

論文

バイアス誤差を改善するエントロピーの推定量

横田 康成[†]

An Entropy Estimator Improving the Bias Error

Yasunari YOKOTA[†]

あらまし エントロピー推定は、信号処理、システム同定をはじめとする様々な分野で有用な解析法の一つとして行われている。しかし、一般的に利用されているエントロピー推定量は、実は不偏推定量にはなっていない。本論文では、エントロピー推定量の偏り、つまりバイアス誤差を、エントロピー関数のテイラー展開を利用して近似的に定式化した。この近似式を整理し、その中に含まれる係数を推定することにより、標本数のべき級数の形で表現される近似式を導出した。導出したバイアス誤差の近似式の符号を反転したものをエントロピー推定量に加えたものを新たなエントロピー推定量として提案した。提案したエントロピー推定量では、バイアス誤差が相殺されるため、その発生を抑えることが可能となり、従来のエントロピー推定量と比べ、統計的にも有意にバイアス誤差が減少した。

キーワード エントロピー推定, 不偏推定量, バイアス誤差, テイラー展開, 平均 2 乗誤差

1. ま え が き

エントロピーは、多くの研究者が認める基本的かつ重要な概念であり、その推定は、様々な分野で有用な解析法の一つとして行われている。例えば、最近、脳・神経科学の分野において、ある外界情報に対し、神経細胞が伝えているエントロピーの推定が行われ、その結果、神経細胞は従来の予想よりも多くの情報量を伝えていることが報告された [1], [2]。

有限種類の記号を独立に出力する情報源のエントロピーは、各記号の生起確率の対数の -1 倍の期待値と定義されている。エントロピーを各記号の生起確率の関数と見たものは、エントロピー関数と呼ばれる。各記号の生起確率が未知である場合、この情報源から出力された有限長の記号列を観測し、この記号列に含まれる各情報源記号の相対度数をその生起確率の推定値とし、この推定値をエントロピー関数に代入することにより、エントロピーを推定することが多い。しかし、こうしたエントロピー推定量は、実は、真のエントロピーに対する最小 2 乗推定量になっていないだけでなく [3], 不偏推定量にもなっていない [4]。前述し

たように、神経細胞が伝えている情報量を推定する場合 [1], [2] など、推定量を高め、あるいは低めに推定すると、その結果導かれる結論が大きく変わることになる。こうした場合、推定量の偏り、つまりバイアス誤差を抑制したエントロピー推定を行うことが強く要求される。こうしたことから、一般的なエントロピー推定量に比べ、バイアス誤差の少ないエントロピー推定量の開発が望まれている。

本論文では、まず、一般的に用いられているエントロピー推定量のバイアス誤差を、エントロピー関数のテイラー展開を利用して、近似的に定式化する。こうして得られるバイアス誤差の近似式はテイラー展開の打ち切り次数の変化に対して不安定であることから、この近似式を、観測した記号列の長さ、つまり標本数のべき級数として整理した式を新たに導く。この標本数のべき級数として表現されたバイアス誤差の近似式は、バイアス誤差を精度良く表現できることを示す。

更に、一般的に用いられているエントロピー推定量に、そのバイアス誤差を補償するように、導出されたバイアス誤差の符号を反転して加えたものを新たなエントロピー推定量として提案する。提案法では、テイラー展開により近似されたバイアス誤差を用いたエントロピー推定量に比べ、次数の変化に対して安定したエントロピー推定が行えることを示す。また、提案し

[†] 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

たエントロピー推定量では、従来の推定量よりも、統計的にも有意にバイアス誤差を減少させられることを示す。

2. エントロピー推定量のバイアス誤差

2.1 テイラー展開によるバイアス誤差の近似

M 種類の情報源記号 $a_i, i = 1, \dots, M$ をそれぞれ確率 $r_i, i = 1, \dots, M$ 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、

$$H = \text{ent}(r_1, \dots, r_M) \quad (1)$$

与えられる。ただし、 $\text{ent}(r_1, \dots, r_M)$ は多変数に拡張されたエントロピー関数であり、 $0 \leq r_i \leq 1$, $\sum_{i=1}^M r_i = 1$ なる $r_i, i = 1, \dots, M$ に対して、

$$\text{ent}(r_1, \dots, r_M) \equiv - \sum_{i=1}^M r_i \log_2 r_i$$

と定義される。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。一般に、情報源記号 $a_i, i = 1, \dots, M$ が、 N 個の標本中、 $N_i, i = 1, \dots, M$ 個ずつ含まれていたとき、情報源記号 a_i の生起確率の推定値を $\hat{r}_i = N_i/N$ とし、この情報源のエントロピーを

$$\hat{H} = \text{ent}(\hat{r}_1, \dots, \hat{r}_M) \quad (2)$$

として推定する。これを一般的なエントロピーの推定量と呼ぶことにし、まず、そのバイアス誤差

$$E[\hat{H}] - H$$

を定式化する。ただし、 $E[\cdot]$ は、 M, r_1, \dots, r_M, N を条件とする $\hat{M}, \hat{r}_1, \dots, \hat{r}_M$ に関する \cdot の条件付期待値を表す。これは、本来、 $E[\cdot]_{\hat{M}, \hat{r}_1, \dots, \hat{r}_M | M, r_1, \dots, r_M, N}$ と書くべきであるが、誤解するおそれがない限り、記述を簡素化するため、単に $E[\cdot]$ と表現する。

この情報源から観測された N 個の標本中に、情報源記号 a_1, \dots, a_M が、それぞれ N_1, \dots, N_M 個含まれる確率は、多項分布 $P_{r_1, \dots, r_M, N}(N_1, \dots, N_M)$ で与えられるから、エントロピーの推定量 \hat{H} の期待値は、

$$\begin{aligned} E[\hat{H}] &= E[\text{ent}(\hat{r}_1, \dots, \hat{r}_M)] \\ &= \sum_{(N_1, \dots, N_M) \in S} \text{ent}\left(\frac{N_1}{N}, \dots, \frac{N_M}{N}\right) \\ &\quad \cdot P_{r_1, \dots, r_M, N}(N_1, \dots, N_M) \end{aligned} \quad (3)$$

となる。ただし、 S は、 $N_i, i = 1, \dots, N$ のとり得るすべての組合せからなる集合

$$\begin{aligned} S &\equiv \{(N_1, \dots, N_M) \mid N_1 + \dots + N_M = N\} \\ N_i &\in \{0, 1, \dots, N\}, i = 1, \dots, M \end{aligned}$$

である。式 (3) を直接解くことは困難であるから、エントロピー関数のテイラー展開を利用して解くことにする。エントロピー関数 $\text{ent}(\hat{r}_1, \dots, \hat{r}_M)$ を情報源記号の生起確率の真値 $r_i, i = 1, \dots, M$ の周りでテイラー展開すると、

$$\begin{aligned} \text{ent}(\hat{r}_1, \dots, \hat{r}_M) &= \text{ent}(r_1, \dots, r_M) \\ &\quad + \sum_{j=1}^M \frac{\partial \text{ent}(r_1, \dots, r_M)}{\partial r_j} (\hat{r}_j - r_j) \\ &\quad + \sum_{j=1}^M \sum_{k=1}^M \frac{1}{2} \frac{\partial^2 \text{ent}(r_1, \dots, r_M)}{\partial r_j \partial r_k} \\ &\quad \cdot (\hat{r}_j - r_j)(\hat{r}_k - r_k) \\ &\quad + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \frac{1}{3!} \frac{\partial^3 \text{ent}(r_1, \dots, r_M)}{\partial r_i \partial r_j \partial r_k} \\ &\quad \cdot (\hat{r}_i - r_i)(\hat{r}_j - r_j)(\hat{r}_k - r_k) \\ &\quad + \dots \end{aligned} \quad (4)$$

となる。ただし、

$$\begin{aligned} \frac{\partial \text{ent}(r_1, \dots, r_M)}{\partial r_j} &= -\log_2 r_j - \frac{1}{\log 2} \\ \frac{\partial^2 \text{ent}(r_1, \dots, r_M)}{\partial r_j \partial r_k} &= -\frac{1}{r_j \log 2} \delta_{j,k} \\ \frac{\partial^3 \text{ent}(r_1, \dots, r_M)}{\partial r_i \partial r_j \partial r_k} &= \frac{1}{r_j^2 \log 2} \delta_{i,j,k} \\ &\vdots \\ \frac{\partial^n \text{ent}(r_1, \dots, r_M)}{\partial r_i \dots \partial r_l} &= \frac{(-1)^{n-1} (n-2)!}{r_j^{n-1} \log 2} \delta_{i, \dots, l} \\ &\vdots \end{aligned}$$

である。式 (4) を式 (3) に代入すれば、エントロピー推定におけるバイアス誤差は、

$$\begin{aligned} E[\hat{H}] - H &= \sum_{j=1}^M \left(-\log_2 r_j - \frac{1}{\log 2} \right) E[(\hat{r}_j - r_j)] \end{aligned}$$

$$\begin{aligned}
& + \sum_{j=1}^M \frac{1}{2} \frac{-1}{r_j \log 2} E[(\hat{r}_j - r_j)^2] \\
& + \sum_{j=1}^M \frac{1}{3!} \frac{1}{r_j^2 \log 2} E[(\hat{r}_j - r_j)^3] \\
& + \dots \\
& + \sum_{j=1}^M \frac{(-1)^{n-1}}{(n-1)n r_j^{n-1} \log 2} E[(\hat{r}_j - r_j)^n] \\
& + \dots
\end{aligned} \quad (5)$$

となる. 式 (5) より, 式 (2) で与えられるエントロピーの推定量の期待値 $E[\hat{H}]$ は, \hat{r}_j の真値 r_j の周りの中心モーメント $E[(\hat{r}_j - r_j)^k]$, $k = 2, 3, \dots$ の関数として表されることがわかる. $N_j = \hat{r}_j N$ は 2 項分布 $B_{r_j, N}(N_j)$, $j = 1, \dots, M$ に従う確率変数であるから, 式 (5) では, $\hat{r}_j N$, $j = 1, \dots, M$ を, 多項分布に従う確率変数としてではなく, それぞれ独立に 2 項分布に従う確率変数として扱うことが可能になっている.

ところで, 2 項分布 $B(r, N)$ に従う確率変数 \mathbf{X} の中心モーメント $\mu_k = E[(\mathbf{X} - rN)^k]$ に関し, 以下の漸化式が成り立つ.

$$\mu_{k+1} = r(1-r) \left(Nk\mu_{k-1} + \frac{d\mu_k}{dr} \right) \quad (6)$$

最初の数項 $\mu_0 = 1$, $\mu_1 = 0$, $\mu_2 = r(1-r)N, \dots$ をこの漸化式に代入することにより, \mathbf{X} の k 次中心モーメントは, k が偶数のとき, たかだか $N^{k/2}$ 次, k が奇数のとき, たかだか $N^{(k-1)/2}$ 次の N のべき多項式になることがわかる. \hat{r}_j は \mathbf{X}/N と考えられるから, \hat{r}_j の k 次中心モーメントは, \mathbf{X} の中心モーメント μ_k を N^k で除したものとなる. したがって, \hat{r}_j の k 次中心モーメントは, k が偶数のとき, たかだか $1/N^{k/2}$ 次, k が奇数のとき, たかだか $1/N^{(k+1)/2}$ 次のべき多項式になる. 具体的に, 最初の数項を書き下せば次式となる.

$$\begin{aligned}
E[\hat{r}_j] &= r_j \\
E[(\hat{r}_j - r_j)^2] &= \frac{1}{N} (r_j(1-r_j)) \\
E[(\hat{r}_j - r_j)^3] &= \frac{1}{N^2} (r_j(1-r_j)(1-2r_j)) \\
E[(\hat{r}_j - r_j)^4] &= \frac{1}{N^3} (r_j(1-r_j)) \\
&\quad \cdot (1+3r_j(1-r_j)(N-2)) \\
&\vdots
\end{aligned} \quad (7)$$

式 (7) を式 (5) に代入することにより, 式 (2) で与えられるエントロピー推定量 \hat{H} のバイアス誤差は,

$$\begin{aligned}
E[\hat{H}] - H &= \frac{1}{2N \log 2} \sum_{j=1}^M (r_j - 1) \\
&+ \frac{1}{6N^2 \log 2} \sum_{j=1}^M \frac{1}{r_j} (1 - 3r_j + 2r_j^2) \\
&- \frac{1}{12N^3 \log 2} \sum_{j=1}^M \frac{1}{r_j^2} (r_j - 1)(-3Nr_j + 3Nr_j^2 - 1 + 6r_j - 6r_j^2) \\
&- \frac{1}{20N^4 \log 2} \sum_{j=1}^M \frac{1}{r_j^3} (r_j - 1)(2r_j - 1)(10Nr_j^2 - 12r_j^2 - 10Nr_j + 12r_j - 1) \\
&+ \frac{1}{30N^5 \log 2} \sum_{j=1}^M \frac{1}{r_j^4} (r_j - 1)(1 - 30r_j - 30N^2r_j^3 + 15N^2r_j^4 - 130r_j^4N + 15N^2r_j^2 + 120r_j^4 + 25Nr_j - 240r_j^3 + 150r_j^2 - 155Nr_j^2 + 260Nr_j^3) \\
&+ \dots
\end{aligned} \quad (8)$$

として与えられる. 式 (8) の右辺で示されるバイアス誤差の各項は, いずれも $1/N^k$, $k \geq 1$ を含むから, バイアス誤差 $E[\hat{H}] - H$ は, $N \rightarrow \infty$ で 0 になる. したがって, 式 (2) で与えられるエントロピーの推定量 \hat{H} は, 漸近的には不偏推定量となるものの, 有限の標本数 N に対しては不偏性をもたないことがわかる.

式 (8) は, その右辺を N, r_j の次数ごとにまとめることにより, 次式で表現し直すことができる.

$$\begin{aligned}
E[\hat{H}] - H &= \frac{1}{\log 2} \sum_{m=2}^{\infty} \sum_{n=\lceil m/2 \rceil}^{m-1} \frac{1}{N^n} \sum_{k=-1}^{n-1} a_{m,n,k} \sum_{j=1}^M r_j^{-k}
\end{aligned} \quad (9)$$

ただし, $\lceil i \rceil$ は, i 以上の最小の整数を表す. また, $a_{m,n,k}$ は, テイラー展開 m 次の項における $N^{-n} \sum_{j=1}^M r_j^{-k}$ を含む項の係数であり, $m = 2, 3, \dots, 6$ についての値を表 1 に示した.

式 (9) の右辺において, $m \leq L$ で打ち切ったもの, つまり, テイラー展開を $L (\geq 2)$ 次までで打ち切るこ

表 1 係数 $a_{m,n,k}$ の値
Table 1 Value of coefficients $a_{m,n,k}$.

m	n	k	$a_{m,n,k}$	m	n	k	$a_{m,n,k}$
2	1	-1	1/2	6	3	-1	1/2
2	1	0	-1/2	6	3	0	-3/2
3	2	-1	1/3	6	3	1	3/2
3	2	0	-1/2	6	3	2	-1/2
3	2	1	1/6	6	4	-1	-13/3
4	2	-1	-1/4	6	4	0	13
4	2	0	1/2	6	4	1	-83/6
4	2	1	-1/4	6	4	2	6
4	3	-1	1/2	6	4	3	-5/6
4	3	0	-1	6	5	-1	4
4	3	1	7/12	6	5	0	-12
4	3	2	-1/12	6	5	1	13
5	3	-1	-1	6	5	2	-6
5	3	0	5/2	6	5	3	31/30
5	3	1	-2	6	5	4	-1/30
5	3	2	1/2				
5	4	-1	6/5				
5	4	0	-3				
5	4	1	5/2				
5	4	2	-3/4				
5	4	3	1/20				

とにより近似されたバイアス誤差を

$$A(r, N, L) = \frac{1}{\log 2} \sum_{m=2}^L \sum_{n=\lceil m/2 \rceil}^{m-1} \frac{1}{N^n} \sum_{k=-1}^{n-1} a_{m,n,k} \sum_{j=1}^M r_j^{-k} \quad (10)$$

とおく。ただし、 $r \equiv (r_1, \dots, r_M)$ とする。また、 $L=0$ における $A(r, N, L)$ は、バイアス誤差を 0 と近似した場合を意味するものとし、 $A(r, N, 0) = 0$ とする。

2.2 テイラー展開によるバイアス誤差の近似精度

テイラー展開によるバイアス誤差の近似精度を見るため、近似されたバイアス誤差 $A(r, N, L)$ と真のバイアス誤差 $E[\hat{H}] - H$ の差 $A(r, N, L) - (E[\hat{H}] - H)$ を様々な r, N, L に対して調べた。例として、 $r = (0.01, 0.99)$, $r = (0.1, 0.9)$, $r = (0.5, 0.5)$, $r = (0.01, 0.02, 0.97)$ とした際のバイアス誤差の近似誤差 $A(r, N, L) - (E[\hat{H}] - H)$ を図 1(a)~(d) にそれぞれ示す。各図とも、標本数 N の関数として、 $L=0, 2, 3, 4, 5, 6$ に対する結果をそれぞれ記号 \circ , \square , \diamond , $*$, $+$, \times を付した実線で示した。ただし、 $L=0$ では、バイアス誤差 $E[\hat{H}] - H$ そのもの (の -1 倍) を表す。これより、 $r = (0.5, 0.5)$ を除き、次数 $L=4$ で最も近似精度が高くなり、これ以上に次数 L を上げてもバイアス誤差の近似精度が高ならないことがわ

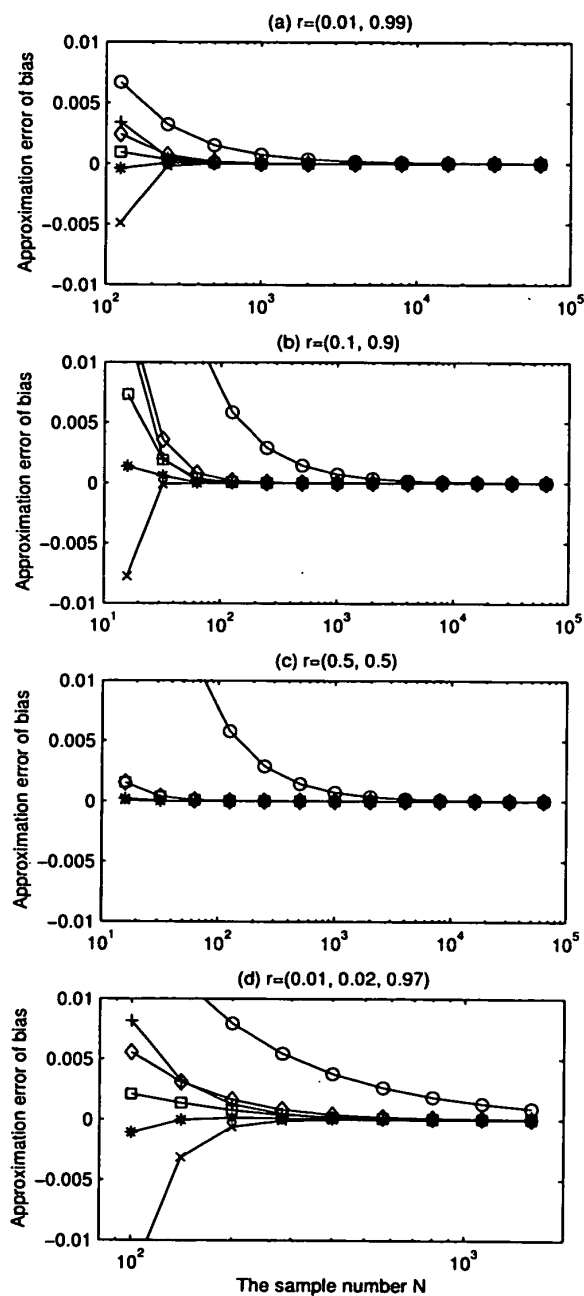


図 1 テイラー展開によるバイアス誤差の近似誤差
Fig. 1 Approximation error of bias error by Taylor expansion.

かる。打ち切り次数 L の増加とともにバイアス誤差の近似精度が高ならないことは一見不自然に感じられるかもしれないが、これが不自然ではないことの理由は、以下のように説明できる。

一般に、関数 $y = f(x)$ において、 $R_n = \frac{h^n}{n!} f^{(n)}(x_0 + \theta h)$ として、十分小さな h に対し、 $\lim_{n \rightarrow \infty} |R_n| = 0$, $0 < \theta < 1$ となるならば、関数 $y = f(x)$ の $x = x_0$ の周りでのテイラー展開は

$f(x_0 + h)$ の近似といわれ、 $|R_n|$ をその誤差という。実際、関数 $-\hat{r}_j \log \hat{r}_j$ については、 r_j , $0 < r_j \leq 1$ の周りでのテイラー展開は、この関数の近似となっている。しかしながら、テイラー展開による関数近似で保証されるのは、十分小さな h 、つまり、 r_j の近傍での近似精度であり、与えられた範囲 $0 \leq \hat{r}_j \leq 1$ 全体での近似精度、例えば平均 2 乗誤差最小の意味での精度を保証するものではない。標本数 N が小さいときには、生起確率の推定値 \hat{r}_j は、真値 r_j から離れて幅広く分布するため、バイアス誤差を近似する際に、関数 $-\hat{r}_j \log \hat{r}_j$ のテイラー展開による近似の精度が保証されない範囲を使うことも起こり得る。したがって、図 1 に示したように、バイアス誤差の近似精度が、打ち切り次数 L の増加とともに減少しないことは何ら不自然なことではない。

2.3 べき級数によるバイアス誤差の近似

2.2 で述べたとおり、テイラー展開を用いたバイアス誤差の近似 (式 (10)) では、 $L = 4$ において最大の近似精度が得られるものの、安定した近似であるとはいえない。ところで、式 (9) は、右辺を標本数 N のべき項でまとめることにより、次式で表現し直すことができる。

$$E[\hat{H}] - H = \frac{1}{\log 2} \sum_{n=1}^{\infty} \frac{1}{N^n} \sum_{k=-1}^{n-1} b_{n,k} \sum_{j=1}^M r_j^{-k} \quad (11)$$

上式において、 $b_{n,k}$ は、 $N^{-n} \sum_{j=1}^M r_j^{-k}$ が含まれるすべての項をまとめて得られる項の係数を意味する。 $n = 1, \dots, 5$ についての係数 $b_{n,k}$ の値を表 2 に示す。

さて、式 (11) は、適当な N の次数で打ち切った場合、バイアス誤差の近似式とみなすことができる。しかし、各次数のべき関数が直交していないため、こうして得られるバイアス誤差の近似式は、バイアス誤差の近似の平均 2 乗誤差を最小にしている保証はない。そこで、式 (11) の右辺を $n \leq K$ で打ち切ったものを

N のべき級数とみなし、

$$B(r, N, K) = \sum_{n=1}^K \frac{1}{N^n} \sum_{k=-1}^{n-1} b_{n,k,K} \sum_{j=1}^M r_j^{-k} \quad (12)$$

と表現する。係数 $b_{n,k,K}$ は、バイアス誤差の真値 $(E[\hat{H}] - H)$ を最もよく表現するような $B(r, N, K)$ を求めたいのであるから、次式で与えられる近似誤差の平均 2 乗値 J を最小にするように決定する。

$$J = E_{M,r,N} \{ [B(r, N, K) - (E[\hat{H}] - H)]^2 \} \quad (13)$$

式 (12) を式 (13) に代入することにより、

$$J = E_{M,r,N} \left[\left(\sum_{n=1}^K \frac{1}{N^n} \sum_{k=-1}^{n-1} b_{n,k,K} \sum_{j=1}^M r_j^{-k} - (E[\hat{H}] - H) \right)^2 \right] \quad (14)$$

が得られる。ここで、

$$R_{n,k} \equiv \frac{1}{N^n} \sum_{j=1}^M r_j^{-k}$$

$$S \equiv E[\hat{H}] - H$$

とおけば、式 (14) は、

$$J = E_{M,r,N} \left[\left(\sum_{n=1}^K \sum_{k=-1}^{n-1} b_{n,k,K} R_{n,k} - S \right)^2 \right] \quad (15)$$

と書ける。期待値 $E[\cdot]$ は、情報源記号の数、各情報源記号の生起確率、標本数の分布、すなわち対象に依存する。そこで、 M, r_1, \dots, r_M, N の組を適当な分布に従うように多数用意し、それらに対する集合平均で期待値 $E[\cdot]$ を置き換えるものとする。 J を最小にする係数 $b_{n,k,K}$ は、それが線形パラメータであることから、正規方程式を解くことにより容易に算出可能である。また、 $E[\hat{H}]$ は、式 (3) により、多項分布を利用して計算できる。

具体的に、 $K = 1, \dots, 5$ のそれぞれについて、 $M \in \{2, 3\}$, $0 \leq r_i \leq 1$ ($\sum_{i=1}^M r_i = 1$), $6 \leq N \leq 1500$ の範囲でそれぞれ独立に一樣分布に従う乱数により M, r_1, \dots, r_M, N の組を 3 万組生成し、それらを用いて係数 $b_{n,k,K}$ を決定した。その結果を小数点以下 3 けたを四捨五入して表 3 に示す。この表で示した係数 $b_{n,k,K}$ をもつ式 (12) をべき級数による

表 2 係数 $b_{n,k}$, $n = 1, \dots, 6$ の値
Table 2 Value of coefficients $b_{n,k}$, $n = 1, \dots, 6$.

n	$k = -1$	0	1	2	3	4	5
1	1/2	-1/2	0	0	0	0	0
2	1/12	0	-1/12	0	0	0	0
3	0	0	1/12	-1/12	0	0	0
4	-1/120	0	-1/12	1/4	-19/120	0	0
5	0	0	1/12	-7/12	19/20	-9/20	0
6	1/252	0	-1/12	5/4	-95/24	9/2	-863/504

表 3 決定された係数 $b_{n,k,K}$
Table 3 The determined coefficients $b_{n,k,K}$.

	$K = 2$	3	4	5
$b_{2,-1,K}$	0.35	-0.05	-0.07	0.22
$b_{2,0,K}$	-0.25	0.31	0.17	-0.18
$b_{2,1,K}$	-0.10	-0.27	-0.13	-0.05
$b_{3,-1,K}$		-0.11	0.11	-0.16
$b_{3,0,K}$		0.43	0.82	-1.13
$b_{3,1,K}$		-0.56	-0.39	2.22
$b_{3,2,K}$		0.18	-0.31	-0.92
$b_{4,-1,K}$			-1.00	-2.07
$b_{4,0,K}$			1.08	3.48
$b_{4,1,K}$			-1.02	2.01
$b_{4,2,K}$			0.11	-5.22
$b_{4,3,K}$			0.34	1.59
$b_{5,-1,K}$				2.41
$b_{5,0,K}$				-0.85
$b_{5,1,K}$				-1.96
$b_{5,2,K}$				-1.22
$b_{5,3,K}$				3.04
$b_{5,4,K}$				-0.73

バイアス誤差の近似式と呼ぶことにする。便宜上、 $K=0$ における $B(r, N, K)$ を、バイアス誤差を 0 と近似した場合、つまり $B(r, N, 0) = 0$ とおく。また、表 3 には示していないが、 $b_{1,-1,K} \cong 0.5$, $b_{1,0,K} \cong -0.5$, $K = 1, 2, 3, \dots$ であつたため、 $b_{1,-1,K} = 1/2$, $b_{1,0,K} = -1/2$, $K = 1, 2, 3, \dots$ とおいた。したがって、 $B(r, N, K)|_{K=1} = A(r, N, L)|_{L=2}$ となる。

2.4 べき級数によるバイアス誤差の近似精度

べき級数によるバイアス誤差の近似精度を見るため、近似されたバイアス誤差 $B(r, N, K)$ と真のバイアス誤差 $E[\hat{H}] - H$ の差 $B(r, N, K) - (E[\hat{H}] - H)$ を様々な r, N, K に対して調べた。例として、 $r = (0.01, 0.99)$, $r = (0.1, 0.9)$, $r = (0.5, 0.5)$, $r = (0.01, 0.02, 0.97)$ とした際のバイアス誤差の近似誤差 $B(r, N, K) - (E[\hat{H}] - H)$ を図 2(a)~(d) にそれぞれ示す。各図とも、標本数 N の関数として、 $K = 0, 1, 2, 3, 4, 5$ に対する結果をそれぞれ記号 $\circ, \square, \diamond, *, +, \times$ を付した実線で示した。図 1 に示したテイラー展開を利用したバイアス誤差の近似とは異なり、次数 $K = 4$ で真のバイアス誤差をほぼ表現できており、かつ $K \geq 4$ としても、近似精度が悪化するという不安定性もないことがわかる。

3. バイアス誤差を改善するエントロピー推定量

3.1 推定量の定式化と推定精度評価

2.2 で導出したバイアス誤差の近似式 (式 (12)) を

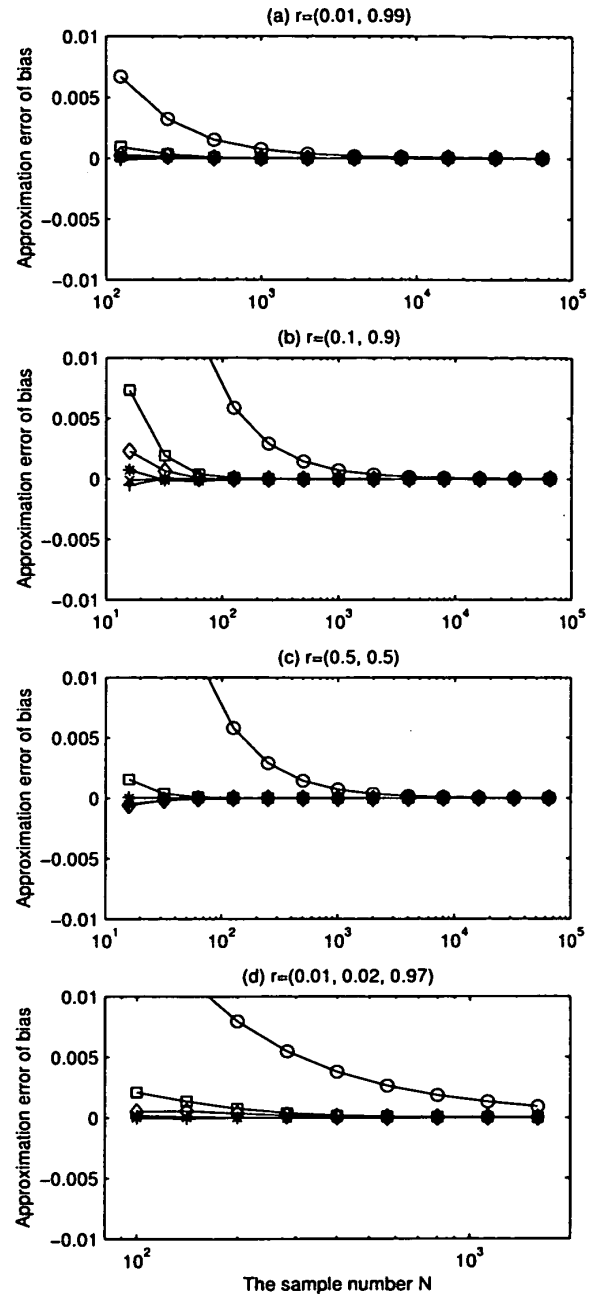


図 2 テイラー展開後、 N のべきの項にまとめることにより定式化されたバイアス誤差の近似精度

Fig. 2 Approximation accuracy of bias error formulated by gathering terms with equal power of N after Taylor expanding.

一般的なエントロピー推定量から減じたものは、バイアス誤差を打ち消し、その発生を抑制したエントロピー推定量になることが期待される。しかし、バイアス誤差の近似式は、情報源記号の真の生起確率 $r = (r_1, \dots, r_M)$ の関数となっており、実際には、真の生起確率は未知であるから、これを推定値に置き換えなければならない。そこで、バイアス誤差の近似式

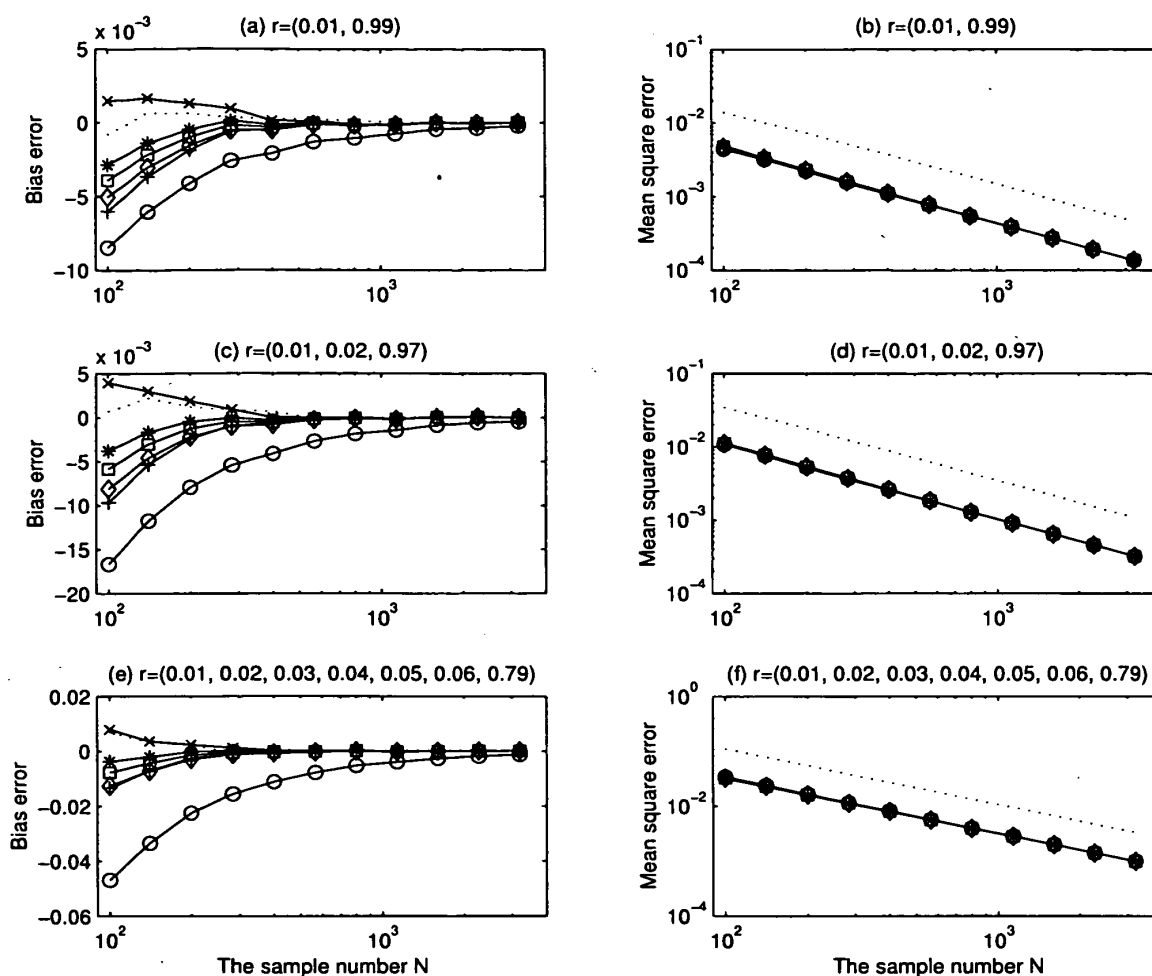


図3 テイラー展開法によるエントロピー推定量のバイアス誤差と平均2乗誤差
Fig. 3 Bias error and mean square error of entropy estimator by Taylor expansion method.

$B(r, N, K)$ において、情報源記号の真の生起確率 r をその推定値 $\hat{r} \equiv (\hat{r}_1, \dots, \hat{r}_M)$ に置き換えて得られる。

$$\hat{H}_K^{(P)} = \hat{H} - B(\hat{r}, N, K) \quad (16)$$

を新たなエントロピー推定量として提案する。ただし、 \hat{M} は、実際の標本中に現れた情報源記号の種類の数を表す。比較のため、べき級数による近似式 $B(\hat{r}, N, K)$ の代わりにテイラー展開による近似式 $A(\hat{r}, N, L)$ を用いた推定量

$$\hat{H}_L^{(T)} = \hat{H} - A(\hat{r}, N, L) \quad (17)$$

についても検討する。便宜上、 $\hat{H}_K^{(P)}$ 、 $\hat{H}_L^{(T)}$ として与えられるエントロピー推定法を、それぞれべき級数法、テイラー展開法と呼ぶことにする。

情報源記号の真の生起確率をそれらの推定値に置き換えることによるエントロピー推定精度への影響を調べるため、また、 $M = 2, 3$ の範囲で推定

された表3に示される係数 $b_{n,k,K}$ が $M > 3$ に対しても適用可能か否かを調べるため、数値実験を行った。 $r = (0.01, 0.99)$, $(0.01, 0.02, 0.97)$, 及び $(0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.79)$ なる生起確率をもつ情報源記号を出力する独立生起情報源から、 N , $100 \leq N \leq 4000$ 個の標本を観測し、それらの標本から、二つの手法でそれぞれエントロピーを推定する。こうした試行を100万回繰り返し、それぞれのエントロピー推定値の真値からのずれの平均値、つまりバイアス誤差、及び真値の周りの分散、すなわち平均2乗誤差を算出した。テイラー展開法、べき級数法のそれぞれに対する結果を、図3、図4にそれぞれ示す。それぞれの図の左側の列は、エントロピー推定値のバイアス誤差（図中、Bias error）、右側の列は、平均2乗誤差（図中、Mean square error）を表す。図3中、記号 \circ , \square , \diamond , $*$, $+$, \times を付した実線は、それぞれ推定量 $\hat{H}_L^{(T)}$, $L = 0, 2, 3, 4, 5, 6$ に対する結果、図4中、

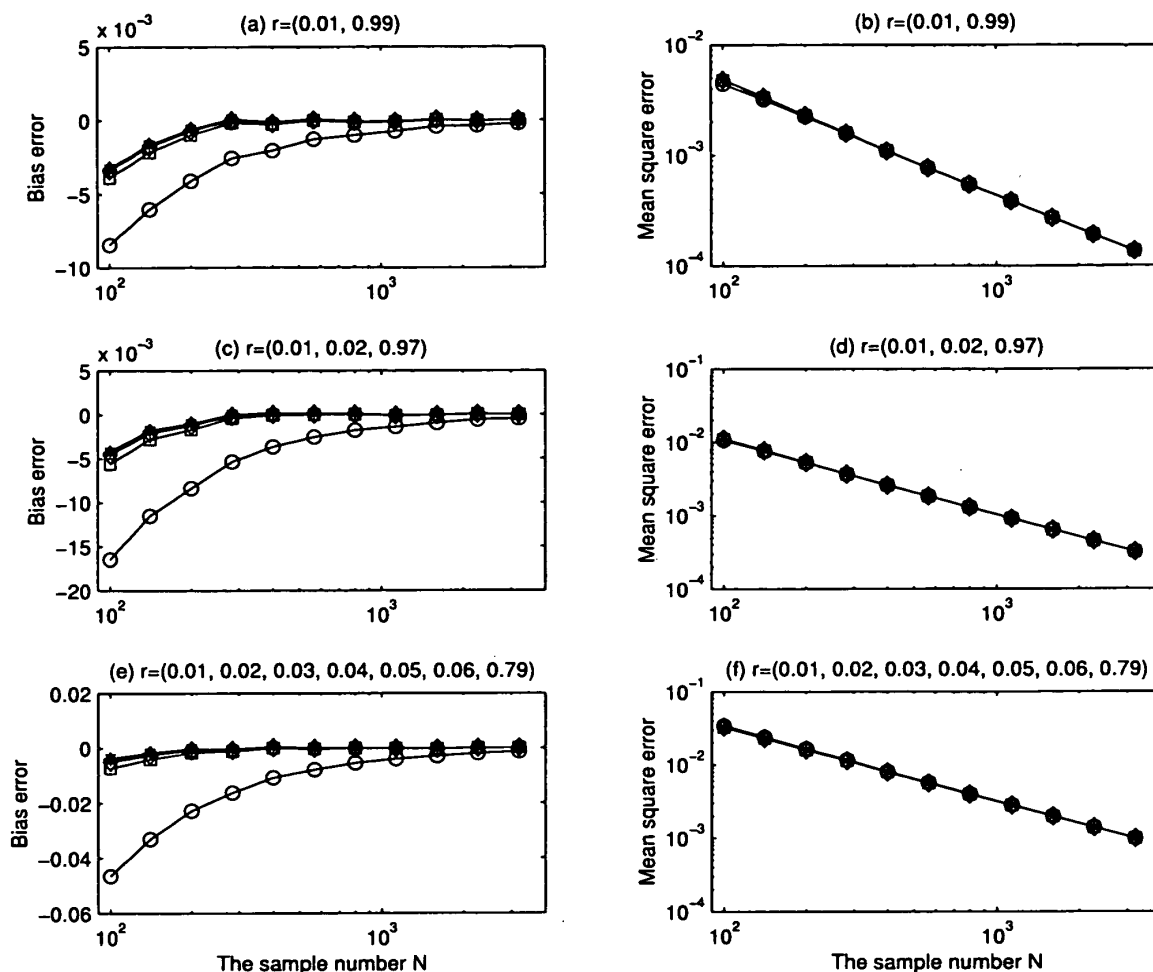


図4 ベキ級数法によるエントロピー推定量のバイアス誤差と平均2乗誤差
Fig.4 Bias error and mean square error of entropy estimator by power series method.

記号 \circ , \square , \diamond , $*$, $+$, \times を付した実線は、それぞれ推定量 $\hat{H}_K^{(P)}$, $K = 0, 1, 2, 3, 4, 5$ に対する結果を表している。ただし、 $\hat{H}_0^{(T)} = \hat{H}_0^{(P)} = \hat{H}$, $\hat{H}_2^{(T)} = \hat{H}_1^{(P)}$ である。これらの図より、以下のことがわかる。(1) いずれのエントロピー推定量を用いても、平均2乗誤差はほとんど変化しない。(2) エントロピー推定量 $\hat{H}_L^{(T)}$ では、次数 L の増加とともに、バイアス誤差は振動的に変化し、不安定である。(3) エントロピー推定量 $\hat{H}_K^{(P)}$ では、打ち切り次数 K の増加とともにバイアス誤差が次第に減少し、 $K \geq 4$ でほぼ飽和する。(4) エントロピー推定量 $\hat{H}_L^{(T)}$ では、最適な打ち切り次数 $L = 4$ を選択すれば、バイアス誤差の大きさは、エントロピー推定量 $\hat{H}_K^{(P)}$, $K \geq 4$ に対するものと等しいか、わずかに小さい程度である。(5) エントロピー推定量 $\hat{H}_K^{(P)}$ は、 $M = 2, 3$ に対して推定された係数 $b_{n,k,K}$ を使っているにもかかわらず、 $M = 7$ に対しても有効である。

提案するエントロピー推定量 $\hat{H}_K^{(P)}$ が実用的な M の範囲の問題にも適用可能であることを示すため、情報源記号の数 M が大きい場合についても数値実験を行った。大きな M を定め、 $\sum_{j=1}^M r_j = 1$ のもとで $0 \leq r_j \leq 1$ の範囲で一様分布に従うように真の生起確率 r_1, \dots, r_M を決め、この生起確率 r_1, \dots, r_M のもとで実際に N 個の標本を生成し、その中に含まれている各情報源記号の数から、提案法によりエントロピー推定を行い、バイアス誤差を算出した。これを r_1, \dots, r_M を変えて試行を繰り返し、バイアス誤差の平均値を算出した。一例として、 $N = 500$, $M = 100$ とした際の結果を図5に示す。この図の縦軸はバイアス誤差の平均値であり、黒色の棒グラフはエントロピー推定量 $\hat{H}_L^{(T)}$ 、白抜き棒グラフはエントロピー推定量 $\hat{H}_K^{(P)}$ に対する結果を表している。横軸は、 $\hat{H}_L^{(T)}$, $\hat{H}_K^{(P)}$ におけるそれぞれ L と K を表し、標本数 N の絶対値最大次数が等しいものをグループ化してある。

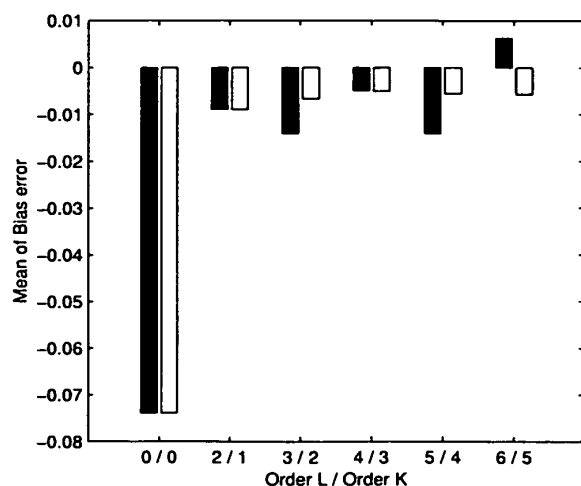


図5 $N = 500$, $M = 100$ の場合のエントロピー推定量のバイアス誤差の平均。テイラー展開法：黒色の棒グラフ， N のべき級数法：白抜きの棒グラフ

Fig. 5 Mean of bias error in entropy estimate when $N = 500$, $M = 100$. Bias errors in entropy estimates by Taylor expansion method and power series method are shown by black and white bar graph, respectively.

これより，エントロピー推定量 $\hat{H}_L^{(T)}$ では， $L = 4$ でバイアス誤差が最小になり， $L > 4$ では逆にバイアス誤差が増える傾向にあり，推定量としては L の変化に非常に不安定であるといえる．一方，提案するエントロピー推定量 $\hat{H}_K^{(P)}$ では， $K = 3$ でバイアス誤差が最小になるものの， $K > 3$ としてもそれほどバイアス誤差は悪化せず，安定した推定精度が得られている．

3.2 従来のバイアス誤差改善法との比較

一般的なエントロピー推定量がバイアス誤差をもち，それが標本数 N に依存することは，既に知られていた．そこで，Strong ら [5] は，様々な標本数 N でエントロピーを \hat{H} として推定し，標本数 N と推定されたエントロピー \hat{H} の間の関係を $1/N$ の多項式で最小 2 乗フィッティング，すなわち多項式の係数を最小 2 乗推定し，エントロピーを $1/N \rightarrow 0$ へ外挿することにより推定する手法を考案した．本論文では，標本数 N と推定されたエントロピー \hat{H} の期待値 $E[\hat{H}]$ の関係を $1/N$ の多項式である式 (9)，あるいは式 (11) で与えている．すなわち，本論文では，Strong らが用いたモデル，つまり $1/N$ の多項式が妥当であることを理論的に裏づけ，これまで対象ごとに推定され，未知であった係数を各情報源記号の生起確率 r_1, \dots, r_M の関数として与えたことになる．

Strong らの手法では，標本数 N と推定されたエン

トロピー \hat{H} の関係をフィッティングしているが，これは，本来，標本数 N とエントロピーの推定値の期待値 $E[\hat{H}]$ の関係として表現すべきものである．しかしながら，一つの標本系列しか存在しない場合には，個々のエントロピー推定値を利用するしかなく，そのため，Strong らの手法では，精度の悪い当てはめを行うことになり，結果として最終的に得られたエントロピー推定値のばらつき，つまり平均 2 乗誤差が大きくなることになる．Strong らは， $1/N$ の 2 次多項式を使うことを推奨しているが，その際の推定値のバイアス誤差と平均 2 乗誤差を図 3 に記号を付していない点線で重ねて示した．バイアス誤差はよく補償されているが，平均 2 乗誤差が約 3 倍に悪化している．また，図 3 中には示していないが，Strong らの手法で $1/N$ の 1 次式を使った場合には，平均 2 乗誤差はほとんど悪化せず，一方，バイアス誤差は， $1/N$ までを使った提案法 $\hat{H}_1^{(P)} (= \hat{H}_2^{(T)})$ にはほぼ一致する結果が得られている．バイアス誤差を抑えても平均 2 乗誤差を 3 倍に悪化させては実用的とはいえないので，結局，一つの標本系列しか存在しない場合には，Strong らの 1 次式を用いた手法，つまり $\hat{H}_1^{(P)} (= \hat{H}_2^{(T)})$ が，これまで用いられてきた最も精度の高いエントロピー推定法であるといえる．

最適な次数を選択した際の提案法 $\hat{H}_4^{(P)}$ が，従来法 ($\cong \hat{H}_1^{(P)} = \hat{H}_2^{(T)}$) と比較し，バイアス誤差を有意に減少させることを確かめるため，これらのバイアス誤差の差の T 検定を行った．その結果，図 3，図 4 で示した範囲 $2 \leq M \leq 7$, $100 \leq N \leq 4000$ ，図 5 で示した $M = 100$, $N = 500$ において，いずれも有意水準 0.01 で有意差が検出された．このことより，提案法は，従来法に比べ統計的に有意にバイアス誤差を減少させられるといえる．

4. む す び

一般的に用いられている独立生起情報源のエントロピー推定量は不偏性をもたない．本論文では，こうした一般的なエントロピー推定量を情報源記号の生起確率の真値の周りでテイラー展開することにより，エントロピー推定量のバイアス誤差を定式化した．その結果，バイアス誤差は，標本数のべき級数として表現でき，各次数の係数は，情報源記号の生起確率の様々なべき乗平均の線形結合として表現できることを示した．更に，定式化したバイアス誤差の符号を反転させ，一般的なエントロピーの推定量に付加した新たなエント

論文／バイアス誤差を改善するエントロピーの推定量

ロピー推定量を提案した。べき級数の打ち切り回数により、線形結合の係数の最適値が異なるため、それらを数値実験により決定した。具体的に数値実験を行うことにより、提案したエントロピーの推定量では、従来法に比べ、平均 2 乗誤差をほとんど悪化させることなく、バイアス誤差を統計的にも有意に減少させられることを示した。

謝辞 本研究の一部は、文部科学省・平成 11-14 年度科学技術振興調整費「視覚系のニューロインフォーマティクスに関する研究」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, 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, 2000.
- [3] Y. Yokota, "An entropy estimator with least square error," Proc. 24th Annual International Conference of the IEEE EMBS, Houston, pp.188-189, Oct. 2002.
- [4] Y. Yokota, "An entropy estimator improving bias error," Proc. 2nd European Medical & Biological Engineering Conference, Vienna, Dec. 2002.
- [5] S.P. Strong, R. Koberle, R.R. de R. van Steveninck, and W. Bialek, "Entropy and information in neural spike trains," Phys. Rev. Lett., vol.80, no.1, pp.197-200, 1998.

(平成 13 年 11 月 26 日受付, 14 年 8 月 19 日再受付)



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 同大学院博士後期課程了。工博。同年名古屋工大・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。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。