SCIS 2020 2020 Symposium on
Cryptography and Information Security
Kochi, Japan, Jan. 28 – 31, 2020
The Institute of Electronics,
Information and Communication Engineers

Maurer's Test を基にした Highly Sensitive Statistical Universal Test における参照分布の分散の導出

The variance of the reference distribution of Highly Sensitive Universal Test constructed on the basis of Maurer's Test

引間 泰成 * 岩崎 淳 * 梅野 健 *
Yasunari Hikima Atsushi Iwasaki Ken Umeno

あらまし Maurer's universal statistical test は 2 値ビット列の非ランダム性を検出する統計検定であり，NIST SP 800–22 に採用されている．Maurer's test では，ビット列のエントロピーに基づく参照分布を計算する．さらに検出力を高めるため，系列中の "1" を一定の確率で "0" に変換することによって，"1" の発生頻度が $\hat{q}$ であるように系列を変換し，変換後の系列に対して検定を行う手法が提案された．この手法では，p 値の算出の際に参照分布の分散が必要であるが，一般の $\hat{q}$ に対する分散を計算する方法は提示されておらず，擬似乱数を用いたシミュレーションによって算出された値が使用されている．乱数の検定を行う上で，定数であるパラメータの算出という本質的でない部分に乱数を使用していることは好ましいとは言えない．そこで本研究では，参照分布の分散を理論的に導出する．

キーワード 乱数，NIST SP 800–22

1 序論

Maurer's universal statistical test[1] は与えられた 2 値ビット列の非ランダム性を検出する手法であり，NIST SP 800–22 の初版から最新版までを通して採用されている乱数検定手法である [2, 3]．同様の手法に Coron による検定手法がある [4]．両者の検定はそれぞれ次のように行われる．まず，検定対象のビット列 $x^n = x_1, x_2, \ldots, x_n$ を重なりのない L ビットごとのブロックに分割し，先頭から Q ブロックを初期化用セグメントとし，残りの K ブロックを検定用セグメントとする．なお，簡単のため $n = L \times (Q + K)$ が成り立つとする．また，ビット列の k 番目のブロックを b_k で表す．すなわち，$b_k = x_{L(k-1)+1}, x_{L(k-1)+2}, \ldots, x_{Lk}$ である．各ブロックごとに，それと同一の値であるブロックが何ブロック前にあるかを表す変数を以下で定義する：

$$A_{n-Q} = \begin{cases} n, \text{ if } b_{n-m} \neq b_n \text{ for } 1 \leq m \leq n-1, \\ \min\{m \mid m \geq 1, b_{n-m} = b_n\}, \text{ otherwise.} \end{cases} \tag{1}$$

式 (1) で定義される A_n を用いて，Maurer による参照分布 $f_M : \{0,1\}^n \to \mathbb{R}$ および Coron による参照分布 $f_C : \{0,1\}^n \to \mathbb{R}$ はそれぞれ次のように定義される：

$$f_M(x^n) = \frac{1}{K} \sum_{n=1}^{K} \log_2 A_n, \tag{2}$$

$$f_C(x^n) = \frac{1}{K} \sum_{n=1}^{K} g(A_n). \tag{3}$$

ここで，式 (3) における関数 $g : \mathbb{N} \to \mathbb{R}$ は

$$g(m) = (\log_2 \mathrm{e}) \sum_{k=1}^{m-1} \frac{1}{k} \tag{4}$$

で定義される．参照分布 $f_M(x^n)$ および $f_C(x^n)$ から p 値への変換はそれぞれ以下で与えられる：

$$p\text{-value}_M = \mathrm{erfc}\left(\left| \frac{f_M(x^n) - \mathbb{E}[f_M(x^n)]}{\sqrt{2}\sigma_M} \right| \right), \tag{5}$$

$$p\text{-value}_C = \mathrm{erfc}\left(\left| \frac{f_C(x^n) - \mathbb{E}[f_C(x^n)]}{\sqrt{2}\sigma_C} \right| \right). \tag{6}$$

ここに，σ_M^2 および σ_C^2 は帰無仮説の下での $f_M(x^n)$ および $f_C(x^n)$ の分散を表し，erfc は次式で定義される相補誤差関数を表す：

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty \mathrm{e}^{-t^2} \, \mathrm{d}t. \tag{7}$$

* 京都大学大学院情報学研究科，〒606-8317 京都府京都市左京区吉田本町 36-1, Graduate School of Informatics, Kyoto University 36-1 Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN

2 値ビット列 x^n の各ビットが独立同分布に従い，“1” である確率と “0” である確率がともに 0.5 であるとき，式 (2) で定義される参照分布は次のような関係を満たすことが知られている [1]：

$$\lim_{L\to\infty}\left[\mathbb{E}[f_M(x^n)] - L\times H(q)\right] = C. \tag{8}$$

ここに，$C = \int_0^\infty e^{-\xi}\log_2\xi\,\mathrm{d}\xi \simeq -0.8327462$ を表し，$H(q)$ は “1” である確率が q，“0” である確率が $1-q$ に対応する 2 値エントロピーを表す．しかしながら，式 (8) は $q \neq 0.5$ の場合には成り立たないことが示されている [5]．一方，式 (3) で定義される参照分布は任意の $q \in (0,1)$ に対して，次のような関係を満たす [4]：

$$\mathbb{E}[f_C(x^n)] = L\times H(q). \tag{9}$$

両者の検定は，式 (8) または式 (9) が $q = 0.5$ において成立することを基礎として系列の非ランダム性を評価している．しかしながら，2 値エントロピー関数の導関数 $\frac{\mathrm{d}}{\mathrm{d}q}H(q) = \log_2\frac{1-q}{q}$ は $q = 0.5$ において 0 になることから，$q = 0.5$ からの偏りを高い感度で検出することは困難であることが指摘されている [6]．そこで，検定対象である 2 値ビット列 x^n における “1” を一定の確率で “0” に変換し，変換後の系列 $\hat{x}^n$ に対して Coron と同様の検定を行う手法 (“Highly Sensitive Universal Statistical Test”) が提案された [6]．この手法によって，Maurer や Coron の手法よりも高い感度で系列の非ランダム性を検出できることが数値実験を通して示唆されている．Highly Sensitive Test では，検定を行う上で次式で定義される p 値を計算する：

$$p\text{-value} = \mathrm{erfc}\left(\left|\frac{f_C(\hat{x}^n) - L\times H(\hat{q})}{\sqrt{2}\sigma_C(\hat{q})}\right|\right). \tag{10}$$

ここに，$\hat{q}$ は変換後の系列 $\hat{x}^n$ における各ビットが “1” である確率を表す．もし，$\hat{q} = 0.5$ であれば，分散 $\sigma(\hat{q})^2$ は σ_C^2 に一致し，次式で与えられることが知られている [5]：

$$\sigma_C^2 = c(L,K)^2\frac{\mathrm{Var}[g(A_n)]}{K}. \tag{11}$$

ここに，$c(L,K)$ は次式で近似される定数である：

$$c(L,K)^2 \simeq d(L) + \frac{e(L)\times 2^L}{K}. \tag{12}$$

式 (12) における $d(L)$ および $e(L)$ の値は文献 [4] において表で与えられている．一方，$\hat{q} \neq 0.5$ の場合における分散 $\sigma_C(\hat{q})^2$ を求める式は提示されておらず，文献 [6] では擬似乱数を用いたシミュレーションによって算出された数値を用いている．乱数の検定を行う上で，定数であるパラメータの算出という本質的でない部分に乱数を用いることは好ましいとは言えない．そこで本研究では，式 (3) で定義される参照分布の分散 $\sigma_C(\hat{q})^2$ を任意の $\hat{q} \in (0,1)$ に対して理論的に導出する．

2 分布

帰無仮説の下で $Q \to \infty$ とすると，系列 $\{A_k\}_{k=1}^K$ は明らかに stationary ergodic である．すなわち，任意の m, n について，$\{A_k\}_{k=n}^{n+m}$ の同時分布は n に依らず，m にのみ依存する．本節では，Highly Sensitive Universal Statistical Test における帰無仮説の下での $f_C(\hat{x}^n)$ の分散を計算するために必要な A_n の分布および (A_n, A_{n+k}) の同時分布を導出する．このとき，$Q \to \infty$ を考えると，A_k の分布は k に依らず，$(A_k, A_{k'})$ の同時分布は $k'-k$ にのみ依存する．

2.1 周辺分布の導出

ここでは，$\langle A_n = i\rangle$ となる確率を求める．この事象は，n 番目のブロックと $n-i$ 番目のブロックが一致し，かつ，それらの間のブロックとは一致しないような事象である．そのような事象を $\mathcal{M}$ で表すと，$\mathcal{M}$ は

$$\mathcal{M} = \langle b_{n-i} = b_n, b_{n-i+1} \neq b_n, \ldots, b_{n-1} \neq b_n\rangle \tag{13}$$

と書ける．すると，求める周辺分布は次のように計算することができる：

$$\Pr[A_n = i] = \sum_{r=0}^{L}\Pr[\mathcal{M} \mid l(b_n) = r]\times\Pr[l(b_n) = r]. \tag{14}$$

ここに，$l(b)$ はブロック $b \in \{0,1\}^L$ における “1” の個数を表す．また，

$$\Pr[\mathcal{M} \mid l(b_n) = r] = \mathcal{Q}_r\times(1-\mathcal{Q}_r)^{i-1}, \tag{15}$$

$$\Pr[l(b_n) = r] = \binom{L}{r}\mathcal{Q}_r. \tag{16}$$

ここに，$\mathcal{Q}_r = \hat{q}^r(1-\hat{q})^{L-r}$ であり，以降でも同様の記号を用いる．式 (14)〜(16) より，事象 $\mathcal{M}$ が生起する確率は次式で与えられる：

$$\Pr[A_n = i] = \sum_{r=0}^{L}\binom{L}{r}\mathcal{Q}_r^2(1-\mathcal{Q}_r)^{i-1}. \tag{17}$$

2.2 同時分布の導出

ここでは，$\langle A_n = i,\, A_{n+k} = j\rangle$ となる確率，すなわち，$\Pr[A_n = i, A_{n+k} = j]$ を求める．以下では，集合 $\{0,1\}^L$ を B^L で表す．なお，本結果は文献 [5] の拡張である．

2.2.1 $1 \le j \le k-1$ のとき

事象 $\langle A_n = i \rangle$ は事象 $\langle A_{n+k} = j \rangle$ に影響を及ぼさない．したがって，求める同時分布は式 (17) より次のように計算される：

$$\begin{aligned}
&\Pr[A_n = i, A_{n+k} = j] \\
&= \Pr[A_n = i] \times \Pr[A_{n+k} = j] \\
&= \left(\sum_{r=0}^{L} \binom{L}{r} \mathcal{Q}_r^2 (1-\mathcal{Q}_r)^{i-1} \right) \\
&\quad \times \left(\sum_{r=0}^{L} \binom{L}{r} \mathcal{Q}_r^2 (1-\mathcal{Q}_r)^{j-1} \right).
\end{aligned} \tag{18}$$

2.2.2 $j = k$ のとき

まず，各 $b \in B^L$ に対して，

$$\begin{aligned}
e_2(b) = \langle & b_{n-i} = b, b_n = b, b_{n+k} = b, \\
& b_{n-i+1} \neq b, \ldots, b_{n-1} \neq b, \\
& b_{n+1} \neq b, \ldots, b_{n+k-1} \neq b \, \rangle
\end{aligned} \tag{19}$$

とおく．これを用いると，$\langle A_n = i,\, A_{n+k} = j \rangle$ が生起する事象 $\mathcal{E}_2$ は次のように表される：

$$\mathcal{E}_2 = \bigcup_{b \in B^L} e_2(b). \tag{20}$$

いま，各ブロックは互いに独立であるので，事象 $e_2(b)$ が生起する確率は，$r = l(b)$ として，次のように計算される：

$$\Pr\left[e_2(b)\right] = \mathcal{Q}_r^3 \times (1-\mathcal{Q}_r)^{i+k-2}. \tag{21}$$

よって，事象 $\mathcal{E}_2$ が生起する確率は次のように求めることができる：

$$\begin{aligned}
\Pr[\mathcal{E}_2] &= \Pr\left[\bigcup_{b \in B^L} e_2(b) \right] \\
&= \sum_{b \in B^L} \Pr[e_2(b)] \\
&= \sum_{b \in B_0^L \cup \cdots \cup B_L^L} \Pr[e_2(b)] \\
&= \sum_{r=0}^{L} \sum_{b \in B_r^L} \Pr\left[e_2(b)\right] \\
&= \sum_{r=0}^{L} \left(\sharp B_r^L \right) \Pr\left[e_2(b)\right] \\
&= \sum_{r=0}^{L} \binom{L}{r} \Pr\left[e_2(b)\right].
\end{aligned} \tag{22}$$

ここに，$B_r^L := \{ b \in B^L \mid l(b) = r \}$，$\sharp S$ は集合 S の要素の個数を表す．以上より，求める同時分布は次のように表される：

$$\begin{aligned}
&\Pr\left[A_n = i,\, A_{n+k} = j\right] \\
&= \Pr[\mathcal{E}_2] \\
&= \sum_{r=0}^{L} \binom{L}{r} \Pr\left[e_2(b)\right] \\
&= \sum_{r=0}^{L} \binom{L}{r} \mathcal{Q}_r^3 (1-\mathcal{Q}_r)^{i+k-2}.
\end{aligned} \tag{23}$$

2.2.3 $k+1 \le j \le k+i-1$ のとき

まず，各 $b_1, b_2 \in B^L$ に対して，

$$\begin{aligned}
&e_3(b_1, b_2) := \\
&\quad \langle b_{n-i} = b_1, b_n = b_1, b_{n+k-j} = b_2, b_{n+k} = b_2 \rangle \\
&\wedge \langle b_{n-i+1} \neq b_1, \ldots, b_{n+k-j-1} \neq b_1 \rangle \\
&\wedge \langle b_{n+k-j+1} \neq b_1, \ldots, b_{n-1} \neq b_1 \rangle \\
&\wedge \langle b_{n+k-j+1} \neq b_2, \ldots, b_{n-1} \neq b_2 \rangle \\
&\wedge \langle b_{n+1} \neq b_2, \ldots, b_{n+k-1} \neq b_2 \rangle
\end{aligned} \tag{24}$$

とおく．これを用いると，$\langle A_n = i,\, A_{n+k} = j \rangle$ が生起する事象 $\mathcal{E}_3$ は次のように表される：

$$\mathcal{E}_3 = \bigcup_{b_1 \in B^L} \bigcup_{b_2 \in B^L \setminus \{b_1\}} e_3(b_1, b_2). \tag{25}$$

いま，各ブロックは互いに独立であるので，事象 $e_3(b_1, b_2)$ が生起する確率は，$r_1 = l(b_1)$, $r_2 = l(b_2)$ として，次のように計算される：

$$\begin{aligned}
\Pr\left[e_3(b_1, b_2)\right] = & \mathcal{Q}_{r_1}^2 \times \mathcal{Q}_{r_2}^2 \times (1-\mathcal{Q}_{r_1})^{i-j+k-1} \\
& \times (1-\mathcal{Q}_{r_1}-\mathcal{Q}_{r_2})^{j-k-1} \\
& \times (1-\mathcal{Q}_{r_2})^{k-1}.
\end{aligned} \tag{26}$$

よって，事象 $\mathcal{E}_3$ が生起する確率は次のように求めることができる：

$$\begin{aligned}
\Pr[\mathcal{E}_3] =& \Pr\left[\bigcup_{b_1\in B^L}\bigcup_{b_2\in B^L\setminus\{b_1\}} e_3(b_1,b_2)\right] \\
=& \sum_{b_1\in B^L}\sum_{b_2\in B^L\setminus\{b_1\}} \Pr\left[e_3(b_1,b_2)\right] \\
=& \sum_{r_1=0}^{L}\sum_{b_1\in B^L_{r_1}}\sum_{r_2=0}^{L}\sum_{b_2\in B^L_{r_2}\setminus\{b_1\}} \Pr\left[e_3(b_1,b_2)\right] \\
=& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\sum_{b_1\in B^L_{r_1}}\sum_{b_2\in B^L_{r_2}} \Pr\left[e_3(b_1,b_2)\right] \\
&+ \sum_{r_1=0}^{L}\sum_{b_1\in B^L_{r_1}}\sum_{b_2\in B^L_{r_1}\setminus\{b_1\}} \Pr\left[e_3(b_1,b_2)\right] \\
=& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\binom{L}{r_1}\binom{L}{r_2}\Pr\left[e_3(b_1,b_2)\right] \\
&+ \sum_{r_1=0}^{L}\sum_{r_2\in\{r_1\}}\binom{L}{r_1}\left\{\binom{L}{r_1}-1\right\}\Pr\left[e_3(b_1,b_2)\right].
\end{aligned} \tag{27}$$

以上より，求める同時分布は次のように計算される：

$$\begin{aligned}
&\Pr\left[A_n=i, A_{n+k}=j\right] \\
=&\Pr[\mathcal{E}_3] \\
=& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\binom{L}{r_1}\binom{L}{r_2}\Pr\left[e_3(b_1,b_2)\right] \\
&+ \sum_{r_1=0}^{L}\sum_{r_2\in\{r_1\}}\binom{L}{r_1}\left\{\binom{L}{r_1}-1\right\}\Pr\left[e_3(b_1,b_2)\right].
\end{aligned} \tag{28}$$

ここで，$\Pr\left[e_3(b_1,b_2)\right]$ は式 (26) で与えられる．

2.2.4 $j=k+i$ のとき

2 つの事象 $\langle A_n=i\rangle$ と $\langle A_{n+k}=j\rangle$ は同時に生起しない．したがって，求める同時分布は

$$\Pr[A_n=i, A_{n+k}=j]=0 \tag{29}$$

である．

2.2.5 $j\geq k+i+1$ のとき

まず，各 $b_1,b_2\in B^L$ に対して，

$$\begin{aligned}
&e_5(b_1,b_2) := \\
&\quad\langle b_{n+k-j}=b_1, b_{n-i}=b_2, b_n=b_2, b_{n+k}=b_1\rangle \\
&\wedge\langle b_{n+k-j+1}\neq b_1,\ldots,b_{n-i-1}\neq b_1\rangle \\
&\wedge\langle b_{n-i+1}\neq b_1,\ldots,b_{n-1}\neq b_1\rangle \\
&\wedge\langle b_{n-i+1}\neq b_2,\ldots,b_{n-1}\neq b_2\rangle \\
&\wedge\langle b_{n+1}\neq b_1,\ldots,b_{n+k-1}\neq b_1\rangle
\end{aligned} \tag{30}$$

とおく．これを用いると，$\langle A_n=i,\, A_{n+k}=j\rangle$ が生起する事象 $\mathcal{E}_5$ は次のように表される：

$$\mathcal{E}_5=\bigcup_{b_1\in B^L}\bigcup_{b_2\in B^L\setminus\{b_1\}} e_5(b_1,b_2). \tag{31}$$

いま，各ブロックは互いに独立であるので，事象 $e_5(b_1,b_2)$ が生起する確率は，$r_1=l(b_1)$, $r_2=l(b_2)$ として，次のように計算される：

$$\begin{aligned}
\Pr\left[e_5(b_1,b_2)\right] =& \mathcal{Q}_{r_1}^2\times\mathcal{Q}_{r_2}^2\times(1-\mathcal{Q}_{r_1})^{-i+j-k-1} \\
&\times(1-\mathcal{Q}_{r_1}-\mathcal{Q}_{r_2})^{i-1} \\
&\times(1-\mathcal{Q}_{r_1})^{k-1}.
\end{aligned} \tag{32}$$

よって，事象 $\mathcal{E}_5$ が生起する確率は次のように求めることができる：

$$\begin{aligned}
\Pr[\mathcal{E}_5] =& \Pr\left[\bigcup_{b_1\in B^L}\bigcup_{b_2\in B^L\setminus\{b_1\}} e_5(b_1,b_2)\right] \\
=& \sum_{b_1\in B^L}\sum_{b_2\in B^L\setminus\{b_1\}} \Pr\left[e_5(b_1,b_2)\right] \\
=& \sum_{r_1=0}^{L}\sum_{b_1\in B^L_{r_1}}\sum_{r_2=0}^{L}\sum_{b_2\in B^L_{r_2}\setminus\{b_1\}} \Pr\left[e_5(b_1,b_2)\right] \\
=& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\sum_{b_1\in B^L_{r_1}}\sum_{b_2\in B^L_{r_2}} \Pr\left[e_5(b_1,b_2)\right] \\
&+ \sum_{r_1=0}^{L}\sum_{b_1\in B^L_{r_1}}\sum_{b_2\in B^L_{r_1}\setminus\{b_1\}} \Pr\left[e_5(b_1,b_2)\right] \\
=& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\binom{L}{r_1}\binom{L}{r_2}\Pr\left[e_5(b_1,b_2)\right] \\
&+ \sum_{r_1=0}^{L}\sum_{r_2\in\{r_1\}}\binom{L}{r_1}\left\{\binom{L}{r_1}-1\right\}\Pr\left[e_5(b_1,b_2)\right].
\end{aligned} \tag{33}$$

以上より，求める同時分布は次のように計算される：

$$\begin{aligned}
&\Pr\left[A_n=i, A_{n+k}=j\right] \\
=&\Pr[\mathcal{E}_5] \\
=& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\binom{L}{r_1}\binom{L}{r_2}\Pr\left[e_5(b_1,b_2)\right] \\
&+ \sum_{r_1=0}^{L}\sum_{r_2\in\{r_1\}}\binom{L}{r_1}\left\{\binom{L}{r_1}-1\right\}\Pr\left[e_5(b_1,b_2)\right].
\end{aligned} \tag{34}$$

ここで，$\Pr\left[e_5(b_1,b_2)\right]$ は式 (32) で与えられる．

3 参照分布の分散

本節では，前節の結果を用いて，式 (3) で定義される参照分布の分散 $\sigma_C(\hat{q})^2$ を任意の $\hat{q}\in(0,1)$ に対して導出す

る．また，計算機実験を通して様々な $\hat{q}$ に対して分散が正しく計算できていることを確認する．

3.1 分散の導出

確率変数 X の分散は次式で定義される：

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] \tag{35}$$
$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{36}$$

また，任意の確率変数 $X_1, X_2, \ldots, X_n$ に対して，確率変数の和の分散は次のように与えられる：

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}[X_i, X_j]. \tag{37}$$

ここに，$\mathrm{Cov}[X_i, X_j]$ は X_i と X_j の共分散を表し，次式で定義される：

$$\mathrm{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \times \mathbb{E}[X_j]. \tag{38}$$

式 (37) を用いることで，式 (3) で定義される参照分布 $f_C(\hat{x}^n)$ の分散は次のように計算される：

$$\begin{aligned}
\sigma_C(\hat{q})^2 =& \mathrm{Var}[f_C(\hat{x}^n)] \\
=& \mathrm{Var}\left[\frac{1}{K}\sum_{n=Q+1}^{K+Q} g(A_n)\right] \\
=& \frac{1}{K^2}\left(\sum_{n=Q+1}^{K+Q} \mathrm{Var}[g(A_n)]\right. \\
& \left. + 2\sum_{1 \le i < j \le K} \mathrm{Cov}[g(A_{Q+i}), g(A_{Q+j})]\right) \\
=& \frac{1}{K^2}\left(K \times \mathrm{Var}[g(A_n)]\right. \\
& \left. + 2\sum_{k=1}^{K-1} (K-k)\mathrm{Cov}[g(A_n), g(A_{n+k})]\right).
\end{aligned} \tag{39}$$

式 (39) の最後の等式は，stationary ergodic の仮定より $g(A_n)$ の分散は n に依らず，$g(A_{Q+i})$ と $g(A_{Q+j})$ の共分散が $k = j - i$ にのみ依存することより従う．

式 (39) における $\mathrm{Var}[g(A_n)]$ は式 (36) より次のように計算される：

$$\mathrm{Var}[g(A_n)] = \mathbb{E}[\{g(A_n)\}^2] - (\mathbb{E}[g(A_n)])^2. \tag{40}$$

ここで，$\mathbb{E}[\{g(A_n)\}^2]$ は期待値の定義より

$$\mathbb{E}[\{g(A_n)\}^2] = \sum_{i=1}^{\infty} \{g(i)\}^2 \Pr[A_n = i] \tag{41}$$

表 1　各 q に対する $D_K(q)$ の値

q	$D_{5000}(q)$	$D_{10000}(q)$
0.33	1.871108549131227	1.867364036867157
0.4	1.331215863269216	1.328691718766912
0.5	1.030286874123153	1.028395217703927

である．なお，式 (41) における $\Pr[A_n = i]$ は式 (17) で与えられる．また，$\mathbb{E}[g(A_n)]$ は stationary ergodic の仮定の下で次のようにして求めることができる：

$$\begin{aligned}
\mathbb{E}[g(A_n)] &= \mathbb{E}\left[\frac{1}{K} \times K g(A_n)\right] \\
&= \mathbb{E}\left[\frac{1}{K}\sum_{n=Q+1}^{Q+K} g(A_n)\right] \\
&= \mathbb{E}[f_C(\hat{x}^n)] \\
&= L \times H(\hat{q}).
\end{aligned} \tag{42}$$

以上より，$\mathrm{Var}[g(A_n)]$ は次のように表される：

$$\mathrm{Var}[g(A_n)] = \sum_{i=1}^{\infty} \{g(i)\}^2 \Pr[A_n = i] - \{L \times H(\hat{q})\}^2. \tag{43}$$

一方，式 (39) における共分散 $\mathrm{Cov}[g(A_n), g(A_{n+k})]$ は式 (38) より次のように計算される：

$$\begin{aligned}
&\mathrm{Cov}\left[g(A_n),\, g(A_{n+k})\right] \\
&= \mathbb{E}[g(A_n)g(A_{n+k})] - \mathbb{E}[g(A_n)] \times \mathbb{E}[g(A_{n+k})] \\
&= \sum_{i,j \ge 1} g(i)g(j)\Pr\left[A_n = i,\, A_{n+k} = j\right] - \left\{L \times H(\hat{q})\right\}^2.
\end{aligned} \tag{44}$$

ここで，同時分布 $\Pr\left[A_n = i,\, A_{n+k} = j\right]$ は 2 節で算出した通りである．

3.2 計算機実験

ここでは，$\sigma_C(\hat{q})^2$ が正しく計算できることを確認する．各パラメータは，$L = 4$, $Q = 10 \cdot 2^L$ とし，式 (41) と式 (44) における無限和は 10^6 で打ち切ることとする．図 1 は $\hat{q} = 0.33, 0.4, 0.5$ の場合における参照分布の分散を各 K に対して求めた値をプロットしたものである．図 1 より，分散は $\mathcal{O}\left(\frac{1}{K}\right)$ で減少することが見てとれる．また，ある K を固定したときのその係数 $D_K(q)$ の値を表 1 にまとめた．表 1 からも $K = 5000$ 程度で分散が $\frac{\mathrm{const}}{K}$ に収束していることが確認できる．なお，$\hat{q} = 0.33$ としたのは，文献 [6] において検出力を高める上で最適であるとされていたことに因る．

次に，参照分布の分散の値を擬似乱数生成器から生成される 2 値ビット列を用いてシミュレーションした結果に

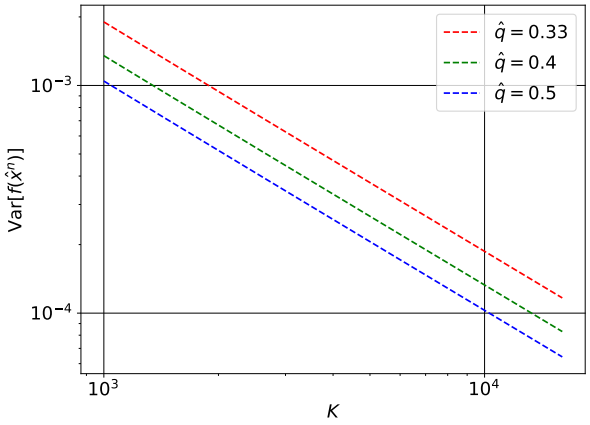

図 1　参照分布の分散

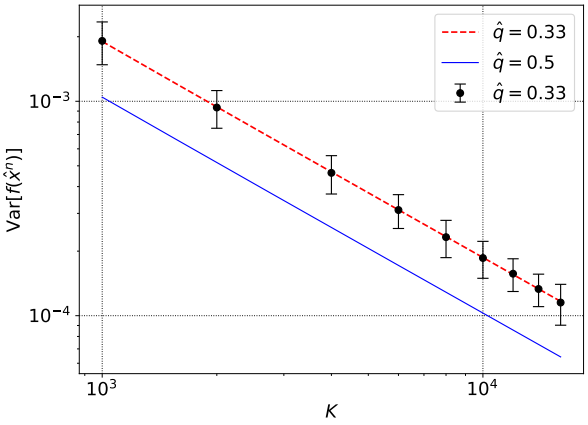

図 2　擬似乱数を用いたシミュレーションに基づく参照分布の分散．実線および破線は $\hat{q} = 0.5$ および $\hat{q} = 0.33$ における参照分布の分散を表す．点はシミュレーションによって算出された不偏分散の平均値，エラーバーはシミュレーションによって算出された不偏分散のばらつきをそれぞれ表す．

ついて述べる．実験手順は以下の通りである．まず，パラメータとして L, Q, K, $\hat{q}$ をそれぞれ設定する．次に，擬似乱数生成器によって 2 値ビット列を M 本生成し，各系列 $(x^{n,1}, x^{n,2}, \ldots, x^{n,M})$ に対して式 (3) で定義される参照分布 $f_i = f_C(x^{n,i})$ $(i = 1, 2, \ldots, M)$ を計算する：

$$u^2 = \frac{1}{M-1} \sum_{i=1}^{M} \left(f_i - \overline{f}\right)^2. \tag{45}$$

ここに，$\overline{f}$ は次式で定義される：

$$\overline{f} = \frac{1}{M} \sum_{i=1}^{M} f_i. \tag{46}$$

以上を計 N 回繰り返し，得られた N 個の不偏分散 $u_1^2, u_2^2, \ldots, u_N^2$ の平均値を次式で計算し，参照分布の分散のシミュレーション結果とする：

$$\overline{u}^2 = \frac{1}{N} \sum_{i=1}^{N} u_i^2. \tag{47}$$

図 2 は本実験に対する結果を表したものである．ここでは，擬似乱数生成器としてメルセンヌ・ツイスタ [7] を使用し，各パラメータは，$L = 4$, $Q = 10 \times 2^L$, $\hat{q} = 0.33$, $M = 1000$, $N = 30$ を与えた．図 2 より，本実験による不偏分散の平均値は式 (39) による結果と高精度で一致することが確認できる．

4　まとめ

本稿では，Maure's Universal Statistical Test を基にした Highly Sensitive Universal Statistical Test における参照分布の分散を理論的に導出した．ここで，求める分散は式 (39) で与えられる (再掲)：

$$\sigma_C(\hat{q})^2 = \frac{1}{K}\mathrm{Var}[g(A_n)] + \frac{2}{K^2} \sum_{k=1}^{K-1} (K-k)\mathrm{Cov}[g(A_n), g(A_{n+k})].$$

分散の具体的な値は，$\mathrm{Var}[g(A_n)]$ と $\mathrm{Cov}[g(A_n), g(A_{n+k})]$ を，式 (17) で与えられる周辺分布 $\Pr[A_n = i]$ と式 (18),(23),(28),(29),(34) で与えられる同時分布 $\Pr[A_n = i, A_{n+k} = j]$ を用いて計算し，代入することで求まる．

また，計算機実験を通して，導出した式によって参照分布の分散が正しく計算されることを確認した．

参考文献

[1] U. M. Maurer, "A universal statistical test for random bit generators," *Journal of cryptology*, vol. 5, no. 2, pp. 89–105, 1992.

[2] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, and E. Barker, "A statistical test suite for random and pseudorandom number generators for cryptographic applications," Tech. Rep., Booz-Allen and Hamilton Inc Mclean Va, 2001.

[3] L. E. Bassham III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, *et al.*, "Sp 800-22 rev. 1a. a statistical test suite for random

and pseudorandom number generators for cryptographic applications," 2010.

[4] J.-S. Coron, "On the security of random sources," in *International Workshop on Public Key Cryptography*, pp. 29–42, Springer, 1999.

[5] J.-S. Coron and D. Naccache, "An accurate evaluation of maurer' s universal test," in *International Workshop on Selected Areas in Cryptography*, pp. 57–71, Springer, 1998.

[6] H. Yamamoto and Q. Liu, "Highly sensitive universal statistical test," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 700–704, IEEE, 2016.

[7] M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 8, no. 1, pp. 3–30, 1998.