情報理論09

藤田 一寿

津山工業高等専門学校情報工学科 講師
電気通信大学先進理工学科 協力研究員

ver. 20150612

情報源のエルゴード性

統計力学におけるエルゴード仮説

時間平均と集団（アンサンブル）平均は一致する

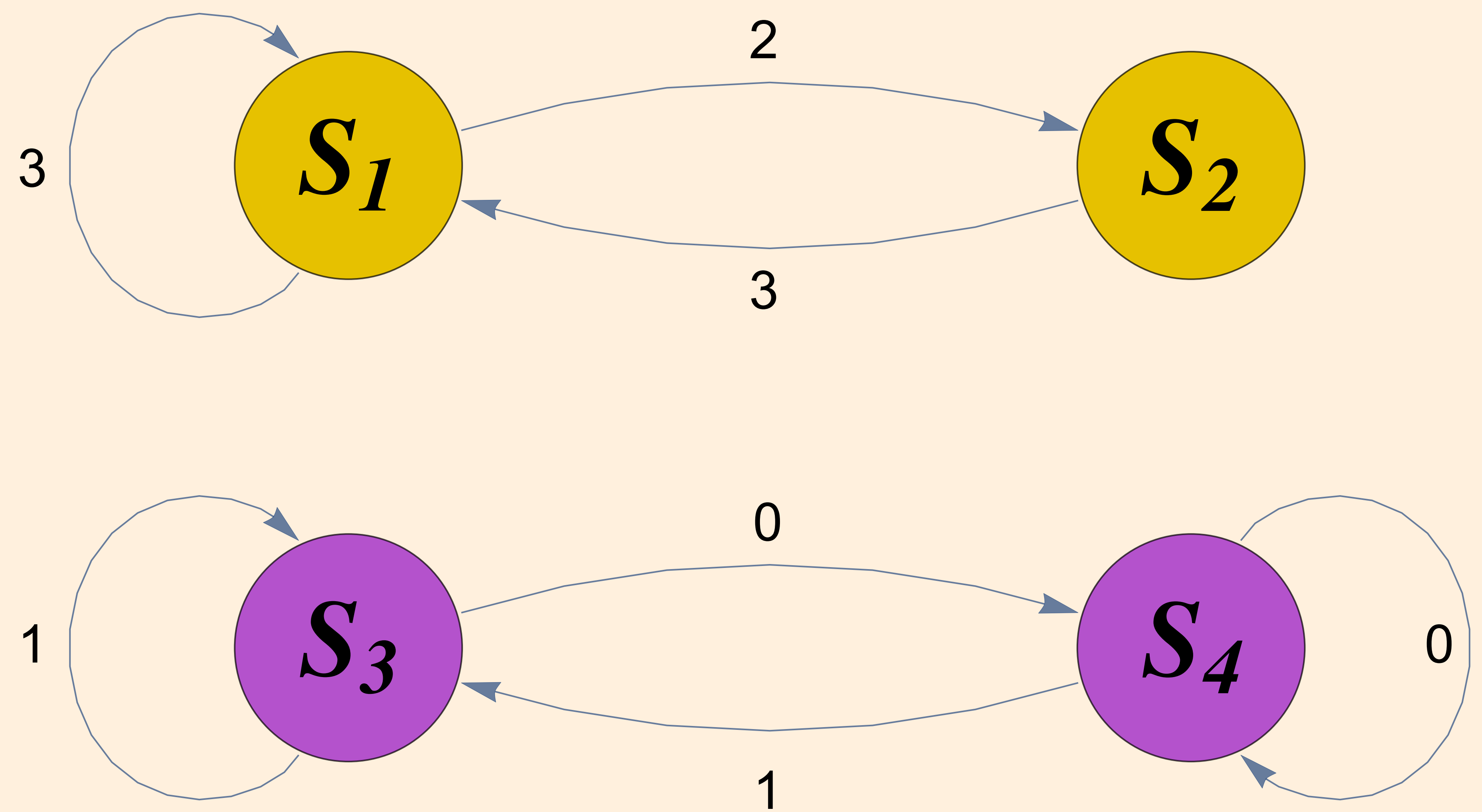
サイコロを１００万回ふった時の平均と、
１００万個のサイコロをふった時の平均は
一致する。

エルゴード的情報源

- 閉部分集合で、小さな閉部分集合を含まない。
  - ある状態になる確率が0とならない。
- この閉部分集合の測度（確率）だけ１であって他の集合の測度（確率）は０である。
- 非周期的である。
- この閉部分集合に含まれる文字列をエルゴード系列と呼ぶ。

非周期的で分離不可能なマルコフ情報源はエルゴード的である。エルゴード的情報源であるならば、時間平均（エントロピー）と集団平均（エントロピー）は等しくなる。

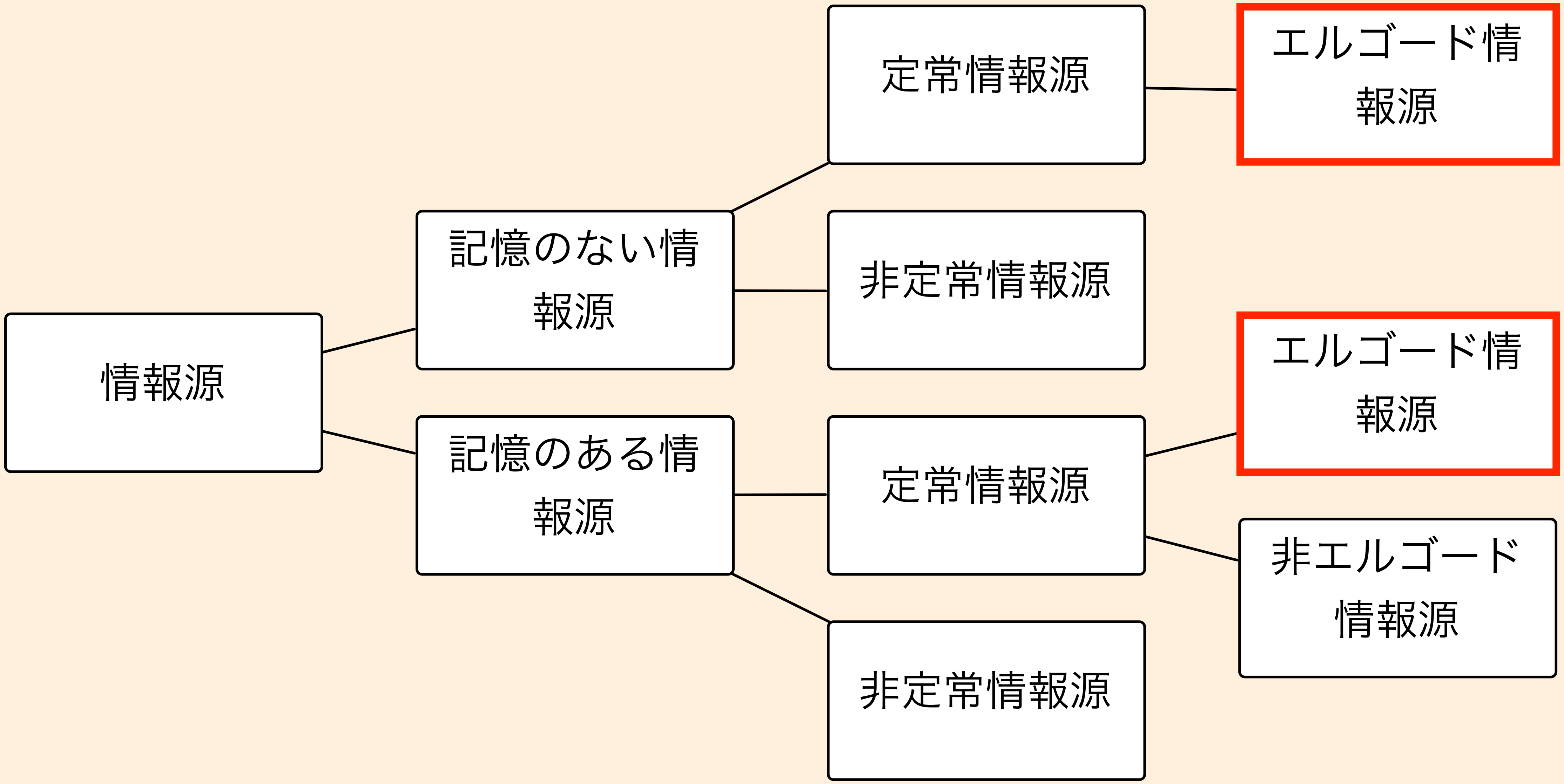
エルゴード的ではない情報源



4つの文字0, 1, 2, 3を発生させるマルコフ情報源

この図でS1かS2の状態にある確率とS3とS4の状態にある確率は等しいとする。

情報源の分類


情報源
記憶のない情報源
記憶のある情報源
定常情報源
非定常情報源
定常情報源
非定常情報源
エルゴード情報源
エルゴード情報源
非エルゴード情報源

情報源の冗長度

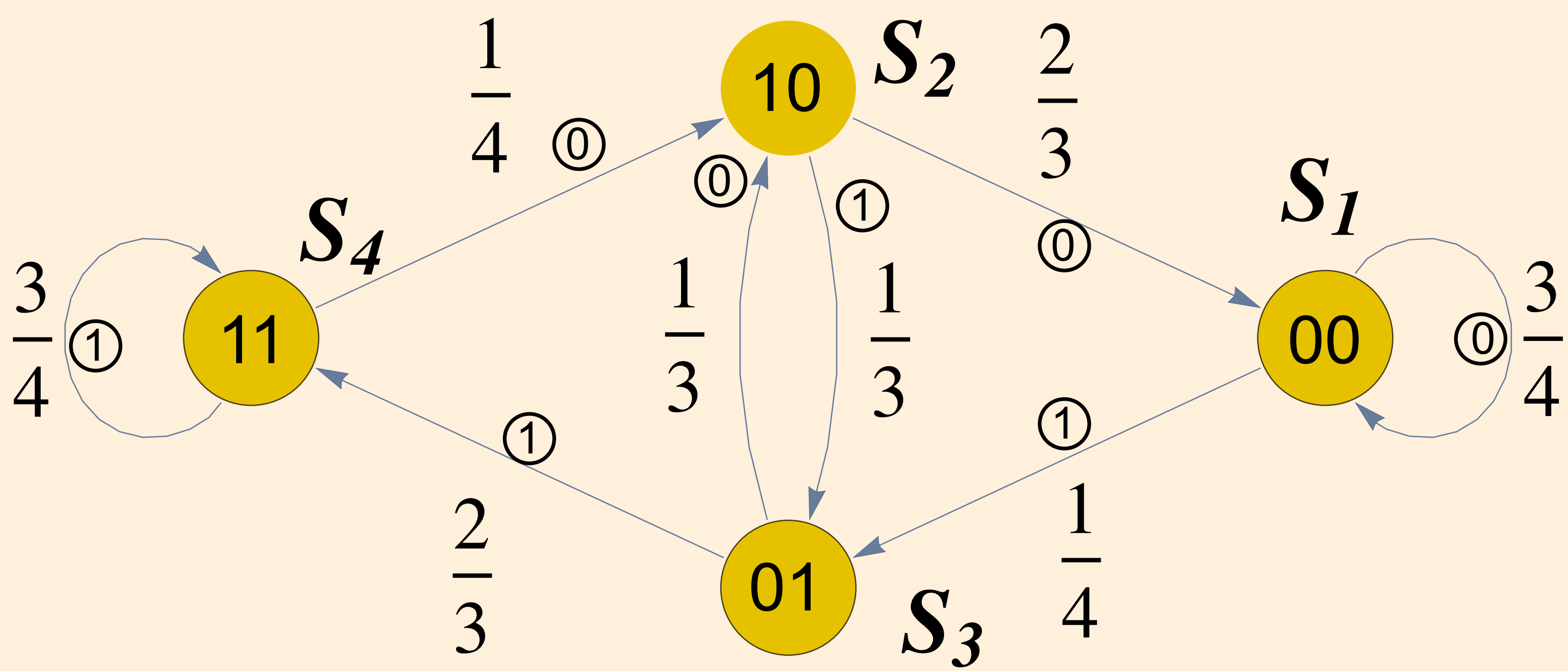
冗長

- 必要以上に物事が多く無駄なこと、長いこと、またはその様子。
  - wikipediaより

情報源のエントロピー

情報源のエントロピーは、次に出てくる文字のエントロピーの期待値であるので、

$$H = \frac{3}{22} \times 0.92 \times 2 + \frac{8}{22} \times 0.81 \times 2$$
$$= 0.84$$

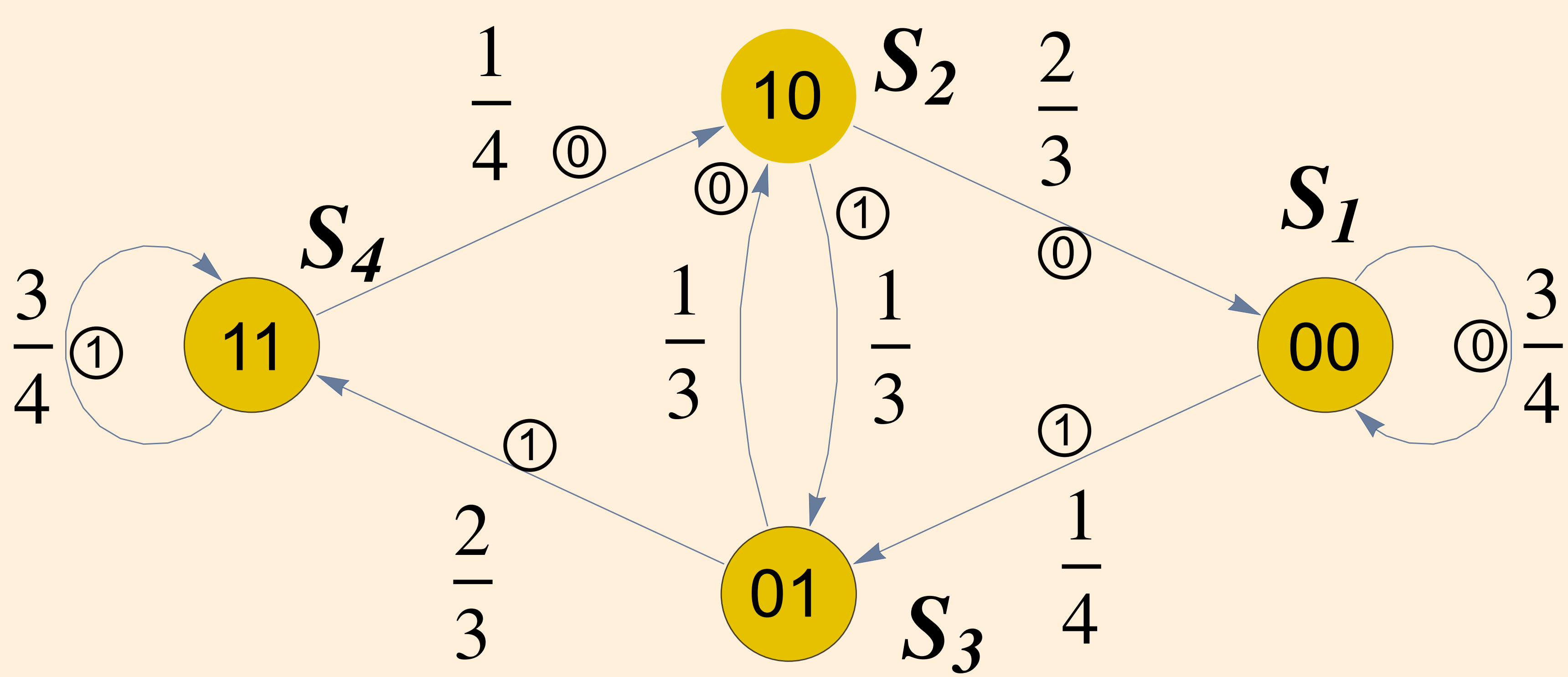


文字の発生確率が独立な場合との比較

この情報源は、実は０と１が出る確率は等しい。

しかし、エントロピーは１ビットではない。

前後の文字の発生確率が独立かつ確率が等しい場合の2元情報源のエントロピーは１ビットになる



この情報源は前後の文字の発生確率は独立ではないため、１ビットよりエントロピーが小さくなる。

０が出た場合、次も０になりやすい。

１が出た場合、次も１になりやすい。

言い方を変えると、００となりやすいので、０が出たあとに０が発生しても得られる情報は少ないと言える。

その分、エントロピーが文字の前後関係が独立な場合より少ない。

情報源の最大エントロピー

A1, …, Akのk個の文字を持っている情報源は、以前に発生した文字が何であろうと関係なく次の文字が出る場合最大のエントロピーを持つ。この時のエントロピーは

$$
\begin{aligned}
H_0 &= -\frac{1}{k}\log\frac{1}{k} \times k \\
&= \log k
\end{aligned}
$$

となる。

冗長度

- 実際のエントロピーをHとするとH0 - H分情報源が持つことができる最大のエントロピーより少ないということになる。逆に考えると、情報源は実際に吐き出す文字のエントロピーに比べ、余分にエントロピーを持つことが可能であると解釈できる（冗長であると考えられる）。

- そこでH0-HとH0の比を冗長度rと定義する。

$$r = \frac{H_0 - H}{H_0}$$

冗長度を考察する

例で用いた情報源の持つことのできる最大のエントロピーは１ビット

実際は0.84ビット

冗長度は0.16

これは、最も効率のよい情報源を用いれば文字数を16%減らすことができる（圧縮できる）ことを示す。

情報源の大数の法則

文字列の発生確率

情報源がk種類の文字を発生させるとすると、文字列の長さがN個であった場合、その文字列は全部で k^N 個ある。

文字列 $x_1 x_2 \ldots x_N$ が発生する確率を

$$p(x_1 x_2 \ldots x_N)$$

と書く。

文字が発生する確率

もし、過去に $...x_1 x_2 ... x_{i-1}$ という文字列が発生したとすると、次に発生する文字xiの発生確率は

$$p(x_i|...x_1 x_2 ... x_{i-1})$$

と書ける。よってxiが発生した時に得られる情報量は

$$I_i = -\log p(x_i|...x_{i-2} x_{i-1})$$

となる。

- ではIiの期待値は

$$\bar{I}_i = -\sum_{x_i} p(x_i|...x_{i-2}x_{i-1}) \log p(x_i|...x_{i-2}x_{i-1})$$

- さらに、これまで出た文字列について平均すると、それが情報源の１文字あたりのエントロピーになる。

$$H = -\sum_{...x_{i-1}} \sum_{x_i} p(...x_{i-1}) p(x_i|...x_{i-1}) \log p(x_i|...x_{i-1})$$

過去のすべての文字列がわかっているとき

- 過去の文字列を x^∞ と表すと、文字列 $x_1 x_2 ... x_N$ が生成される確率は

$$p(x_1 x_2 ... x_N | x^\infty) = p(x_1 | x^\infty) p(x_2 | x^\infty) ... p(x_N | x^\infty)$$

- と書ける。よって、情報量Iは

$$\begin{aligned} I &= -\log(p(x_1 | x^\infty) p(x_2 | x^\infty) ... p(x_N | x^\infty)) \\ &= \sum_{i=1}^{N} I_i \end{aligned}$$

- IiはだいたいIi期待値の期待値であるHと仮定すると

$$I \simeq NH$$

- となることが期待できる。

$$I = -\log p = NH$$

- エントロピーHは情報量の期待値
- つまり、文字列の長さを非常に大きくすれば情報量の平均がエントロピーに近づくと考えられる。

$$\frac{I}{N} \to H$$

$$I \simeq (H \pm \varepsilon)N$$

- 文字列の出現確率をpとすると

$$I = -\log p \simeq (H \pm \varepsilon)N$$

$$(H-\varepsilon)N < -\log p < (H+\varepsilon)N$$

$$|-\frac{\log p}{N} - H| < \varepsilon$$

Nが大きくなると情報量の平均がエントロピーに近づく。逆に言えば、エントロピーをNで割ったものが情報量Iとなる。さらに言うと、出現確率が、

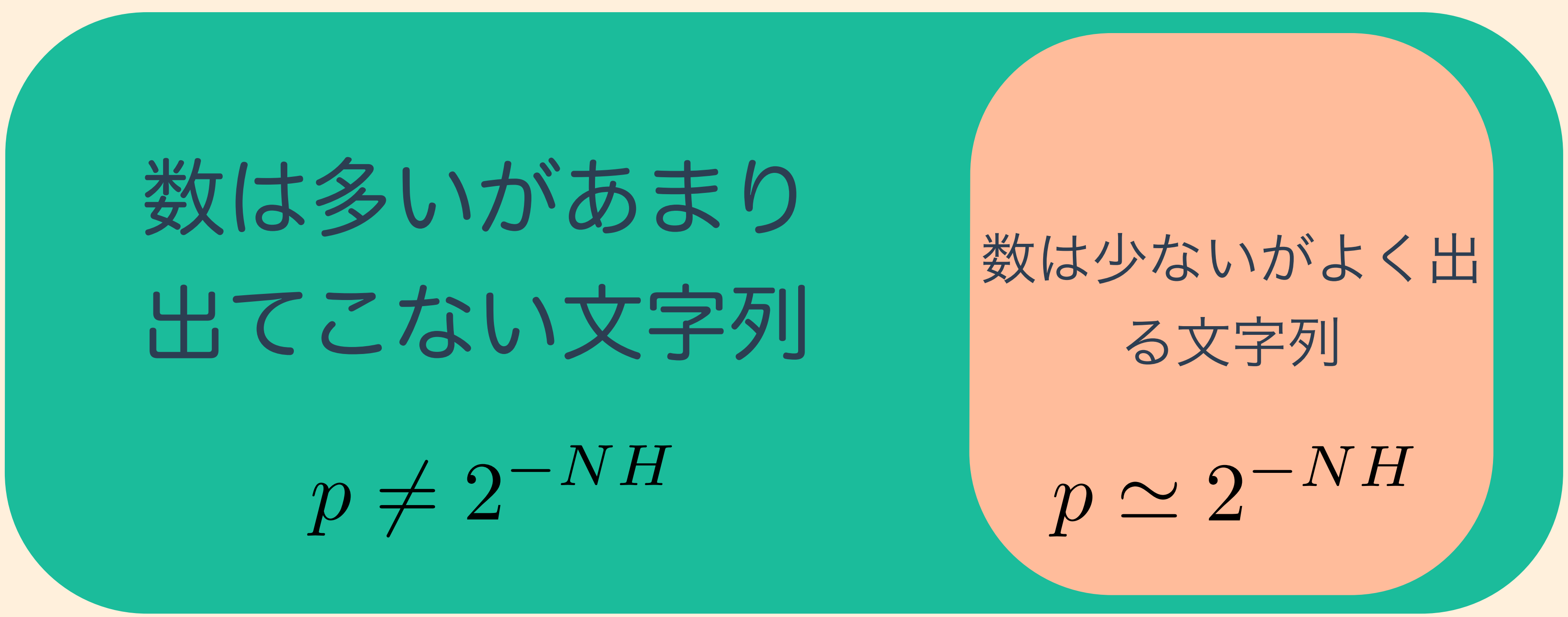
$$-\log p \simeq NH$$

$$p \simeq 2^{-NH}$$

となる文字列がほとんどとなる。

任意のε>0、δ>0に対して、ある N_0 が存在し、$N \geq N_0$ なら、 $|-\frac{\log p}{N} - H| < \varepsilon$ に属する文字列の発生する確率を1-δより大きくすることができる。

$$\mathrm{Prob}\left\{ \left| \frac{-\log p}{N} - H \right| \geq \varepsilon \right\} < \delta$$



文字列全体