縮む列があれば伸びる列があり、圧縮の余地を作るのは列ではなく源の偏りである

— 長さ n ビットの列 2^n 通りに対し、それより短い列は $2^n - 1$ 通りで、いつも 1 通り足りない — 分離子は、相手にしているのが、あり得るすべての列を平等に並べた集まりか、偏った源が出す列か —

[論文751]

山岸 勇輝 (Yuuki YAMAGISHI)
Yamagishi Complexity Torus Lab, Japan｜山岸勇輝ラボ
2026年9月

— From thinking to syncing. —

要旨

圧縮器はたいていのファイルを縮める。けれど、どんな列でも縮める可逆な圧縮器は作れない。縮められない理由と、それでも縮む理由を、数え上げとエントロピーで分ける。新しい定理も法則も主張しない。 第一に、短い列は、いつも 1 通り足りない——32 ビットの列 4294967296 通りに対し、短い列は 4294967295 通り（第2節）。 第二に、だから縮まない列が必ず在る——縮む列があれば、どこかで伸びる列がある（第2節）。 第三に、これが本稿の芯である。すべての列を平等に並べると、大きく縮む列はわずかしか無い——20 ビットの列で 1 バイト縮められるのは、すべての列の 0.007812 未満（第2節）。 第四に、縮める割合は、縮める量に対して指数で落ちる——k ビット縮む割合は 2^{1-k} を超えない（第2節）。 第五に、それでも縮むのは、源が偏っているからである——1 の出る確率 0.90 で 0.468996、0.99 で 0.080793 ビット／記号（第3節）。 第六に、分離子は、相手にしているのが、あり得るすべての列を平等に並べた集まりか、偏った源が出す列かである——「よく縮む」は、どの列の集まりを相手にしているかを言うまで意味が決まらない（第4節）。 キーワード：可逆圧縮、鳩の巣原理、数え上げ、圧縮の限界、エントロピー、源の偏り。

> 本稿の射程（射程注記）：新しい定理も法則も主張しない——短い列の数え上げによる可逆圧縮の限界、ほとんどの列が大きくは縮まないこと、記憶の無い源のエントロピーが縮められる下限であることは、いずれも既知である（カバー＆トーマス 2006、シプサー 2012）。圧縮器を扱わない——個別の圧縮器の性能や手法は扱わない。列の長さの区切りを問わない——数え上げは長さの違う列をすべて使ってよいとした上限で、区切りの情報に要る分は入れていない。源を限る——偏りは記憶の無い二記号の源だけで見る。計算不能性を扱わない——一つの列の最短の記述が計算できないことは論文252 の側で、引くだけにする。実在のデータを主張しない——個別のファイルは扱わない。既刊との関係：論文252 は、「情報量」が四つあり要るものが違うことを示した——その射程注記は、圧縮器が縮められなかったことは短く書けないことの証拠にならないとし、下界は出せないと明記していた。そこでは一つの列について下界が出なかった。ここでは列の集まりについて、数え上げで下界が出る。どの列かは言えないが、大きく縮む列がわずかしか無いことは言える。一つの列の下界と、列の集まりの下界になる。加えたのは、短い列が 1 通り足りないことを三つの長さで数えたこと、k ビット縮む割合の上限を 2^{1-k} と並べ、1 バイトで **0.007812** 未満と示したこと、源の偏りごとのエントロピーを並べたこと、分離子を「すべての列の集まりか、偏った源の列か」に置いたことである。

1. 記法と問い

設定：可逆な圧縮は、列を別の列に写し、元に戻せる写し方である。違う列を同じ列に写すと戻せないので、写し方は一対一でなければならない。長さ n の列と、それより短い列を数える。

$$\#\{\text{長さ } n\} = 2^n, \qquad \#\{\text{長さ } n \text{ 未満}\} = 2^n - 1, \qquad \frac{\#\{\text{長さ } n-k \text{ 以下}\}}{2^n} = \frac{2^{n-k+1}-1}{2^n} < 2^{1-k}$$

記憶の無い二記号の源で 1 の出る確率を p とすると、1 記号あたりのエントロピーは
$h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ である。

問い：どんな列でも縮める可逆な圧縮は在るのか。無いなら、なぜ圧縮は効くのか。

2. 数え上げが、圧縮を禁じる

計算 2.1：三つの長さで、その長さの列と、それより短い列を数えた。

長さ	その長さの列	それより短い列
8 ビット	**256**	**255**
16 ビット	**65536**	**65535**
32 ビット	**4294967296**	**4294967295**

> **観察 2.2**
>
> 短い列は、いつも 1 通り足りない。だから長さ n の列をすべて短い列に一対一で写すことはできず、どんな可逆な圧縮にも縮まない列が在る——縮む列があれば、そのぶん伸びる列が要る。

計算 2.3：長さ 20 ビットの列のうち、k ビット以上縮められる列の割合の上限を出した。

k	割合の上限	2^{1-k}
1	9.999990×10^{-1}	1.000000
2	4.999990×10^{-1}	5.000000×10^{-1}
4	1.249990×10^{-1}	1.250000×10^{-1}
8	7.811546×10^{-3}	7.812500×10^{-3}
16	2.956390×10^{-5}	3.051758×10^{-5}

> **観察 2.4**
>
> これが本稿の芯である。すべての列を平等に並べると、大きく縮む列はわずかしか無い。1 バイト（8 ビット）縮められる列は、すべての列の 0.007812 未満である——どの列が縮むかは圧縮器が決めるが、縮む列の数は圧縮器に依らず数え上げが決める。

> **観察 2.5**
>
> 縮める割合は、縮める量に対して指数で落ちる。割合の上限はどの k でも 2^{1-k} を超えない——1 ビット縮めるのはほとんどの列にできても、16 ビット縮めるのは 2.956390×10^{-5} 以下の列にしかできない。

3. それでも縮むのは、源が偏っているから

計算 3.1：記憶の無い二記号の源で、1 の出る確率ごとにエントロピーを出した。

1 の出る確率	エントロピー（ビット／記号）
0.50	1.000000
0.70	0.881291
0.90	0.468996
0.99	0.080793

> **観察 3.2**
>
> 偏りの無い源は、縮められない。確率 0.50 でエントロピーは 1.000000 で、平等に並べた集まりと同じである——数え上げの禁止は、偏りの無い源の話そのものである。

> **観察 3.3**
>
> 偏った源の列は、大きく縮む。確率 0.90 で 0.468996、0.99 で 0.080793 ビット／記号まで縮められる——偏った源は、平等に並べたときにはわずかしか無い「縮む列」ばかりを出す。圧縮の余地は、列ではなく源が作っている。

4. 分離子

> **観察 4.1**
>
> 分離子は、相手にしているのが、あり得るすべての列を平等に並べた集まりか、偏った源が出す列かである。前者では数え上げが圧縮を禁じ、後者ではエントロピーが余地を与える——「よく縮む」は、どの列の集まりを相手にしているかを言うまで意味が決まらない。

> **観察 4.2**
>
> 論文252 と並べる。252 は、圧縮器が縮められなかったことは短く書けないことの証拠にならず、一つの列については下界が出ないことを明記した——ここでは列の集まりについて、数え上げで下界が出る。どの列かは言えないが、1 バイト縮む列が 0.007812 未満であることは言える。

> **観察 4.3**
>
> 言わないこと。個別の圧縮器も、列の区切りに要る分も、記憶のある源も、計算不能性も扱わない——示したのは、圧縮を禁じる数え上げと、圧縮を許す偏りが、別の相手についての話だということである。

5. 点呼と結び

1. **4294967296** 対 **4294967295**（計算 2.1）：**1** 通り足りない。
2. 縮まない列が在る（観察 2.2）：伸びる列が要る。
3. **8** ビットで $\mathbf{7.811546 \times 10^{-3}}$（計算 2.3）：数え上げの上限。
4. **0.007812** 未満（観察 2.4）：本稿の芯。
5. $\mathbf{2^{1-k}}$ を超えない（観察 2.5）：指数で落ちる。
6. 偏り無しで **1.000000**（計算 3.1）：縮められない。
7. 確率 **0.70** で **0.881291**（計算 3.1）：少し縮む。
8. 確率 **0.90** で **0.468996**（計算 3.1）：半分より下。
9. 確率 **0.99** で **0.080793**（計算 3.1）：源が余地を作る。
10. すべての列の集まりか、偏った源の列か（観察 4.1）：分離子・論文**252**。

結び：縮む列があれば伸びる列があり、圧縮の余地を作るのは列ではなく源の偏りである。長さ n ビットの列 2^n 通りに対し、短い列は $2^n - 1$ 通りで、**32** ビットでも **4294967296** 対 **4294967295** と **1** 通り足りない——だからどんな可逆な圧縮にも縮まない列が在る。長さ **20** ビットの列のうち **8** ビット以上縮められるのは数え上げの上限で $\mathbf{7.811546 \times 10^{-3}}$、すべての列の **0.007812** 未満で、k ビット縮む割合は $\mathbf{2^{1-k}}$ を超えない——縮む列の数は、圧縮器に依らず数え上げが決める。それでも圧縮が効くのは源が偏っているからで、**1** の出る確率 **0.90** で **0.468996**、**0.99** で **0.080793** ビット／記号まで縮められる。分離子は、相手にしているのが、あり得るすべての列を平等に並べた集まりか、偏った源が出す列かである。既刊との位置——論文**252** は一つの列について下界が出ないことを明記した。ここは列の集まりについて、数え上げで下界を出した。正直に言えば——個別の圧縮器も、列の区切りに要る分も、記憶のある源も扱っていない。

参考文献

1. Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley.
2. Sipser, M. (2012). *Introduction to the Theory of Computation* (3rd ed.). Cengage Learning.

3. 山岸勇輝 (2026). 『「情報量」は四つあり、要るものが違う』[論文252]. Zenodo. 10.5281/zenodo.22161148

— From thinking to syncing. —

山岸 勇輝 (Yuuki YAMAGISHI)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)