数理解析研究所講究録
1032巻 1998年 142-147

i.i.d. 2値系列の頻度分布について

九大システム情報科学研究科　香田　徹　(Tohru Kohda)
九大システム情報科学研究科　藤崎 礼志　(Hiroshi Fujisaki)

Abstract: 筆者らは，先に，独立同分布（independent and identically distributed; i.i.d.）2値系列が等確率である場合について，生成される確率ベクトルに関するパタンとクラスの頻度を評価した．本稿では，等確率でないi.i.d. 2値系列から生成される確率ベクトルに関するパタンとクラスの頻度を評価し，それらの経験分布が形成するガウス分布の分散の理論値を与えた．

1　はじめに

暗号生成器から生成される系列が独立同分布（以下，i.i.d. と略称する）の場合，暗号文を解読することは原理的に不可能であろう．0，1の2値系列がi.i.d. のとき，ある長さ m のパタンの頻度に関する頻度分布はガウス分布に従うという中心極限定理が成立する．したがって，確率論的立場による暗号強度評価法として，暗号文のパタンの頻度の期待値と分散を調べる評価法が考えられる．このテストは m 次および $2m$ 次の相関関数を同時に調べることになる．筆者らは，先に，i.i.d. 2値系列が等確率である場合について，系列から生成される確率ベクトルに関するパタンとクラスの頻度の期待値と分散を理論的に評価した[2]．小文では，i.i.d. 2値系列が等確率でない一般の場合について，確率ベクトルに関するパタンとクラスの頻度の期待値と分散を理論的に評価した*．

2　i.i.d. 2値系列の頻度分布に関する中心極限定理

ω を初期値とするi.i.d. 2値確率変数を $\{X_n(\omega)\}_{n=0}^{\infty}$ としよう．このとき，確率変数

$$\overline{X}_n(\omega) = 1 - X_n(\omega) \tag{1}$$

を導入する．また，これらの期待値を各々，

$$\left.\begin{array}{l} \mathbf{E}[X_n] = p \\ \mathbf{E}[\overline{X}_n] = 1-p \end{array}\right\} \tag{2}$$

と定義する．ここで，$0 \leq p \leq 1$ である．任意の m ビットからなるビット列を

$$\vec{U} = U_0 U_1 \cdots U_{m-1}, \qquad U_n \in \{0,1\}, (0 \leq n \leq m-1) \tag{3}$$

とすると，$\vec{U}$ は 2^m 種類存在する．その r 番目のビットパタンを

$$\vec{u}^{(r)} = u_0^{(r)} u_1^{(r)} \cdots u_{m-1}^{(r)}, \qquad u_n^{(r)} \in \{0,1\}, (0 \leq n \leq m-1) \tag{4}$$

*本小文の内容は，[3] で発表した．

とする．初期値ωの2値確率変数の系列$\{X_n(\omega)\}_{n=0}^{\infty}$に対してあるパタン$\vec{u}^{(r)}$の発生頻度を評価するために，まず2値確率変数

$$Y_n(\omega;u_k^{(r)}) = X_{n+k}(\omega) \oplus u_k^{(r)} = X_{n+k}(\omega)u_k^{(r)} + \overline{X}_{n+k}(\omega)\overline{u}_k^{(r)} \tag{5}$$

を導入する．ただし，$\oplus$は排他的論理和（2を法とする加算）であり，

$$\overline{u}_n^{(r)} = 1 - u_n^{(r)} \tag{6}$$

である．さらに2値確率変数

$$Y_n(\omega;\vec{u}^{(r)}) = \prod_{k=0}^{m-1} Y_n(\omega;u_k^{(r)}) \tag{7}$$

を導入すると，$\{X_n(\omega)\}_{n=0}^{T+m-1}$中に存在する$\vec{u}^{(r)}$の個数は

$$M_T(\omega;\vec{u}^{(r)}) = \sum_{n=0}^{T-1} Y_n(\omega;\vec{u}^{(r)}) \tag{8}$$

で与えられる．

確率変数$Y_n(\omega;\vec{u}^{(r)})$は$m-1$従属であるが，確率論より，中心極限定理の成立が保証される [1]．すなわち，確率変数

$$Z_T(\omega;\vec{u}^{(r)}) = \frac{M_T(\omega;\vec{u}^{(r)}) - T\mathbf{E}[Y_n(\vec{u}^{(r)})]}{\sqrt{T}} \tag{9}$$

を導入すると，$Z_T(\omega;\vec{u}^{(r)})$の分布は，平均値0，分散$\sigma^2$のガウス分布

$$\phi(\xi) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\xi^2}{2\sigma^2}\right] \ (-\infty < \xi < \infty) \tag{10}$$

に近づく．ここで，分散σ^2は，

$$\sigma^2 = \lim_{T\to\infty}\left[\frac{1}{T}\mathbf{E}[M_T^2] - T\{\mathbf{E}[Y_n]\}^2\right] \tag{11}$$

で与えられる．

3 i.i.d. 2値系列から生成される確率ベクトルのパタンに関する分布

以下，$Z_T(\omega;\vec{u}^{(r)})$の分散を評価しよう．まず，確率変数$Y_n(\omega;\vec{u}^{(r)})$の平均値は，

$$\mathbf{E}[Y_n(\vec{u}^{(r)})] = \prod_{k=0}^{m-1}\mathbf{E}[Y_n(u_k^{(r)})] = \prod_{k=0}^{m-1}\{p\,u_k^{(r)} + (1-p)\,\overline{u}_k^{(r)}\} \tag{12}$$

であり，一方，$Z_T(\vec{u}^{(r)})$ の分散は

$$\sigma^2(Z_T(\vec{u}^{(r)})) = \mathbf{E}[Z_T^2(\vec{u}^{(r)})] = \frac{1}{T}\mathbf{E}[M_T^2(\vec{u}^{(r)})] - T\{\mathbf{E}[Y_n(\vec{u}^{(r)})]\}^2 \tag{13}$$

で与えられる．このとき，

$$M_T^2(\omega;\vec{u}^{(r)}) = \sum_{l=0}^{T-1}\sum_{n=0}^{T-1} I_{l,n}^{r,r}(\omega), \tag{14}$$

$$\mathbf{E}[M_T^2(\vec{u}^{(r)})] = \sum_{l=0}^{T-1}\sum_{n=0}^{T-1} \mathbf{E}[I_{l,n}^{r,r}] \tag{15}$$

となる．ただし，

$$I_{l,n}^{r,s}(\omega) = Y_l(\omega;\vec{u}^{(r)})Y_n(\omega;\vec{u}^{(s)}) \tag{16}$$

と定義した．ここで，$\mathbf{E}[I_{l,n}^{r,s}]$ を考えよう．

case 1) $l = n$ の時

$$\begin{aligned}\mathbf{E}[I_{n,n}^{r,s}] &= \mathbf{E}\Big[\prod_{k=0}^{m-1}(X_{n+k}u_k^{(r)}u_k^{(s)} + \overline{X}_{n+k}\overline{u}_k^{(r)}\overline{u}_k^{(s)})\Big] \\ &= \prod_{k=0}^{m-1}\mathbf{E}\Big[X_{n+k}u_k^{(r)}u_k^{(s)} + \overline{X}_{n+k}\overline{u}_k^{(r)}\overline{u}_k^{(s)}\Big] \\ &= \prod_{k=0}^{m-1}\{p\,u_k^{(r)}u_k^{(s)} + (1-p)\,\overline{u}_k^{(r)}\overline{u}_k^{(s)}\}. \end{aligned} \tag{17}$$

case 2) $l = n+i, \quad (1 \le i \le m-1)$ の時

$$\begin{aligned}\mathbf{E}[I_{n+i,n}^{r,s}] &= \mathbf{E}\Big[\prod_{k=0}^{i-1}Y_n(u_k^{(s)})Y_{n+m}(u_{k+m-i}^{(r)})\prod_{k=0}^{m-i-1}(X_{n+i+k}u_k^{(r)}u_{k+i}^{(s)} + \overline{X}_{n+k+i}\overline{u}_k^{(r)}\overline{u}_{k+i}^{(s)})\Big] \\ &= \prod_{k=0}^{i-1}\mathbf{E}[Y_n(u_k^{(s)})]\mathbf{E}[Y_{n+m}(u_{k+m-i}^{(r)})] \\ &\quad\times\prod_{k=0}^{m-i-1}\mathbf{E}[X_{n+i+k}u_k^{(r)}u_{k+i}^{(s)} + \overline{X}_{n+k+i}\overline{u}_k^{(r)}\overline{u}_{k+i}^{(s)}] \\ &= \prod_{k=0}^{i-1}\{p\,u_k^{(s)} + (1-p)\,\overline{u}_k^{(s)}\}\{p\,u_{k+m-i}^{(r)} + (1-p)\,\overline{u}_{k+m-i}^{(r)}\} \\ &\quad\times\prod_{k=0}^{m-i-1}\{p\,u_k^{(r)}u_{k+i}^{(s)} + (1-p)\,\overline{u}_k^{(r)}\overline{u}_{k+i}^{(s)}\}. \end{aligned} \tag{18}$$

case 3) $l = n-i, \quad (1 \le i \le m-1)$ の時

case 2) において l と n との役割を入れ換えると，

$$\begin{aligned}\mathbf{E}[I_{l,l+i}^{r,s}] &= \prod_{k=0}^{i-1}\{p\,u_k^{(r)} + (1-p)\,\overline{u}_k^{(r)}\}\{p\,u_{k+m-i}^{(s)} + (1-p)\,\overline{u}_{k+m-i}^{(s)}\} \\ &\quad\times\prod_{k=0}^{m-i-1}\{p\,u_k^{(s)}u_{k+i}^{(r)} + (1-p)\,\overline{u}_k^{(s)}\overline{u}_{k+i}^{(r)}\}. \end{aligned} \tag{19}$$

case 4) $l = n + i \quad (|i| > m - 1)$ の時

$$\begin{aligned}\mathbf{E}[I^{r,r}_{n+i,n}] &= \mathbf{E}[Y_{n+i}(\vec{u}^{(r)})]\mathbf{E}[Y_n(\vec{u}^{(s)}_k)] \\ &= \prod_{k=0}^{m-1}\{p\,u^{(r)}_k + (1-p)\,\overline{u}^{(r)}_k\}\{p\,u^{(s)}_k + (1-p)\,\overline{u}^{(s)}_k\} \qquad (20)\end{aligned}$$

となる．ここで，case i の場合の数を $\#(\text{case}\,i)$ と表すと

$$\left.\begin{aligned}&\#(\text{case}\,1) = T \\ &\#(\text{case}\,2) = \#(\text{case}\,3) = T - i \\ &\#(\text{case}\,4) = (T-m)(T-m+1)\end{aligned}\right\} \qquad (21)$$

であるから，

$$\begin{aligned}\mathbf{E}[M^2_T(\vec{u}^{(r)})] &= T\mathbf{E}[Y_n(\vec{u}^{(r)})] \\ &\quad + \sum_{i=1}^{m-1}\Bigg(2(T-i)\prod_{k=0}^{i-1}\mathbf{E}[Y_n(u^{(r)}_k)]\mathbf{E}[Y_n(u^{(r)}_{k+m-i})] \\ &\quad \times \prod_{k=0}^{m-i-1}\{pu^{(r)}_k u^{(r)}_{k+i} + (1-p)\overline{u}^{(r)}_k\overline{u}^{(r)}_{k+i}\}\Bigg) \\ &\quad + (T-m)(T-m+1)\{\mathbf{E}[Y_n(\vec{u}^{(r)})]\}^2 \qquad (22)\end{aligned}$$

となる．したがって，式 (13) より分散は

$$\begin{aligned}\sigma^2(Z_T(\vec{u}^{(r)})) &= \mathbf{E}[Y_n(\vec{u}^{(r)})] \\ &\quad + \sum_{i=1}^{m-1}\Bigg(\frac{2(T-i)}{T}\prod_{k=0}^{i-1}\mathbf{E}[Y_n(u^{(r)}_k)]\mathbf{E}[Y_n(u^{(r)}_{k+m-i})] \\ &\quad \times \prod_{k=0}^{m-i-1}\{pu^{(r)}_k u^{(r)}_{k+i} + (1-p)\overline{u}^{(r)}_k\overline{u}^{(r)}_{k+i}\}\Bigg) \\ &\quad + \frac{(T-m)(T-m+1)}{T}\{\mathbf{E}[Y_n(\vec{u}^{(r)})]\}^2 \\ &\quad - T\{\mathbf{E}[Y_n(\vec{u}^{(r)})]\}^2 \qquad (23)\end{aligned}$$

となる．ここで，$T \to \infty$ とすると，

$$\begin{aligned}\sigma^2(Z_\infty(\vec{u}^{(r)})) &= \mathbf{E}[Y_n(\vec{u}^{(r)})] \\ &\quad + \sum_{i=1}^{m-1}\Bigg(2\prod_{k=0}^{i-1}\mathbf{E}[Y_n(u^{(r)}_k)]\mathbf{E}[Y_n(u^{(r)}_{k+m-i})] \\ &\quad \times \prod_{k=0}^{m-i-1}\{pu^{(r)}_k u^{(r)}_{k+i} + (1-p)\overline{u}^{(r)}_k\overline{u}^{(r)}_{k+i}\}\Bigg) \\ &\quad + (1-2m)\{\mathbf{E}[Y_n(\vec{u}^{(r)})]\}^2 \qquad (24)\end{aligned}$$

となる．

4 i.i.d. 2値系列から生成される確率ベクトルのクラスに関する分布

$\vec{u}^{(r)}$ のパタン数は，m が大となると膨大となるので，事象の数を減らすため，各パタンに存在する1の個数でクラス分けすることを考える．パタン中に，1が ℓ 個，0が $m-\ell$ 個存在する m 次元2値ベクトルの全ての集合を $C^{(\ell)}$ と定義する．$C^{(\ell)}$ は，全てのパタンを元とする m 次元2値ベクトルの集合 $\{\vec{u}^{(r)}\}_{r=0}^{2^m-1}$ の部分集合である．ω を初期値とする i.i.d. 2値確率変数の列 $\{X_n(\omega)\}_{n=0}^{\infty}$ に対して m 次元ベクトル

$$\vec{X}_n(\omega) = (X_n(\omega), X_{n+1}(\omega), \cdots, X_{n+m-1}(\omega))^T \tag{25}$$

が $C^{(\ell)}$ の要素となる確率を考察する．

$$Y_n(\omega; C^{(\ell)}) = \sum_{\vec{u}^{(r)} \in C^{(\ell)}} Y_n(\omega; \vec{u}^{(r)}) \tag{26}$$

とおいて，$T > m$ とすると，確率変数

$$M_T(\omega; C^{(\ell)}) = \sum_{n=0}^{T-1} Y_n(\omega; C^{(\ell)}) = \sum_{\vec{u}^{(r)} \in C^{(\ell)}} \sum_{n=0}^{T-1} \prod_{k=0}^{m-1} Y_n(\omega; u_k^{(r)}) \tag{27}$$

は，$\{X_n(\omega)\}_{n=0}^{T+m-1}$ に対する $\vec{X}_n(\omega) \in C^{(\ell)}$ となる個数である．また，

$$\begin{aligned} \mathbf{E}[Y_n(C^{(\ell)})] &= \mathbf{E}\Big[\sum_{\vec{u}^{(r)} \in C^{(\ell)}} Y_n(\vec{u}^{(r)})\Big] = \sum_{\vec{u}^{(r)} \in C^{(\ell)}} \mathbf{E}[Y_n(\vec{u}^{(r)})] \\ &= \binom{m}{\ell} p^{\ell}(1-p)^{m-\ell} \end{aligned} \tag{28}$$

となる．確率変数 $Y_n(\omega; C^{(\ell)})$ の振舞いを調べるために§2と同様に新しい確率変数

$$Z_T(\omega; C^{(\ell)}) = \frac{M_T(\omega; C^{(\ell)}) - T\mathbf{E}[Y_n(C^{(\ell)})]}{\sqrt{T}} \tag{29}$$

を導入する．明らかに

$$\sigma^2(Z_T(C^{(\ell)})) = \mathbf{E}[Z_T^2(C^{(\ell)})] = \frac{1}{T}\mathbf{E}[M_T^2(C^{(\ell)})] - T\{\mathbf{E}[Y_n(C^{(\ell)})]\}^2 \tag{30}$$

である．式 (27) の定義より，

$$M_T^2(\omega; C^{(\ell)}) = \sum_{\vec{u}^{(r)} \in C^{(\ell)}} \sum_{\vec{u}^{(s)} \in C^{(\ell)}} \sum_{l=0}^{T-1} \sum_{n=0}^{T-1} I_{l,n}^{r,s}(\omega) \tag{31}$$

であり，分散は

$$
\begin{aligned}
&\sigma^2(Z_T(C^{(\ell)})) \\
&= \binom{m}{\ell}p^{\ell}(1-p)^{m-\ell} + \sum_{\vec{u}^{(r)}\in C^{(\ell)}}\sum_{\vec{u}^{(s)}\in C^{(\ell)}}\Bigg\{\sum_{i=1}^{m-1}\Bigg\{\frac{(T-i)}{T}\Bigg(\\
&\quad \prod_{k=0}^{i-1}\mathbf{E}[Y_n(u_k^{(s)})]\mathbf{E}[Y_n(u_{k+m-i}^{(r)})]\prod_{k=0}^{m-i-1}\{pu_k^{(r)}u_{k+i}^{(s)} + (1-p)\overline{u}_k^{(r)}\overline{u}_{k+i}^{(s)}\} \\
&\quad + \prod_{k=0}^{i-1}\mathbf{E}[Y_n(u_k^{(r)})]\mathbf{E}[Y_n(u_{k+m-i}^{(s)})]\prod_{k=0}^{m-i-1}\{pu_{k+i}^{(r)}u_k^{(s)} + (1-p)\overline{u}_{k+i}^{(r)}\overline{u}_k^{(s)}\}\Bigg)\Bigg\}\Bigg\} \\
&+ \binom{m}{\ell}^2\frac{(T-m)(T-m+1)}{T}p^{2\ell}(1-p)^{2(m-\ell)} - T\Bigg\{\binom{m}{\ell}p^{\ell}(1-p)^{m-\ell}\Bigg\}^2 \qquad (32)
\end{aligned}
$$

となる．ここで，$T\to\infty$とすると，

$$
\begin{aligned}
&\sigma^2(Z_\infty(C^{(\ell)})) \\
&= \binom{m}{\ell}p^{\ell}(1-p)^{m-\ell} + \sum_{\vec{u}^{(r)}\in C^{(\ell)}}\sum_{\vec{u}^{(s)}\in C^{(\ell)}}\Bigg\{ \\
&\quad \sum_{i=1}^{m-1}\Bigg(\prod_{k=0}^{i-1}\mathbf{E}[Y_n(u_k^{(s)})]\mathbf{E}[Y_n(u_{k+m-i}^{(r)})]\prod_{k=0}^{m-i-1}\{pu_k^{(r)}u_{k+i}^{(s)} + (1-p)\overline{u}_k^{(r)}\overline{u}_{k+i}^{(s)}\} \\
&\quad + \prod_{k=0}^{i-1}\mathbf{E}[Y_n(u_k^{(r)})]\mathbf{E}[Y_n(u_{k+m-i}^{(s)})]\prod_{k=0}^{m-i-1}\{pu_{k+i}^{(r)}u_k^{(s)} + (1-p)\overline{u}_{k+i}^{(r)}\overline{u}_k^{(s)}\}\Bigg)\Bigg\} \\
&+ \binom{m}{\ell}^2(1-2m)p^{2\ell}(1-p)^{2(m-\ell)} \qquad (33)
\end{aligned}
$$

となる．

5 むすび

本稿では，等確率でない一般のi.i.d. 2値系列から生成される確率ベクトルに関するパタンとクラスの頻度を評価し，それらの経験分布が形成するガウス分布の分散の理論値を与えた．得られた分散の理論値を利用して，擬似乱数のi.i.d. 性を評価することができる．

References

[1] P. Billingsley, *Probability and Measure*, John Wiley & Sons, 1995.

[2] 香田 徹, 大賀崇弘, 常田明夫, "2値系列の頻度分布について", 電子情報通信学会技術研究報告, CAS96-10, VLD96-30, 1996.

[3] 香田 徹, 藤崎礼志, 大賀崇弘, "不等確率2値系列の頻度分布について", 電子情報通信学会技術研究報告, NLP97-39, 1997.