum_{i=1}^{M} \frac{n_i + 1}{N + M} \{\Psi(N + M + 1) \\ &\qquad - \Psi(n_i + 2)\} \end{aligned} \tag{5}$$

となる [3]．ただし，$\Psi(\cdot)$ はディガンマ関数を表す．本論文では，$\hat{H}_{\mathrm{LMS}}(\boldsymbol{n})$ を最小二乗推定量と呼ぶ．

2.3 バイアス誤差を抑制するエントロピー推定量

一般的推定量 $\hat{H}_{\mathrm{conv}}(\boldsymbol{n})$ のバイアス誤差は

$$\mathrm{Bias}_{\boldsymbol{r}}(\hat{H}_{\mathrm{conv}}) = \underset{\boldsymbol{n}|\boldsymbol{r}}{E}\left[\hat{H}_{\mathrm{conv}}(\boldsymbol{n})\right] - \mathrm{ent}(\boldsymbol{r}) \qquad (6)$$

と定義される．ただし，$\underset{\boldsymbol{n}|\boldsymbol{r}}{E}[\cdot]$ は，生起確率 $\boldsymbol{r}$ を条件とする生起個数 $\boldsymbol{n}$ に関する条件付期待値を表す．式 (6) で与えられるバイアス誤差を厳密に求めることは困難であるため，エントロピー関数 $\mathrm{ent}(\boldsymbol{r})$ のテイラー展開を用いて，近似的に定式化されている．これにより，一般的推定量 $\hat{H}_{\mathrm{conv}}(\boldsymbol{n})$ は，不偏性をもたないことが理論的に示されている [5], [6]．更に，横田は，L 次のテイラー展開を用いたバイアス誤差の近似 $B_{\mathrm{approx}}(\boldsymbol{r};L)$ を用いて，一般的推定量 $\hat{H}_{\mathrm{conv}}(\boldsymbol{n})$ に含まれるバイアス誤差を抑制するエントロピー推定量

$$\hat{H}_{\mathrm{Bias}}(\boldsymbol{n};L) = \hat{H}_{\mathrm{conv}}(\boldsymbol{n}) - B_{\mathrm{approx}}(\boldsymbol{n}/N;L)$$

を導出した [6]．本論文では，$\hat{H}_{\mathrm{Bias}}(\boldsymbol{n};L)$ をバイアス誤差抑制推定量と呼ぶ．

2.4 バイアス誤差と分散の和のミニマックス推定量

母数の推定問題において，その推定量のバイアス誤差と分散を同時に最小化できず，これらがトレードオフに制約されることがある．Paninsiki は，エントロピー推定量においても，こうしたトレードオフがあり，バイアス誤差を小さく抑える推定量は，推定分散が極めて大きくなることを指摘した [7]．

そこで，Paninsiki は，エントロピー推定量のバイアス誤差と推定分散の和の最大値を最小にする推定量を導出した [7]．この推定量は，バイアス誤差と分散のトレードオフを制御するパラメータ $\alpha, 0 \le \alpha \le 1$ をもち，次式で与えられる．

$$\hat{H}_{\mathrm{MM}}(\boldsymbol{n};\alpha) = \sum_{i=1}^{M} h_M(n_i;\alpha)$$

where,

$$\left(h_M(0;\alpha), h_M(1;\alpha), \ldots, h_M(N;\alpha)\right)^T = (\boldsymbol{X} + \alpha N \boldsymbol{Z})^{-1}\boldsymbol{y}$$

ただし，$\boldsymbol{X}$，$\boldsymbol{Z}$ は正方行列，$\boldsymbol{y}$ は列ベクトルであり，各要素は，それぞれ

$$\boldsymbol{X}_{n,m} = M^2 \int_0^1 B_{N,\boldsymbol{r}}(n) B_{N,\boldsymbol{r}}(m) dr = \frac{M^2 (N!)^2 (n+m)!(2N-n-m)!}{n!m!(N-n)!(N-m)!(2N+1)!},$$

$$\boldsymbol{y}_n = M^2 \int_0^1 B_{N,\boldsymbol{r}}(n)\,(-r\log r)\,dr = \frac{M^2(n+1)\left\{\Psi(N+2) - \Psi(n+2)\right\}}{(N+2)(N+1)}$$

$$\boldsymbol{Z} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 2 & -1 \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

で与えられる．また，$B_{N,r}(n)$ は，二項分布

$$B_{N,r}(n) \equiv \frac{N!}{n!(N-n)!} r^n (1-r)^{N-n}$$

を表す．本論文では，$\hat{H}_{\mathrm{MM}}(\boldsymbol{n})$ をミニマックス推定量と呼ぶ．

3. バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化するエントロピー推定量

本章では，バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化するエントロピー推定量を提案する．

3.1 制約付き誤差を最小にするエントロピー推定量の導出

バイアス誤差の二乗平均

$$J_b = \underset{\boldsymbol{r}}{E}\left[\left\{\underset{\boldsymbol{n}|\boldsymbol{r}}{E}\left[\hat{H}(\boldsymbol{n})\right] - \mathrm{ent}(\boldsymbol{r})\right\}^2\right] = \int \left\{\sum_{\boldsymbol{n}\in S_n} \hat{H}(\boldsymbol{n}) p(\boldsymbol{n}|\boldsymbol{r}) - \mathrm{ent}(\boldsymbol{r})\right\}^2 p(\boldsymbol{r}) d\boldsymbol{r} = \int \left\{\sum_{\boldsymbol{n}\in S_n} \hat{H}(\boldsymbol{n}) P_{N,\boldsymbol{r}}(\boldsymbol{n}) - \mathrm{ent}(\boldsymbol{r})\right\}^2 \Gamma(M) d\boldsymbol{r} \qquad (7)$$

を任意の値 β に制約する条件下で，平均二乗誤差 J_s を最小にする，すなわち，制約付き誤差

$$J = J_s + \lambda(J_b - \beta) \qquad (8)$$

を最小にする推定量 $\hat{H}_{\mathrm{opt}}(\boldsymbol{n};\lambda)$ を導出する．ただし，$\underset{\boldsymbol{r}}{E}[\cdot]$ は，生起確率 $\boldsymbol{r}$ に関する期待値，λ は，ラグランジュの未定乗数を表す．式 (4) と式 (7) を式 (8) に代入し，$\hat{H}_{\mathrm{opt}}(\boldsymbol{m};\lambda)$，$\boldsymbol{m}\in S_m$ で偏微分すると，

$$\frac{\partial J}{\partial \hat{H}_{\mathrm{opt}}(\boldsymbol{m};\lambda)} = 2\int\left\{\hat{H}_{\mathrm{opt}}(\boldsymbol{m};\lambda)-\mathrm{ent}(\boldsymbol{r})\right\}P_{N,\boldsymbol{r}}(\boldsymbol{m})\Gamma(M)d\boldsymbol{r} + 2\lambda\int\left\{\sum_{\boldsymbol{n}\in S_n}\hat{H}_{\mathrm{opt}}(\boldsymbol{n};\lambda)P_{N,\boldsymbol{r}}(\boldsymbol{n}) -\mathrm{ent}(\boldsymbol{r})\right\}P_{N,\boldsymbol{r}}(\boldsymbol{m})\Gamma(M)d\boldsymbol{r}$$

となる．ここで，右辺をゼロとおいて，式を整理すれば，連立方程式

$$\sum_{\boldsymbol{n}\in S_n}\left\{\frac{\lambda}{1+\lambda}\frac{\int P_{N,\boldsymbol{r}}(\boldsymbol{n})P_{N,\boldsymbol{r}}(\boldsymbol{m})d\boldsymbol{r}}{\int P_{N,\boldsymbol{r}}(\boldsymbol{m})d\boldsymbol{r}} + \frac{1}{1+\lambda}\delta_{\boldsymbol{n},\boldsymbol{m}}\right\}\hat{H}_{\mathrm{opt}}(\boldsymbol{n};\lambda) = \frac{\displaystyle\int \mathrm{ent}(\boldsymbol{r})P_{N,\boldsymbol{r}}(\boldsymbol{m})d\boldsymbol{r}}{\displaystyle\int P_{N,\boldsymbol{r}}(\boldsymbol{m})d\boldsymbol{r}},\quad \boldsymbol{m}\in S_m$$

が得られる．この式の両辺の積分を解けば，

$$\sum_{\boldsymbol{n}\in S_n}\left\{\frac{\lambda}{1+\lambda}\frac{N!(N+M-1)!}{(2N+M-1)!}\prod_{i=1}^{M}\frac{(n_i+m_i)!}{n_i!m_i!} + \frac{1}{1+\lambda}\delta_{\boldsymbol{n},\boldsymbol{m}}\right\}\hat{H}_{\mathrm{opt}}(\boldsymbol{n};\lambda) = \sum_{i=1}^{M}\frac{m_i+1}{N+M}\sum_{j=m_i+2}^{N+M}\frac{1}{j},\quad \boldsymbol{m}\in S_m \tag{9}$$

となる．ただし，$\delta_{\boldsymbol{n},\boldsymbol{m}}$ は，離散的デルタ関数

$$\delta_{\boldsymbol{n},\boldsymbol{m}} = \begin{cases}1, & \text{for}\quad \boldsymbol{n}=\boldsymbol{m},\\ 0, & \text{others}\end{cases}$$

を表す．したがって，推定量 $\hat{H}_{\mathrm{opt}}(\boldsymbol{n};\lambda)$ は，連立方程式 (9) の解として与えられる．

しかし，連立方程式 (9) に含まれる変数の個数は，$O(N^{M-1})$ となるため，記号数 M と標本数 N が大きいとき，連立方程式 (9) を解くことは現実的に不可能である．そこで，次節以降では，多くのエントロピー推定量が属する，1 変数関数の和で表現可能なエントロピー推定量のクラスを仮定し，式 (8) で与えられる制約付き誤差 J を最小にするエントロピー推定量を導出する．

3.2 1 変数関数の和で表現可能なエントロピー推定量の推定精度の一般式の導出

一般的な推定量，ベイズ推定量 [3], [4]，バイアス誤差を抑制する推定量 [6]，ミニマックス推定量 [7] は，いずれも適当な一つの 1 変数関数 $f(n), n\in\{0,1,\ldots,N\}$ を定め，それらに生起個数 $n_i, i=1,2,\ldots,M$ を代入したものの和

$$\hat{H}(\boldsymbol{n}) = \sum_{i=1}^{M} f(n_i) \tag{10}$$

で表されることに着目する．多くの有用なエントロピー推定量は，式 (10) の形で表現されるクラスに属すると思われる．そこで，まず，このクラスに属する任意のエントロピー推定量の平均二乗誤差 J_s^* とバイアス誤差の二乗平均 J_b^* の一般式を導出する．

式 (10) を，式 (4) の平均二乗誤差 J_s 及び式 (7) のバイアス誤差の二乗平均 J_b に代入して変形することにより，

$$J_b^* = \sum_{n_1=0}^{N}\sum_{n_2=0}^{N}A_{N,M}(n_1,n_2)f(n_1)f(n_2) + \sum_{n=0}^{N}B_{N,M}(n)f(n)+C_{N,M} \tag{11}$$

$$J_s^* = \sum_{n=0}^{N}D_{N,M}(n)\{f(n)\}^2 + \sum_{n_1=0}^{N}\sum_{n_2=0}^{N-n_1}E_{N,M}(n_1,n_2)f(n_1)f(n_2) + \sum_{n=0}^{N}F_{N,M}(n)f(n)+G_{N,M} \tag{12}$$

が得られる．式中の係数を，式 (13)～(19) に，また，導出過程の概要を付録 **1.**，**2.** に示した．ただし，$\Psi_1(\cdot)$ は，トリガンマ関数を表す．

3.3 バイアス誤差の二乗平均と平均二乗誤差のトレードオフを最適に制御可能なエントロピー推定量の導出

式 (11)，(12) を用いて，制約付き誤差

$$J^* = J_s^* + \lambda(J_b^* - \beta)$$

$$A_{N,M}(n_1,n_2)=\frac{N!N!(M-1)}{n_1!n_2!}\left\{\frac{(n_1+n_2)!(2N+M-n_1-n_2-2)!M}{(N-n_1)!(N-n_2)!(2N+M-1)!}\right.$$
$$\left.+\sum_{a=0}^{N-n_1}\sum_{b=0}^{N-n_2}\frac{(-1)^{a+b}M!(n_1+a)!(n_2+b)!}{a!(N-n_1-a)!b!(N-n_2-b)!(n_1+a+n_2+b+M-1)!}\right\} \tag{13}$$

$$B_{N,M}(n)=\frac{N!2(M-1)}{n!}\left[\frac{(n+1)!(N+M-n-2)!M}{(N-n)!(N+M)!}\Big\{\Psi(n+2)-\Psi(N+M+1)\Big\}\right.$$
$$\left.+\sum_{a=0}^{N-n}\frac{(-1)^a(n+a)!M!}{a!(N-n-a)!(n+a+M)!}\Big\{\Psi(2)-\Psi(n+a+M+1)\Big\}\right] \tag{14}$$

$$C_{N,M}=\frac{2}{M+1}\left[\Big\{\Psi(3)-\Psi(M+2)\Big\}^2+\Psi_1(3)-\Psi_1(M+2)\right]$$
$$+\frac{M-1}{M+1}\left[\Big\{\Psi(2)-\Psi(M+2)\Big\}^2-\Psi_1(M+2)\right] \tag{15}$$

$$D_{N,M}(n)=\frac{(N+M-n-2)!N!M(M-1)}{(N-n)!(N+M-1)!} \tag{16}$$

$$E_{N,M}(n_1,n_2)=\begin{cases}\dfrac{(N+M-n_1-n_2-3)!N!M(M-1)^2(M-2)}{(N-n_1-n_2)!(N+M-1)!} & \text{for } n_2\le N-n_1\\ 0 & \text{others.}\end{cases} \tag{17}$$

$$F_{N,M}(n)=\frac{2(n+1)D_{N,M}(n)}{(N+M)}\Big\{\Psi(n+2)-\Psi(N+M+1)\Big\}$$
$$+\sum_{k=0}^{N-n}\frac{2(k+1)E_{N,M}(n,k)}{(N+M)}\Big\{\Psi(k+2)-\Psi(N+M+1)\Big\} \tag{18}$$

$$G_{N,M}=\frac{2}{M+1}\left[\Big\{\Psi(3)-\Psi(M+2)\Big\}^2+\Psi_1(3)-\Psi_1(M+2)\right]$$
$$+\frac{M-1}{M+1}\left[\Big\{\Psi(2)-\Psi(M+2)\Big\}^2-\Psi_1(M+2)\right] \tag{19}$$

を最小にするように，式 (10) における 1 変数関数 $f(n), n\in\{0,1,\ldots,N\}$ を決定する．そこで，J^* を $f(m), m\in\{0,1,\ldots,N\}$ で偏微分して，右辺をゼロとおけば，連立方程式

$$2\sum_{n=0}^{N}\Big\{\lambda A_{N,M}(n,m)+D_{N,M}(n)\delta_{n,m}+E_{N,M}(n,m)\Big\}f(n)=-\lambda B_{N,M}(m)-F_{N,M}(m)\ \ ,m\in\{0,1,\ldots,N\} \tag{20}$$

が得られる．この連立方程式に含まれる変数の個数は $N+1$ 個であり，式 (9) で示される連立方程式の変数の個数 $O(N^{M-1})$ より少なくなる．したがって，連立方程式を解く計算量は大幅に削減され，現実的な問題として扱うことが可能になる．式 (20) を解くことにより得られる 1 変数関数 $f(n), n\in\{0,1,\ldots,N\}$ は，未定乗数 λ の関数であるため，これを改めて，$f(n;\lambda), n\in\{0,1,\ldots,N\}$ と書くことにする．本論文では，この連立方程式の解 $f(n;\lambda), n\in\{0,1,\ldots,N\}$ を用いて得られる推定量

$$\hat{H}_{\mathrm{proposed}}(\boldsymbol{n};\lambda)=\sum_{i=1}^{M}f(n_i;\lambda) \tag{21}$$

をエントロピー推定量として提案する．この推定量は，バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化することを保証している．未定乗数 λ の値により，バイアス誤差の二乗平均値と平均二乗誤差の配分比を制御可能である．

3.4 提案法による推定誤差の具体例

提案したエントロピー推定量 $\hat{H}_{\rm proposed}(\boldsymbol{n};\lambda)$ の平均二乗誤差とバイアス誤差の二乗平均の関係の具体例を示す．記号数 $M=5,25$，標本数 $N=50,250$ の各組合せ（4 通り）について，提案法 $\hat{H}_{\rm proposed}(\boldsymbol{n};\lambda), 0 \le \lambda < \infty$ のバイアス誤差の二乗平均 J_b^* と平均二乗誤差 J_s^* をそれぞれ式 (11)，式 (12) により算出した．得られたこれら 2 種類の誤差の関係を，図 1 に太い実線で示す．未定乗数 λ の値を調節することにより，推定精度は，この太い実線上で制御される．

参考のため，ミニマックス推定量 $\hat{H}_{\rm MM}(\boldsymbol{n};\alpha), 0 \le \alpha < \infty$，一般的推定量 $\hat{H}_{\rm conv}(\boldsymbol{n})$，最小二乗推定量 $\hat{H}_{\rm LMS}(\boldsymbol{n})$，バイアス誤差抑制推定量 $\hat{H}_{\rm Bias}(\boldsymbol{n};L), L=2,3,\ldots,6$ のそれぞれの推定精度を図 1 に点線，記号 □，記号 ○，記号 × で重ねて示した．ただし，一般的推定量 $\hat{H}_{\rm conv}(\boldsymbol{n})$ については，バイアス誤差の二乗平均 J_b^*，平均二乗誤差 J_s^* ともに極めて大きく，図 1 (c)，(d) に重ねて示すことはできなかった．

図 1 より，提案法は，バイアス誤差の二乗平均 J_b^* を任意の値以下に制約する条件下で，従来法よりも

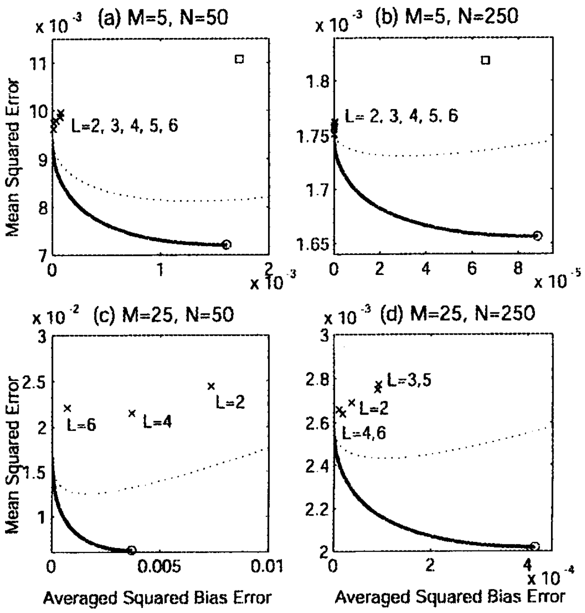

図 1　エントロピー推定量のバイアス誤差の二乗平均 J_b^* と平均二乗誤差 J_s^*（提案法：実線，ミニマックス推定量：点線，一般的推定量：記号 □，最小二乗推定量：記号 ○，バイアス誤差抑制推定量：記号 ×）

Fig. 1　Averaged squared bias error and mean squared error of the entropy estimated by proposed method (solid line), minimax method (dotted line), conventional method (symbol □), least squared method (symbol ○), bias improved method (symbol ×).

平均二乗誤差 J_s^* を小さく抑えられることが分かる．提案法では，未定乗数 λ の値を調節することにより，バイアス誤差の二乗平均 J_b^* と平均二乗誤差 J_s^* のトレードオフを最適に制御できるため，エントロピー推定の用途に応じて決まる二つの誤差に関する所望の条件下で推定を行うことが可能である．

4. 考　　察

4.1 バイアス誤差と平均二乗誤差のトレードオフ関係

本論文では，バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化するエントロピー推定量を導出した．ところで，Paninski は，エントロピー推定においては，バイアス誤差を抑えるほど推定分散が大きくなる，すなわちバイアス誤差と推定分散がトレードオフの関係にあることを指摘している [7]．図 1 に示した具体例を見る限り，バイアス誤差と平均二乗誤差の間においても，トレードオフの関係があるように見受けられる．本節では，バイアス誤差と平均二乗誤差のトレードオフ関係の有無について考察する．

まず，バイアス誤差の二乗平均 J_b と分散 J_v がトレードオフに制約される推定量があり，それが，バイアス誤差の二乗平均をある一定値 J_b 以下に制約する条件のもとで分散 J_v の最小化を達成しているものとする．この推定量の分散 J_v とバイアス誤差の二乗平均 J_b の関係を $J_v = g(J_b)$ と書くことにする．バイアス誤差の二乗平均 J_b と分散 J_v の間にトレードオフ関係があるとは，$g(J_{b_{min}}) > g(J_b)$ を満たす $J_b(>0)$ が存在することを意味する．ただし，$J_{b_{min}}$ は，この推定量が達成し得るバイアス誤差の最小値を表す．平均二乗誤差 J_s は分散 J_v とバイアス誤差の二乗平均 J_b の和であるから，この推定量の平均二乗誤差は，$J_s = g(J_b) + J_b$ と書ける．ゆえに，平均二乗誤差 J_s とバイアス誤差の二乗平均 J_b の間にトレードオフ関係があるための必要十分条件は，$g(J_{b_{min}}) > g(J_b) + J_b$ を満たす $J_b(>0)$ が存在することである．バイアス誤差の二乗平均 J_b は非負なので，バイアス誤差と平均二乗誤差の間にトレードオフ関係が存在するならば，バイアス誤差と分散の間にもトレードオフ関係があることになる．また，平均二乗誤差 J_s とバイアス誤差の二乗平均 J_b の間にトレードオフ関係があるための一つの十分条件は，$g(J_b)$ の導関数を $g'(J_b)$ と書けば，$\lim_{J_b \to J_{b_{min}}} g'(J_b) < -1$ である．エントロピー

推定においては，バイアス誤差を0に近づけようとすると，推定分散が急速に大きくなるため[7]，厳密に証明することはできないもののバイアス誤差と平均二乗誤差の間にトレードオフ関係があることはほぼ確実であると思われる．

4.2 バイアス誤差の二乗平均を任意の値以下に制約する条件のもとで分散を最小化する推定量との関係

バイアス誤差と平均二乗誤差の間にトレードオフ関係があるならば，そして，このトレード関係を満たす範囲で，バイアス誤差の二乗平均を任意の値に制約する条件下で分散を最小化する推定量は，バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化する推定量にもなっている．このことは，以下のように背理法により容易に証明できる．

ある推定量が，バイアス誤差の二乗平均 J_b を任意の値に制約する条件下で分散 J_v を最小化しているものとする．このとき，これらの関係が $J_v = g(J_b)$ で表されるとする．さて，この推定量の平均二乗誤差 J_s' が仮に $J_s' < g(J_b) + J_b$ であるとすると，分散は $J_v' = J_s' - J_b(< g(J_b))$ となり，$g(J_b)$ よりも分散を小さくすることができることになり矛盾を生ずる．同様に，バイアス誤差の二乗平均を任意の値に制約する条件下で平均二乗誤差を最小化する推定量は，バイアス誤差の二乗平均を任意の値に制約する条件下で分散を最小化する推定量であることもいえる．つまり，提案したエントロピー推定量は，バイアス誤差の二乗平均を任意の値以下に制約する条件のもとで分散を最小化する推定量にもなっているといえる．

5. むすび

本論文では，バイアス誤差の二乗平均を任意の許容量以下に抑える条件下で，平均二乗誤差を最小化することを保証するエントロピー推定量を提案した．具体的には，まずはじめに，1変数関数の和で表現可能なエントロピー推定量のクラスを仮定し，このクラスに属するすべての推定量のバイアス誤差と平均二乗誤差の一般式を導出した．このクラスには，一般的な推定量，ベイズ推定量[3], [4]，バイアス誤差を抑制する推定量[6]，ミニマックス推定量[7]などかなり広範なエントロピー推定量が含まれる．このクラスの仮定のもとで，バイアス誤差の二乗平均を任意の値に制約する条件下で，平均二乗誤差を最小化する推定量を導出した．

謝辞　本研究の一部は，文部科学省・平成11～15年度科学技術振興調整費「視覚系のニューロインフォマティクスに関する研究」，及び文部科学省・平成14～16年度科学研究費補助金（若手研究（B），No.14780634）の補助を受けて行われた．

文献

[1] R.R. de R. van Steveninck, G.D. Lewen, S.P. Strong, R. Koberle, and W. Bialek, "Reproducibility and variability in neural spike trans," Science, vol.275, no.21, pp.1805–1808, March 1997.

[2] P. Reinagel and R.C. Reid, "Temporal coding of visual information in the thalamus," J. Neuroscience, vol.20, no.14, pp.5392–5400, July 2000.

[3] D.H. Wolpert and D.R. Wolf, "Estimating functions of probability distributions from a finite set of samples," Phys. Rev. E, vol.52, no.6, pp.6841–6854, Dec. 1995.

[4] 横田康成，志賀元紀，"平均2乗誤差を改善するエントロピー推定量，" 信学論（A），vol.J86-A, no.9, pp.936–944, Sept. 2003.

[5] G. Miller, "Note on the bias of information estimates," Information theory in psychology: Problems and methods, ed. H. Quastler, pp.95–100, Free Press, Glencoe, 1955.

[6] 横田康成，"バイアス誤差を改善するエントロピーの推定量，" 信学論（D-II），vol.J86-D-II, no.1, pp.140–149, Jan. 2003.

[7] L. Paninski, "Estimation of entropy and mutual information," Neural Comput., vol.15, no.6, pp.1191–1253, 2003.

付　　録

1. 平均二乗誤差 J_s^* の導出

1変数関数の和で表現可能なエントロピー推定量の平均二乗誤差 J_b^* を導出する．ここで，$h(r) = -r\log r$ とおき，式(4)に式(10)を代入すると，

$$J_s = \int \sum_{\boldsymbol{n}\in S} P_{N,\boldsymbol{r}}(\boldsymbol{n}) \left[\sum_{i=1}^{M} \{ f(n_i) - h(r_i) \} \right]^2 \Gamma(M) d\boldsymbol{r}$$

となる．ここで，右辺の大括弧を展開すると，

$$J_s = \int \Gamma(M) \sum_{\boldsymbol{n}\in S} P_{N,\boldsymbol{r}}(\boldsymbol{n}) \left[\sum_{i=1}^{M} \Big\{ f(n_i) - h(r_i) \Big\}^2 + 2\sum_{j\neq k} \Big\{ f(n_j) - h(r_j) \Big\} \Big\{ f(n_k) - h(r_k) \Big\} \right] d\boldsymbol{r}$$

となり，大括弧の中の第1項は二項分布，第2項は変数の数が三つの多項分布として計算できる．更に，式を変形すれば，

$$J_s = M(M-1)\sum_{n=1}^{N}\int_0^1 (1-r_1)^{M-2} B_{N,r_1}(n)\Big\{f(n)-h(r_1)\Big\}^2 dr_1$$
$$+ M(M-1)^2(M-2)\sum_{n_1=0}^{N}\sum_{n_2=0}^{N-n_1}\int_0^1\int_0^{r_1} P'_{N,\boldsymbol{r}}(\boldsymbol{n})(1-r_1-r_2)^{M-3} \Big\{f(n_1)-h(r_1)\Big\}\Big\{f(n_2)-h(r_2)\Big\}dr_2dr_1$$
$$= \sum_{n_1=0}^{N}\frac{N!M(M-1)}{n_1!(N-n_1)!}\Bigg[\Big\{f(n_1)\Big\}^2\int_0^1 r_1^{n_1}(1-r_1)^{N+M-n_1-2}dr_1$$
$$+2f(n_1)\int_0^1 r_1^{n_1+1}(1-r_1)^{N+M-n_1-2}\log r_1 dr_1$$
$$+\int_0^1 r_1^{n_1+2}(1-r_1)^{N+M-n_1-2}\Big\{\log r_1\Big\}^2 dr_1\Bigg]$$
$$+\sum_{n_1=0}^{N}\sum_{n_2=0}^{N-n1}\frac{N!M(M-1)^2(M-2)}{n_1!n_2!(N-n_1-n_2)!}\Bigg[f(n_1)f(n_2)\int_0^1\int_0^{1-r_1} r_1^{n_1}r_2^{n_2} (1-r_1-r_2)^{N+M-n_1-n_2-3}dr_2dr_1$$
$$+2f(n_1)\int_0^1\int_0^{1-r_1} r_1^{n_1}r_2^{n_2+1} (1-r_1-r_2)^{N+M-n_1-n_2-3}\log r_2 dr_2dr_1$$
$$+\int_0^1\int_0^{1-r_1} r_1^{n_1+1}r_2^{n_2+1} (1-r_1-r_2)^{N+M-n_1-n_2-3}\log r_1\log r_2 dr_2dr_1\Bigg]$$

となる．式中の積分は，ディリクレ分布の規格化に関する等式

$$\int_{\boldsymbol{r}\in R}\frac{\Gamma(N+M)}{\prod_{i=1}^{M}\Gamma(n_i+1)}\prod_{i=1}^{M} r_i^{n_i}d\boldsymbol{r} = 1$$

を移項した等式

$$\int_{\boldsymbol{r}\in R}\prod_{i=1}^{M} r_i^{n_i}d\boldsymbol{r} = \frac{\prod_{i=1}^{M}\Gamma(n_i+1)}{\Gamma(N+M)}$$

の両辺を n_j で微分して得られる等式

$$\int_{\boldsymbol{r}\in R}\log r_j\prod_{i=1}^{M} r_i^{n_i}d\boldsymbol{r} = \frac{\prod_{i=1}^{M}\Gamma(n_i+1)}{\Gamma(N+M)}\Big\{\Psi(n_j+1)-\Psi(N+M)\Big\} \quad \text{(A·1)}$$

と，式 (A·1) の両辺 n_j で微分して得られる等式

$$\int_{\boldsymbol{r}\in R}\Big\{\log r_j\Big\}^2\prod_{i=1}^{M} r_i^{n_i}d\boldsymbol{r} = \frac{\prod_{i=1}^{M}\Gamma(n_i+1)}{\Gamma(N+M)}\Big[\Big\{\Psi(n_j+1)-\Psi(N+M)\Big\}^2 +\Psi_1(n_j+1)-\Psi_1(N+M)\Big] \quad \text{(A·2)}$$

と，式 (A·1) の両辺を $n_k, j\neq k$ で微分して得られる等式

$$\int_{\boldsymbol{r}\in R}\log r_j\log r_k\prod_{i=1}^{M} r_i^{n_i}d\boldsymbol{r} = \frac{\prod_{i=1}^{M}\Gamma(n_i+1)}{\Gamma(N+M)}\Big[\Big\{\Psi(n_j+1)-\Psi(N+M)\Big\} \Big\{\Psi(n_k+1)-\Psi(N+M)\Big\}-\Psi_1(N+M)\Big] \quad \text{(A·3)}$$

を利用して解くことができ，得られた式を変形することにより，式 (12) を導出できる．

2. バイアス誤差の二乗平均 J_b^* の導出

1 変数関数の和で表現可能なエントロピー推定量のバイアス誤差の二乗平均 J_b^* を導出する．式 (7) に式 (10) を代入すると，

$$J_b = \int\Bigg[\sum_{\boldsymbol{n}\in S}P_{N,\boldsymbol{r}}(\boldsymbol{n})\sum_{i=1}^{M}f(n_i)-\sum_{i=1}^{M}h(r_i)\Bigg]^2\Gamma(M)d\boldsymbol{r}$$

となり，右辺の大括弧を展開し，変形すると，

$$J_b = \Bigg[\int\sum_{i=1}^{M}\Bigg\{\sum_{n_i=0}^{N}B_{N,r_i}(n_i)f(n_i)-h(r_i)\Bigg\}^2 d\boldsymbol{r}$$
$$+2\int\sum_{j\neq k}\Bigg\{\sum_{n_j=0}^{N}B_{N,r_j}(n_j)f(n_j)-h(r_j)\Bigg\} \Bigg\{\sum_{n_k=0}^{N}B_{N,r_k}(n_k)f(n_k)h(r_k)\Bigg\}d\boldsymbol{r}\Bigg]\Gamma(M)$$

$$
\begin{aligned}
&= M(M-1)\Bigg[\sum_{n_1=0}^{N}\sum_{n_2=0}^{N}\frac{N!N!}{n_1!(N-n_1)!n_2!(N-n_2)!} \\
&\int_0^1 r_1^{n_1+n_2}(1-r_1)^{2N+M-n_1-n_2-2}dr_1 f(n_1)f(n_2) \\
&+\sum_{n_1=0}^{N}\frac{N!2}{n_1!(N-n_1)!} \\
&\int_0^1 r_1^{n_1+1}(1-r_1)^{N+M-n_1-2}\log r_1 dr_1 f(n_1) \\
&+\int_0^1 r_1^2(1-r_1)^{M-2}\left\{\log r_1\right\}^2 dr_1\Bigg] \\
&+M(M-1)^2(M-2)\Bigg[\sum_{n_1=0}^{N}\sum_{n_2=0}^{N} \\
&\frac{N!N!}{n_1!(N-n_1)!n_2!(N-n_2)!} \\
&\int_0^1\int_0^{1-r_1} r_1^{n_1}(1-r_1)^{N-n_1}r_2^{n_2}(1-r_2)^{N-n_2} \\
&(1-r_1-r_2)^{M-3}dr_2dr_1 f(n_1)f(n_2) \\
&+\sum_{n_1=0}^{N}\frac{N!2}{n_1!(N-n_1)!}\int_0^1\int_0^{1-r_1} r_1^{n_1}(1-r_1)^{N-n_1} \\
&r_2(1-r_1-r_2)^{M-3}\log r_2 dr_2 dr_1 f(n_1) \\
&+\int_0^1\int_0^{1-r_1} r_1 r_2(1-r_1-r_2)^{M-3} \\
&\log r_1 \log r_2 dr_2 dr_1\Bigg]
\end{aligned}
$$

となる．ここで，$(1-r_1)^{N-n_1}$ と $(1-r_2)^{N-n_2}$ を

$$
(1-r_1)^{N-n_1} = \sum_{a=0}^{N-n_1}\frac{(N-n_1)!}{a!(N-n_1-a)!}(-1)^a r_1^a
$$

及び

$$
(1-r_2)^{N-n_2} = \sum_{b=0}^{N-n_2}\frac{(N-n_2)!}{b!(N-n_2-b)!}(-1)^b r_2^b
$$

と展開し，得られた式に含まれる積分を式 (A·1), (A·2), (A·3) を利用して解き，更に式を変形することにより，式 (11) を導出できる．

（平成 16 年 8 月 18 日受付，11 月 24 日再受付，
12 月 14 日最終原稿受付）

志賀　元紀　（学生員）

平 13 岐阜大・工・応用情報卒．平 15 同大大学院博士前期課程了．現在，同博士後期課程在学中．非線形・非定常システムの解析，学習理論に関する研究に従事．

横田　康成　（正員）

平元豊橋技科大・工・情報卒．平 3 同大大学院修士課程了．平 6 同博士後期課程了．同年名工大・工・電気情報・助手．平 8 岐阜大・工・電子情報・講師．平 10 同大・工・応用情報・助教授．現在に至る．工博．非線形・非正規・非定常システムの同定と解析，符号化などに関する研究に従事．IEEE，計測自動制御学会，日本エム・イー学会各会員．