推定を独立な標本から繰り返す場合に最適なエントロピー推定量

学生員 志賀 元紀* 非会員 横田 康成**

An Optimal Entropy Estimator which is an Average over Plural Independent Estimates

Motoki Shiga*, Student Member, Yasunari Yokota**, Non-member

We cannot record biological signals, e.g., neural signals, over long time, because of biological tiredness and adaptation property. We often adopt the average over plural estimates obtained from samples recorded in different trials as a final estimate to improve the estimation accuracy. For minimizing mean squared error of the entropy estimation, we should optimally balance between bias error and mean squared error of individual estimations, however the optimal balance is unknown. This paper derives the optimal balance between bias error and mean squared error of individual estimations.

キーワード：エントロピー，エントロピー推定量，平均 2 乗誤差

Keywords: entropy, entropy estimator, mean squared error

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⁾。こうしたことから，エントロピーをより精度良く推定する手法が渴望されるようになった。

生体から発生する信号は，生体の疲れや順応が原因になり，長時間の記録をできないことが多い。こうした状況で，エントロピーの推定精度を高めるためには，エントロピー推定を独立な標本から繰り返し，得られた複数の推定値の集合平均を最終的な推定結果にする手法が用いられる⁽¹⁾。

こうした推定法の平均 2 乗誤差を最小化するには，個々の推定のバイアス誤差と平均 2 乗誤差がトレードオフ関係にある⁽²⁾ため，個々の推定のバイアス誤差と平均 2 乗誤差の配分比を最適化しなくてはならない。そこで，志賀らは，バイアス誤差と平均 2 乗誤差の配分比を制御できるエントロピー推定量を提案した⁽²⁾。しかし，最適な配分比が明らかになっていないため，志賀らの手法において，配分比を数値的に最適化しなくてはならなかった。

本論文では，エントロピー推定を独立な標本から繰り返し，その集合平均を最終的な推定値にするエントロピー推

定量の平均 2 乗誤差を導出する。これにより，個々の推定におけるバイアス誤差と平均 2 乗誤差の最適な配分比を明らかにする。

2. 推定を独立に繰り返す場合のエントロピー推定

M 種類の情報源記号を，生起確率 $r = (r_1, r_2, \dots, r_M)$ で出力する独立生起情報源を仮定する。この情報源から出力された N 個の標本を用いてエントロピーを推定する試行を K 回繰り返し，得られた K 個の推定値の集合平均を，最終的な推定値にする場合を考える。

こうしたエントロピー推定量は， $i = 1, \dots, K$ 番目の標本群に含まれる情報源記号の数を $\mathbf{n}^{(i)} = (n_1^{(i)}, n_2^{(i)}, \dots, n_M^{(i)})$ ，その標本群に対するエントロピー推定量を $\hat{H}(\mathbf{n}^{(i)})$ としたとき，

$$\hat{H}_{\text{Avg}}(\mathbf{n}) = \frac{1}{K} \sum_{i=1}^K \hat{H}(\mathbf{n}^{(i)}) \dots \dots \dots (1)$$

となる。式(1)中の $\hat{H}(\mathbf{n}^{(i)})$ には，生起確率の最尤推定値を用いた推定量 $\hat{H}_{\text{conv}}(\mathbf{n}) = -\sum_{i=1}^M (n_i/N) \log(n_i/N)$ などが用いられる。

推定量 $\hat{H}_{\text{Avg}}(\mathbf{n})$ の平均 2 乗誤差を最小にするには，個々の推定 $\hat{H}(\mathbf{n})$ のバイアス誤差と平均 2 乗誤差がトレードオフ関係にある⁽²⁾ため，個々の推定 $\hat{H}(\mathbf{n})$ のバイアス誤差と平均 2 乗誤差の配分比を最適化しなくてはならない。そこで，志賀らは，その配分比を制御できるエントロピー推定量を提案した⁽²⁾。しかし，推定量 $\hat{H}_{\text{Avg}}(\mathbf{n})$ の平均 2 乗誤差を最小化するための，個々の推定 $\hat{H}(\mathbf{n})$ におけるバイアス誤差と平均 2 乗誤差の最適な配分比は明らかにされていない。

3. エントロピー推定量 $\hat{H}_{\text{Avg}}(\mathbf{n})$ の平均 2 乗誤差

情報源記号の真の生起確率 r に関する事前知識がない場

* 岐阜大学工学研究科電子情報システム工学専攻
〒501-1193 岐阜県岐阜市柳戸 1-1
Graduate School of Engineering, Gifu University
Yanagido 1-1, Gifu-shi, Gifu, 501-1193

** 岐阜大学工学部応用情報学科
〒501-1193 岐阜県岐阜市柳戸 1-1
Faculty of Engineering, Gifu University
Yanagido 1-1, Gifu-shi, Gifu, 501-1193

合, その事前確率 $p(r)$ は, r によらず一定, つまり,

$$p(r) = \begin{cases} \Gamma(M), & r \in R, \\ 0, & \text{others} \end{cases} \dots\dots\dots (2)$$

と仮定される。ただし, $\Gamma(\cdot)$ は, ガンマ関数を表し, また, R は, $0 \leq r_i \leq 1, i = 1, 2, \dots, M$ かつ $\sum_{i=1}^M r_i = 1$ を満たすベクトル r の集合を表す。

また, N 個の標本に含まれる情報源記号の出現数 n の確率分布 $p(n|r)$ は, 独立生起情報源の仮定より, 多項分布

$$p(n|r) = P_{N,r}(n) \equiv \frac{N!}{\prod_{i=1}^M n_i!} \prod_{i=1}^M r_i^{n_i} \dots\dots\dots (3)$$

と仮定される。このとき, エントロピー推定量 $\hat{H}_{\text{Avg}}(n)$ の平均 2 乗誤差は,

$$\begin{aligned} & J(K, N, M) \\ &= E_{n,r} [\{\hat{H}_{\text{Avg}}(n) - \text{ent}(r)\}^2] \\ &= \int_R \prod_{i=1}^M \sum_{n^{(i)} \in S} p(n^{(i)}|r) \{\hat{H}_{\text{Avg}}(n) - \text{ent}(r)\}^2 p(r) dr \\ &= \int_R \prod_{i=1}^M \sum_{n^{(i)} \in S} P_{N,r}(n^{(i)}) \{\hat{H}_{\text{Avg}}(n) - \text{ent}(r)\}^2 \Gamma(M) dr \\ &\dots\dots\dots (4) \end{aligned}$$

となる。ただし, S は, $n_i \in \{0, 1, \dots, N\}, i = 1, 2, \dots, M$ かつ $\sum_{i=1}^M n_i = N$ を満たすベクトル n の集合を表す。式 (4) に式 (1) を代入して展開すれば,

$$J(K, N, M) = \frac{(K-1)}{K} J_b(N, M) + \frac{1}{K} J_s(N, M) \dots\dots\dots (5)$$

を得る。ただし, 式中の $J_b(N, M)$ および $J_s(N, M)$ は,

$$J_b(N, M) = \Gamma(M) \int \left(\sum_{n \in S} P_{N,r}(n) \hat{H}(n) - \text{ent}(r) \right)^2 dr, \dots\dots\dots (6)$$

$$J_s(N, M) = \Gamma(M) \int \sum_{n \in S} P_{N,r}(n) \left(\hat{H}(n) - \text{ent}(r) \right)^2 dr \dots\dots\dots (7)$$

である。 $J_b(N, M)$ は, 個々のエントロピー推定量 $\hat{H}(n)$ のバイアス誤差の 2 乗平均を表し, また, $J_s(N, M)$ は, 個々のエントロピー推定量 $\hat{H}(n)$ の平均 2 乗誤差を表している。

したがって, 平均 2 乗誤差 $J(K, N, M)$ を最小にするための, 推定量 $\hat{H}(n^{(i)})$ のバイアス誤差と平均 2 乗誤差の最適な配分比は, $1 : K - 1$ であることが式 (5) よりわかる。

4. 志賀らの手法⁽²⁾において, 最適な配分比を設定した場合の推定精度

志賀らの手法⁽²⁾は, エントロピー推定量 $H(n^{(i)})$ が, 任意の 1 変数関数 $f(n)$ の和

$$H(n^{(i)}) = \sum_{i=1}^M f(n_i) \dots\dots\dots (8)$$

図 1 エントロピー推定量の平均 2 乗誤差

Fig. 1. The mean squared errors of entropy estimators.

になることを仮定した下で, バイアス誤差の 2 乗平均と平均 2 乗誤差の重みつき和, つまり,

$$J_s(N, M) + \lambda J_b(N, M) \dots\dots\dots (9)$$

を最小にするように導出された。具体的には, 連立方程式

$$\begin{aligned} & 2 \sum_{n=0}^N \left\{ \lambda A_{N,M}(n, m) + D_{N,M}(n) \delta_{n,m} \right. \\ & \quad \left. + E_{N,M}(n, m) \right\} f(n) = -\lambda B_{N,M}(m) - F_{N,M}(m), \\ & \quad m \in \{0, 1, \dots, N\} \dots\dots\dots (10) \end{aligned}$$

の解 $f^*(n)$ を式 (8) に代入することで得られる。式 (10) 中の係数 $A, B \sim G$ については, 文献 (2) を参照していただきたい。前節で述べた最適な配分比より, $\lambda = K - 1$ としたとき, 志賀らの手法は, 平均 2 乗誤差 $J(K, N, M)$ を最小にする。

具体的に, 平均 2 乗誤差 $J(K, N, M)$ と試行回数 K の関係を示す。図 1 (a), (b) に, 標本数 $N = 100$, 記号の種類数 $M = 10, 25$ の場合における, 志賀らの手法, 推定量 $\hat{H}_{\text{conv}}(n)$, 最小 2 乗法⁽³⁾の平均 2 乗誤差を, 記号 $+, \circ, \triangle$ でそれぞれ示した。図 1 より, 最適な配分比を設定することにより, 志賀らの手法は, 他の手法より平均 2 乗誤差を小さくできることがわかる。

5. まとめ

本論文では, エントロピーの推定を独立に繰り返し, その集合平均を最終的な推定値にする手法の平均 2 乗誤差を導出した。これにより, 個々の推定におけるバイアス誤差と平均 2 乗誤差の最適な配分比が明らかになった。

(平成 17 年 3 月 15 日受付)

文 献

- (1) P. Reinagel and R.C. Reid: "Temporal coding of visual information in the thalamus", *J. Neuroscience*, Vol.20, No.14, pp.5392-5400 (2000)
- (2) M. Shiga and Y. Yokota: "An Entropy Estimator That Minimizes Mean Squared Error under Condition of Restricted Averaged Squared Bias Error", *IEICE Japan*, Vol.J88-A, No.4, pp.519-527 (2005-4) (in Japanese)
志賀元紀・横田康成:「バイアス誤差の 2 乗平均を任意の値に制約する条件下で平均 2 乗誤差を最小化するエントロピー推定量」, 信学論, **J88-A**, 4, pp.519-527 (2005-4)
- (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 (1995)