

2元行列アンサンブルの見逃し誤り確率の 数値的評価

Numerical Evaluation of Undetected Error Probability of Binary Matrix Ensemble

大杉 晶一

Akihito Osugi

広島市立大学大学院 情報科学研究科

Email: osugi@cm.info.hiroshima-cu.ac.jp

三村 和史

Kazushi Mimura

広島市立大学大学院 情報科学研究科

Email: mimura@hiroshima-cu.ac.jp

Abstract—We numerically evaluate the error exponent of the undetected error probability of a binary matrix ensemble for finite length codewords. We also show that the range where the error detection is valid do not depend on the code length.

I. はじめに

近年の情報化社会の発展において、信頼性の高いデータ通信が求められており、誤り検出方式の重要性は増加している。誤り検出は、データ通信の信頼性を保つために必要な技術のひとつである。誤り検出や誤り訂正を行うとき、誤りが生じたにも関わらず、その誤りが見逃されてしまうという事象が起こり得る。この見逃し誤り事象が起こる確率のことを、見逃し誤り確率という。誤り検出の性質の理解のためには、見逃し誤り確率の評価が重要である。見逃し誤り確率は、重み分布 [1], [2], [3], [4] から評価されることが示され [5], 理論的な解析が進められている。

2元行列アンサンブルについて、見逃し誤り確率は、符号長が無限大の漸近評価が解析されている [5], [6]。符号長が有限場合の解析的な評価は困難で、その場合は数値的な方法が必要となる。本論文では、2元行列アンサンブルのうち、レギュラー疎行列アンサンブルについて、符号語長が有限の場合の見逃し誤り確率の誤り指数の数値的評価を行う。

II. 見逃し誤り確率

2元線形符号 $C(H) \triangleq \{x \in \mathbb{F}_2^m : Hx^t = 0^m\}$ の符号語 x を送信するとする。行列 H は各要素が0または1の $m \times n$ 2元検査行列、 \mathbb{F}_2 は集合 $\{0, 1\}$ の2元有限体、 0^m および太文字は、長さ m のゼロベクトルと2元行ベクトルをそれぞれ表す。また、通信路はビット反転確率 ϵ ($0 < \epsilon < 1/2$) の2元対称通信路を仮定する。検出器は、受信語 $y = x \oplus e$ からシンドローム $s = Hy^t = H(x \oplus e)^t$ を計算して誤り検出を行う。 $e = (e_1, \dots, e_n) \in \mathbb{F}_2^n$ は2元誤りベクトルである。

見逃し誤り事象とは、 $He^t = 0$ かつ $e \neq 0$ となる事象である。これより、見逃し誤り確率 $P_U(H)$ は

$$P_U(H) \triangleq \mathbb{P}[e \in C(H), e \neq 0] \quad (1)$$

と表すことができる。これは、 $w(x)$ はベクトル x のハミング重み(非ゼロ要素数)とすると、 $P_U(H) = \sum_{e \in C(H), e \neq 0^n} \epsilon^{w(e)} (1 - \epsilon)^{n-w(e)}$ となる。

III. 誤り指数

$\mathcal{E}_{m,n}$ を、各行に K 個、各列に C 個の非ゼロ要素を持つ $m \times n$ 行列の集合とする。すべての要素 $H \in \mathcal{G}_{m,n}$ に等確率 $\tilde{P}(H) = 1/N$ を割り当てる。ただし、 N は $\mathcal{E}_{m,n}$

の濃度である。この組 $(\mathcal{E}_{m,n}, \tilde{P})$ を (C, K) -レギュラーアンサンブルと呼ぶ。レギュラーアンサンブルで符号化率は、 $R \triangleq 1 - m/n = 1 - C/K < 1$ となる。

符号化率 R を一定とした $(1-R)n \times n$ 行列のアンサンブル $\mathcal{G}_{(1-R)n,n}$ に対して、平均見逃し誤り確率は n について減少する。このため、見逃し誤り確率の誤り指数が、漸近的な性質の理解で本質的な役割を果たす。誤り指数として、有限サイズのアニール平均誤り指数 $T_G(\epsilon; R, n)$ 、有限サイズのカエンチ平均誤り指数 $\hat{T}_G(\epsilon; R, n)$ の2つを評価する。これらの異なる指標は、アニール平均誤り指数は見逃し誤り確率の外れ値の影響を受けやすいが、カエンチ平均誤り指数は受けにくいという特徴を持つ。

$$T_G(\epsilon; R, n) \triangleq \frac{1}{n} \log_2 \langle P_U(H) \rangle \quad (2)$$

$$\hat{T}_G(\epsilon; R, n) \triangleq \frac{1}{n} (\log_2 P_U(H)) \quad (3)$$

ここで、 $\langle f(H) \rangle \triangleq |\mathcal{H}_{(1-R)n,n}|^{-1} \sum_{H \in \mathcal{H}_{(1-R)n,n}} f(H)$ とおいた。また、 $\mathcal{H}_{(1-R)n,n} \subset \mathcal{E}_{(1-R)n,n}$ は検査行列 H の集合とし、 $|\mathcal{H}_{(1-R)n,n}|$ は集合 $\mathcal{H}_{(1-R)n,n}$ の濃度である。Jensenの不等式より、 $\hat{T}_G(R, n) \leq T_G(R, n)$ の関係がある。

IV. 結果

(3,6)-レギュラーアンサンブルについて、有限サイズのアニール平均誤り指数 $T_G(\epsilon; R, n)$ と有限サイズのカエンチ平均誤り指数 $\hat{T}_G(\epsilon; R, n)$ を数値的に評価する。図1に有限サイズのアニール平均誤り指数を、図2に有限サイズのカエンチ平均誤り指数を示す。それぞれ、横軸はビット反転確率 ϵ である。また、符号化率は $R = 0.5$ 、符号長 $n \in \{14, 16, 18, 20, 22, 24, 26\}$ とした。 $N = 500$ 個の検査行列 H をビット反転確率 $\epsilon \in \{0.01, 0.02, \dots, 0.49\}$ ごとに乱数により生成する。実線は、符号長が無限大の場合の漸近的な誤り指数の解析値である [5], [6]。

ここで、誤りのない誤りベクトルが $e = 0^m$ となる事象を考える。この誤りなし事象の確率を P_N とおくと、 $P_N = (1 - \epsilon)^n \leq 1$ となる。誤りなし確率の誤り指数を $\mathcal{S}(\epsilon)$ とおくと、

$$\mathcal{S}(\epsilon) \triangleq \frac{1}{n} \log_2 P_N = \log_2(1 - \epsilon) \leq 0 \quad (4)$$

となる。ここでアニール平均誤り指数と誤りなし確率の指数の交点(最大値)を ϵ_a^* とし、カエンチ平均誤り指数と誤りなし確率の指数の交点(最大値)を ϵ_q^* とする。それぞれ

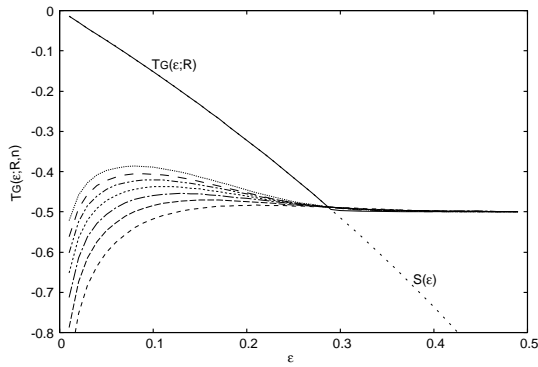


図 1. 有限サイズのアニール平均誤り指数 $T_G(\epsilon; R, n)$. 曲線は上から, $n \in \{26, 24, 22, 20, 18, 16, 14\}$ での値を表す.

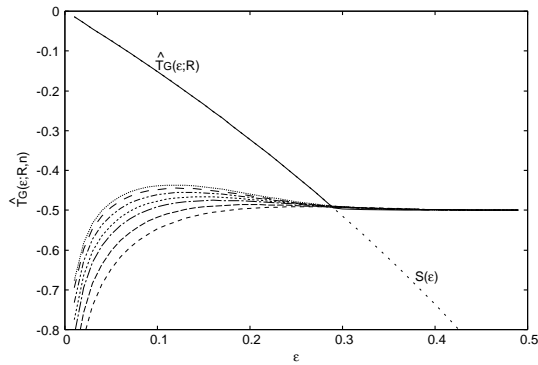


図 2. 有限サイズのクエンチ平均誤り指数 $\hat{T}_G(\epsilon; R, n)$. 曲線は上から, $n \in \{26, 24, 22, 20, 18, 16, 14\}$ での値を表す.

をアニール臨界ビット反転確率, クエンチ臨界ビット反転確率と呼ぶ. ビット反転確率 ϵ が増加するにつれて, 誤り指数は $-(1-R)$ に収束するが, 誤りなし確率は減少する. つまり, $\epsilon > \epsilon^*$ では, 誤りなし確率は見逃し誤り確率よりも指数的に小さいため, 符号長が大きいときは誤り検出が有効なのは $\epsilon < \epsilon^*$ に限られる. このため, ϵ^* は誤り検出が有効となる範囲の指標となる.

V. 考察

図 1, 図 2 の誤り指数 $T_G(\epsilon; R, n)$, $\hat{T}_G(\epsilon; R, n)$ はビット反転確率 ϵ が小さい範囲では, 符号長 n が大きくなるほど漸近的な理論値 $T_G(\epsilon; R)$, $\hat{T}_G(\epsilon; R)$ [5], [6] にそれぞれ近づいている.

ビット反転確率 ϵ が大きい範囲では, 誤り指数は $-(1-R) = -0.5$ となっている. しかし, ビット反転確率 ϵ が小さい範囲では, 数値的に求めた誤り指数は小さく, 分散の大きい. ビット反転確率 $\epsilon = 0$ では見逃し誤り確率は $P_U(H) = -\infty$ となる. 符号語長が大きいとき, 任意の $\epsilon > 0$ ですぐに有限の値をとり, $\epsilon = 0$ で急激に減少するが, 符号語長が短いときは ϵ が小さくなるにつれて緩やかに減少しているからであると考えられる.

図 3 アニール臨界ビット反転確率 ϵ_a^* , クエンチ臨界ビット反転確率 ϵ_q^* を図 3 に示す. 臨界ビット反転確率は, 符号長 n に殆ど依存しないことがわかった.

レギュラーアンサンブルの重み分布は集中することが証明されている [7]. 図 4 に, 符号語長 n に対する σ^2/μ^2 の変化を示す. ここで, σ^2 は $P_U(H)$ の分散, μ は $P_U(H)$ の

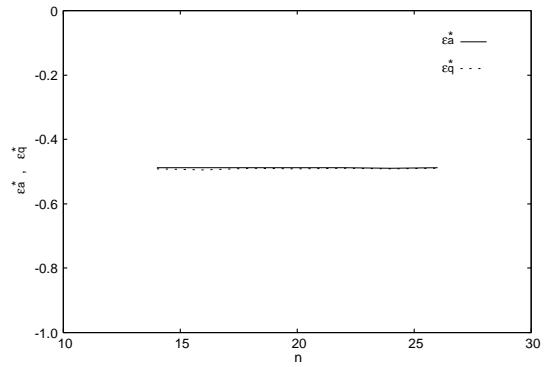


図 3. 臨界ビット反転確率 ϵ_a^* , ϵ_q^* .

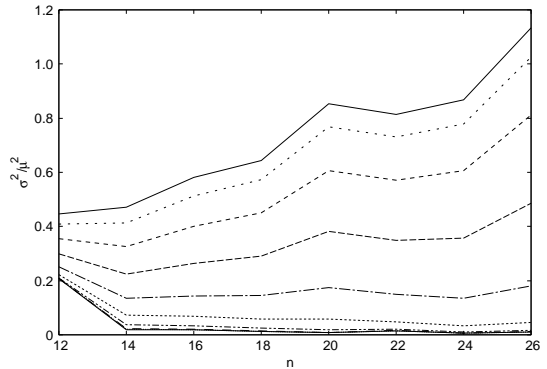


図 4. 規格化された分散 σ^2/μ^2 . 曲線は上から, $\epsilon \in \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.49\}$ での値を表す.

平均である. ϵ が大きい場合は 0 に集中していることが確認できる.

VI. おわりに

本実験では, レギュラーアンサンブルについて, 符号長が有限な時の見逃し誤り確率の誤り指数を数値的に評価した. ビット反転確率の大きい範囲では, 漸近的な評価により得られていた理論の裏付けと定性的に一致した結果を得る事ができた. また, 誤り検出が有効となる範囲の指標は符号語長によらないということがわかった.

参考文献

- [1] S. Litsyn and V. Shevelev, *IEEE Trans. Inf. Theory*, vol. 48, no. 4, 887, Apr. 2002.
- [2] S. Condamine, "Study of the weight enumerator function for a Gallager code, project report," Cavendish Laboratory, University of Cambridge, 2002.
- [3] S. Litsyn and V. Shevelev, "On Ensembles of Low-Density Parity-Check Codes: Asymptotic Distance Distributions," *IEEE Trans. Infor. Theory*, 48, 4, 887 (2003)
- [4] C. Di, A. Montanari and R. Urbanke, *Proc. of ISIT2004*, 102, 2004.
- [5] T. Wadayama, "On the Undetected Error Probability of Binary Matrix Ensembles," *IEEE Trans. Inf. Theory*, 56, 5, 2168 (2010).
- [6] K. Mimura, T. Wadayama, T. Tanaka and Y. Kabashima, "Average Error Exponent of Undetected Error Probability of Binary Matrix Ensembles," to appear in *Proc. of ISIT 2011* (2011)
- [7] V. Rathi, "On the Asymptotic Weight Distribution of Regular LDPC Ensembles," *Proc. of ISIT2005*, 2161 (2005)