ＦＩＰＳ１４０－２（Change Notice 1）乱数検定
－分布関数の導出とＲＰＧ１００の測定結果－

アナンダ　ビターナゲ†　　清水　隆邦‡

２００３年１０月２日

FDK 株式会社　RPG 推進室

〒972-8322 福島県いわき市常磐上湯長谷町釜の前 1

E-mail:　†kalin@fdk.co.jp,　‡tshimizu@fdk.co.jp

導入　米国連邦政府の情報処理設備調達基準ＦＩＰＳ(Federal Information Processing Standards)のうち，FIPS PUB 140-2 には，暗号モジュールに関する基準が設けられている．基準には，乱数に関する項目も設けられており，それは４種類の統計検定で構成されている．ＦＩＰＳには，これら検定において合格とされる統計量の範囲が記述されているものの，棄却率とその根拠となる分布関数は記述されていない．ここでは，各統計検定で扱われる統計量の分布関数を導出し，それを用いて乱数検定における棄却率を算出する．また，RPG100 の乱数により得られた統計量の分布も理論値と併せて示す．

1. The Monobit Test

二値数列 20,000bit 中に現れる，１の値の数を数える．

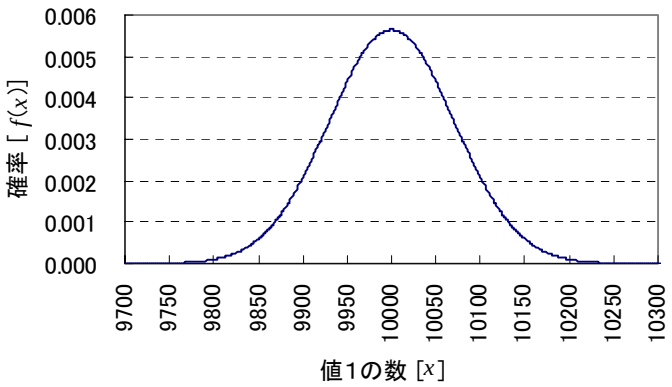
統計量 x　：20,000bit 中の１の数

採択域　　：$9{,}725 < x < 10{,}275$

統計量の分布関数：

これは，確率が 1/2 のベルヌーイ試行を n=20,000 回行う事に同じである．1 の数が x 個であるとき，その並べ方は ${}_nC_x = n!/(n-x)!x!$ 通りであるので，確率は次の二項分布に従う．

$$f(x) = nCx(1/2)^x(1-1/2)^{n-x} = nCx(1/2)^n \qquad (1)$$



検定の棄却率 α　：　$\alpha = 1 - \sum_{x=9726}^{10274} f(x) \cong 0.0001$

2. The Poker Test

二値数列 20,000bit を 4bit ずつのセルに区切り，全部で 5,000 個のセルを作る．セルのパターンは 2^4 = 16 種類あり，各々のパターン(パターン番号を i =0～15 とする)について出現数 g_i をカウントする．

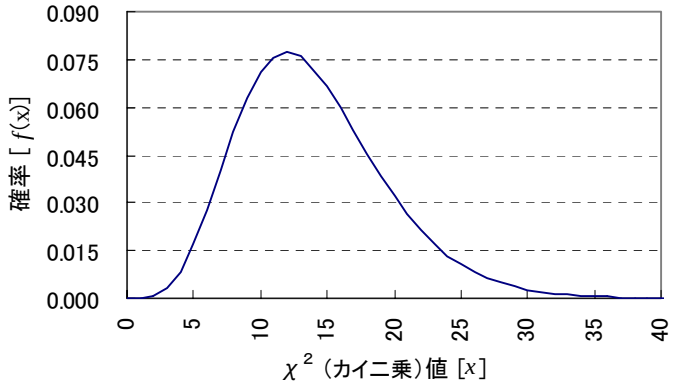
統計量 x : $x = (16/5000) \times \sum_{i=0}^{15} g_i^{\,2} - 5000$

採択域 ：$2.16 < x < 46.17$

統計量の分布関数：

ここでの統計量は χ^2 値であり，上記の式は通常用いられる式の近似となっている．各パターンの出現確率は等しく 1/16 である．パターン数は 16 だが，15 種類についての度数がわかると最後の 1 種類の度数も決まってしまうことから，χ^2 分布の自由度は 15 となっている．

$$f(x) = \frac{1}{2^{15/2}\Gamma(15/2)} x^{15/2-1} e^{-x/2} \qquad \text{ただし，} \quad \Gamma(y) \equiv \int_0^\infty z^{y-1} e^{-z} dz \qquad (2)$$



検定の棄却率 α ： $\alpha = 1 - \int_{2.16}^{46.17} f(x)dx \cong 0.0001$

3. The Runs Test

同じ数値の続きを連というが，この検定は二値数列 20,000bit 内に現れた連の個数をカウントする．検定は連の長さごとに個別に行い，その長さは 1,2,3,4,5,及び 6 以上に分けられている．さらに乱数値の 1 と 0 も区別されているので，全部で１２個の検定となっている．

統計量 x ：20,000bit 中に現れる各長さの連の数

採択域 ：

連の長さ	採択域
1	2,315 ≦ x ≦ 2,685
2	1,114 ≦ x ≦ 1,386
3	527 ≦ x ≦ 723
4	240 ≦ x ≦ 384
5	103 ≦ x ≦ 209
6 以上	103 ≦ x ≦ 209

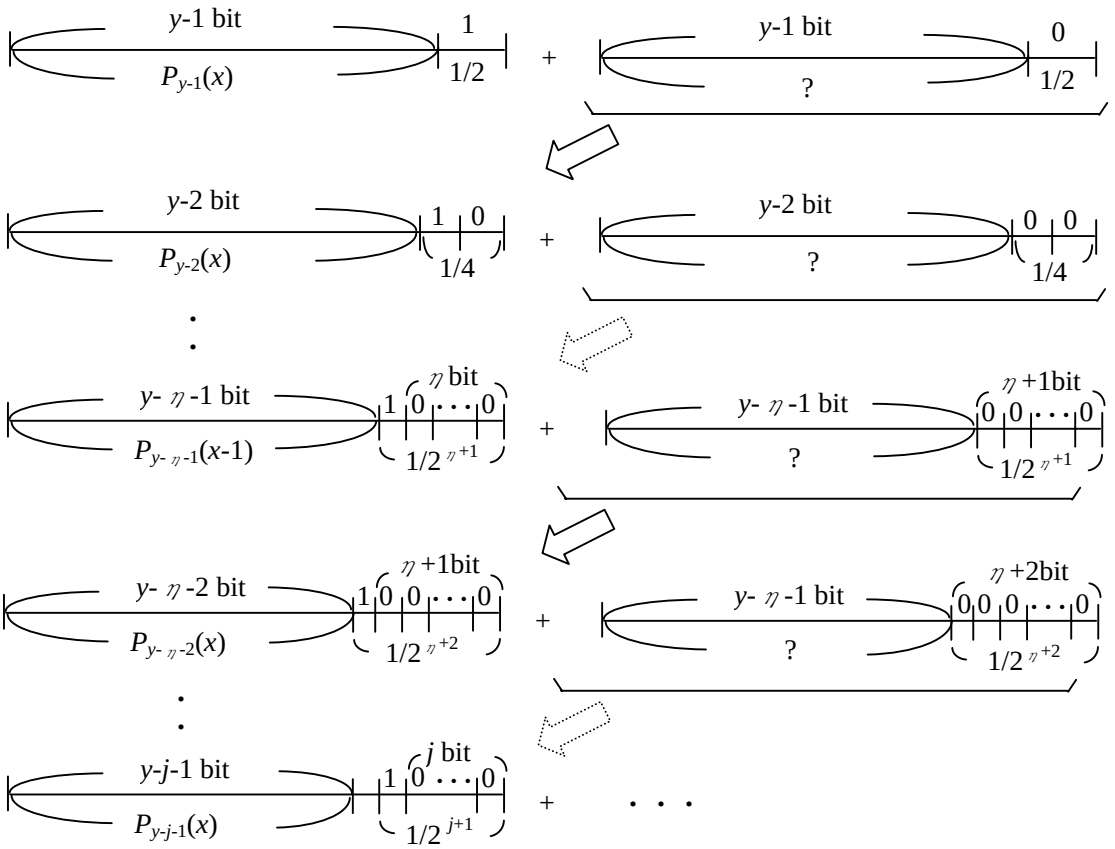
統計量の分布関数：

この検定の様に固定ビット長に現れる連の分布関数は単純でなく，二項係数などを用いた場合には計算時間がかかりすぎて現実的でない．そこで，連の各度数における確率値を漸化式により求めた．

【長さ η の連について】

ビット長 y の数列に所定の長さ η の連が x 個($x \geqq 2$)である確率を $P_y(x)$とする．また，連の数値は 0 のであるとする．

この確率 $P_y(x)$は，y ビットのうちの最右 1 ビットを 1 に固定した場合と 0 に固定した場合のたし合わせで表わすことができる．ただし，最右 1 ビットを 1 に固定した場合に連が x 個である確率は $P_{y-1}(x)/2$ と容易に書けるが，最右 1 ビットを 0 に固定した場合には解らない．理由は，右から二番目に 0 が来るときに，このビットを含む連の長さが変わってしまうためである．そこで，固定するビットを一つ増やして計算を続けることにする．図の二行目左のように，固定ビットの左端が 1 であれば計算は容易で確率は $P_{y-2}(x)/4$ となる．



この作業を固定ビットが y-1 個となるまで続け，$P_{y-1}(x)/2+ P_{y-2}(x)/4+\cdots$という具合に各々の確率を合計すれば $P_y(x)$を求めることができる．ただし，固定ビットの 0 の長さが η であるときは，固定ビットが長さ η の連としてカウントされるため，固定ビット以外では長さ η の連の個数が x-1 個でなくてはならず，確率は $P_{y-\eta-1}(x\text{-}1)/2^{\eta+1}$ となる．固定ビットの 0 の長さが η+1 からは元にもどり，$P_{y-\eta-2}$ ($\underline{x\text{-}1}$)ではなく $P_{y-\eta-2}$ ($\underline{x}$)を使った計算となる．

連の個数 x が 0 と 1 の時は上記の計算と若干異なるが，基本的な考え方は変わらないので説明を省略する．漸化式は以下の通り．

長さ η の連がゼロ個の場合

$$P_k(0)=1 \qquad (\ k=\ 1\ \sim\ \eta-1\ )$$
$$P_k(0)=1-2^{-\eta} \qquad (\ k=\ \eta\ )$$
$$P_k(0)=\sum_{i=1}^{\eta}2^{-i}P_{\eta+1-i}(0)+2^{-(\eta+1)} \qquad (\ k=\ \eta+1\ )$$
$$P_k(0)=\sum_{i=1}^{\eta}2^{-i}P_{k-i}(0)+\sum_{i=\eta+2}^{k-1}2^{-i}P_{k-i}(0)+2^{-(k-1)} \qquad (\ k\ \geqq\ \eta+2\ ) \qquad (3)$$

長さ η の連が 1 個の場合

$$P_k(1)=0 \qquad (\ k=\ 1\ \sim\ \eta-1\ )$$
$$P_k(1)=2^{-\eta} \qquad (\ k=\ \eta\ ,\ \eta+1\ )$$
$$P_k(1)=\sum_{i=1}^{\eta}2^{-i}P_{k-i}(1)+\sum_{i=\eta+2}^{k-\eta}2^{-i}P_{k-i}(1)+2^{-(\eta+1)}P_{k-\eta-1}(0) \qquad (\ k\ \geqq\ \eta+2\ ) \qquad (4)$$

長さ η の連が x 個($x\geqq 2$)の場合

$$P_k(x)=0 \qquad (\ k\ =\ 1\sim\ (\eta+1)x-2\ )$$
$$P_k(x)=2^{-((\eta+1)x-1)} \qquad (\ k\ =\ (\eta+1)x-1\ )$$
$$P_k(x)=\sum_{i=1}^{\eta}2^{-i}P_{k-i}(x)+\sum_{i=\eta+2}^{k-(\eta+1)x+1}2^{-i}P_{k-i}(x)+2^{-(\eta+1)}P_{k-\eta-1}(x-1) \qquad (\ k\ \geqq\ (\eta+1)x\ ) \qquad (5)$$

【長さ η 以上の連について】

ある長さ以上の連すべてについての出現確率も，上記と同様の考え方で導くことができる．異なる点は，固定ビットの 0 の長さが η となってから，ずっと $P(\underline{x\text{-}1})$ を使った計算となることである．

長さ η 以上の連がゼロ個の場合

$$P_k(0)=1 \qquad (\ k=\ 1\sim\ \eta-1\ )$$
$$P_k(0)=1-2^{-\eta} \qquad (\ k=\ \eta\ )$$
$$P_k(0)=\sum_{i=1}^{\eta}2^{-i}P_{k-i}(0) \qquad (\ k\ \geqq\ \eta+1\ ) \qquad (6)$$

長さ η 以上の連が 1 個の場合

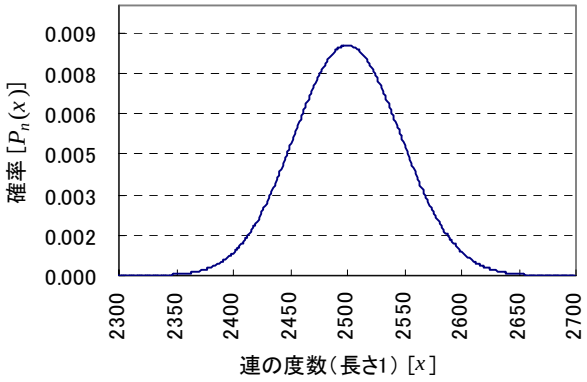
$$P_k(1)=0 \qquad (\ k\ =\ 1\ \sim\ \eta-1\ )$$
$$P_k(1)=2^{-\eta} \qquad (\ k\ =\ \eta\ )$$
$$P_k(1)=\sum_{i=1}^{\eta}2^{-i}P_{k-i}(1)+\sum_{i=\eta+1}^{k-1}2^{-i}P_{k-i}(0)+2^{k-1} \qquad (\ k\ \geqq\ \eta+1\ ) \qquad (7)$$

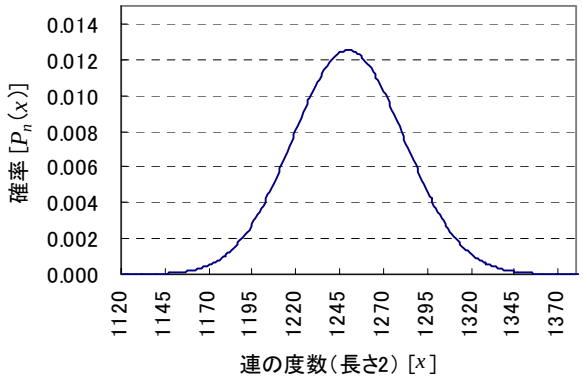
長さ η 以上の連が x 個($x\geqq 2$)の場合

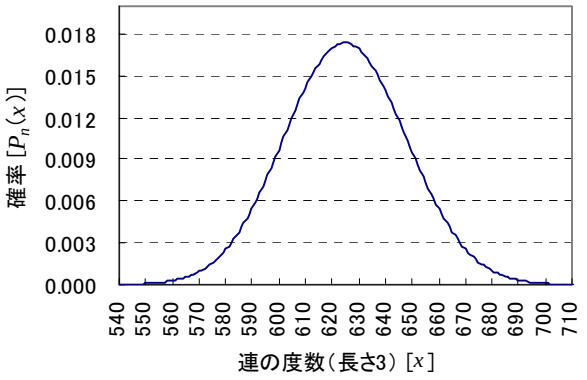
$$P_k(x)=0 \qquad (\ k\ =\ 1\sim\ (\eta+1)x-2\ )$$
$$P_k(x)=2^{-((\eta+1)x-1)} \qquad (\ k\ =\ (\eta+1)x-1\ )$$
$$P_k(x)=\sum_{i=1}^{\eta}2^{-i}P_{k-i}(x)+\sum_{i=\eta+1}^{k-(\eta+1)(x-1)+1}2^{-i}P_{k-i}(x-1) \qquad (\ k\ \geqq\ (\eta+1)x\ ) \qquad (8)$$

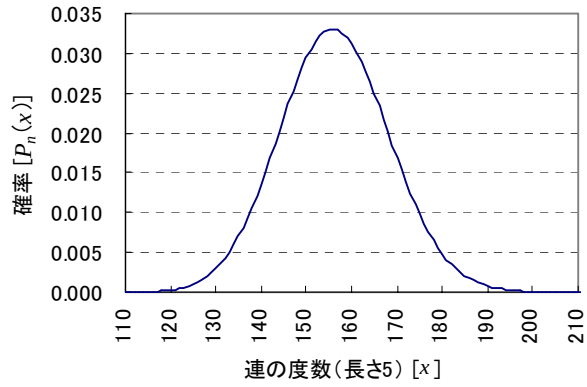
これら漸化式において，2^i の高次を省略することで計算量を減らせることができる．ただし，x=0,1 はできるだけ精度を保つことを勧める．

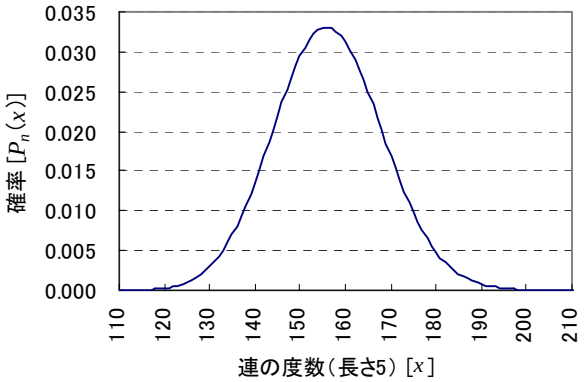
以上で求めた各連の長さ別に得られた確率 $P_n(x)$のうち，n=20000bit のものが FIPS140-2 における Runs Test の統計量の確率分布となる．

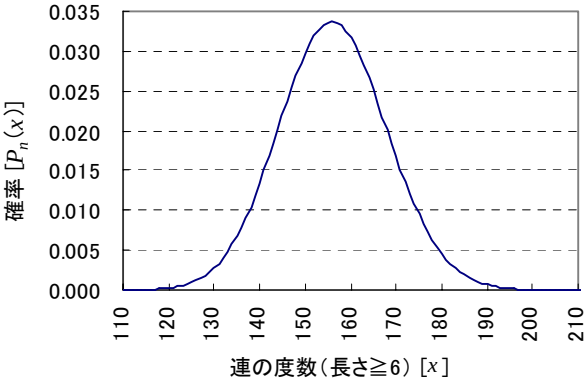












検定の棄却率 α　：　$\alpha_i = 1 - \sum_{x=\min(i)}^{\max(i)} P_n(X)$　　ただし，i は連の長さを表わし、$max(i)$と $min(i)$は連の長さ i における採択域の上限と下限であるとする。また、$n = 20000$ である。

	1 の連 α_1	2 の連 α_2	3 の連 α_3	4 の連 α_4	5 の連 α_5	6 以上の連 α_{6+}
棄却率 α_i	0.00007355	0.00001875	0.00001871	0.00001675	0.00001299	0.00000902

4. The Longruns Test

二値数列 20,000bit 中に現れる連のうち，最長の連の長さを調べる．

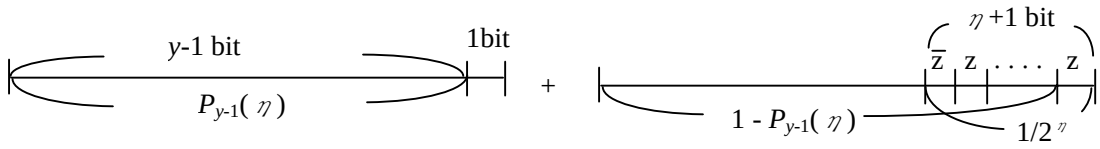
統計量 x ：20,000bit 中で最長の連の長さ（ビット値の区別無し）

採択域 ： $x < 26$

統計量の分布関数：

この検定における確率の計算は，上記 Runs Test の長さ η 以上の連がゼロ個の場合についての計算を用いることはできますが，これより効率のよい計算方法がありますので以下に示します．

ビット長 y の数列に長さ η 以上の連が一つでも現れる確率を $P_y(\eta)$ とする．これは，右端 1bit 減らしたビット長 y-1 の数列に長さ η 以上連が存在する確率と，ビット長 y-1 の数列には長さ η 以上の連が存在せず，かつ右端 1bit を含めて長さ η の連が存在する確率との合計となる．ここで，右端 1bit を含めた場合に長さ η 以上の連としなかったのは，ビット長 y-1 の数列に長さ η 以上の連が存在しないことと矛盾することによる．



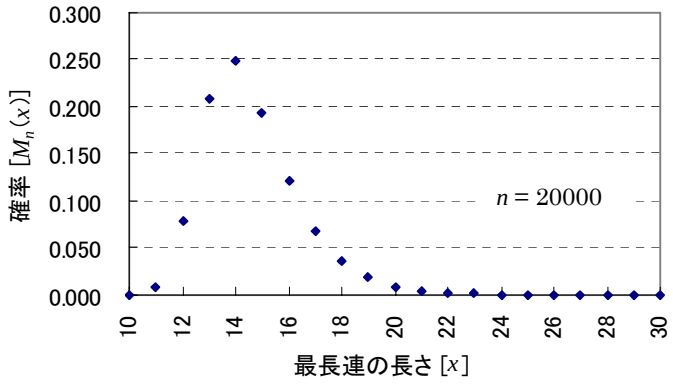
ビット長 y-1 の数列に長さ η 以上の連が存在する確率は $P_{y-1}(\eta)$，存在しない確率は $1 - P_{y-1}(\eta)$ であり，右端 1bit を含めて長さ η の連が存在する確率はビット値の 1 と 0 を区別しないことを考慮して $1/2^{\eta}$ である．

以上をまとめると，次の漸化式となる．

$$
\begin{aligned}
&P_k(\eta) = 0 && (\ k = 1 \sim \eta - 1\ ) \\
&P_k(\eta) = 2^{-(\eta-1)} && (\ k = \eta\ ) \\
&P_k(\eta) = P_{k-1}(\eta) + 2^{-\eta}(1 - P_{k-1}(\eta)) && (\ k \geqq \eta + 1\ )
\end{aligned}
\tag{9}
$$

n bit 中に現れる連のうち，最長の連の長さが η である確率は次式で与えられる

$$M_n(\eta) = P_n(\eta) - P_n(\eta + 1) \tag{10}$$



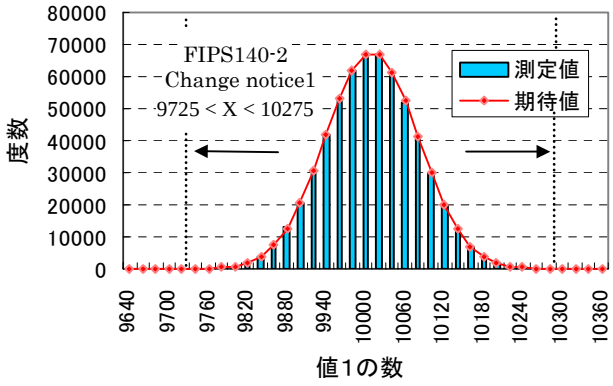
検定の棄却率 α ： $\alpha = P_n \cong 0.000298$ ただし，$n = 20000$, $\eta = 26$.

5 ＲＰＧ１００乱数による検定統計量の分布

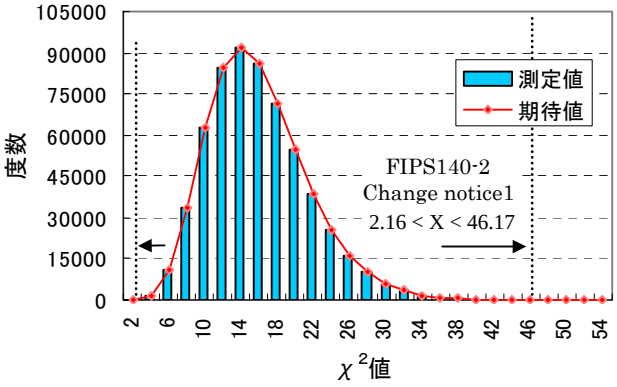
ＦＩＰＳ乱数検定を 60,0000 回行って得られた各種検定統計量の測定データと，前述の計算によって得られた理論値を示す．

（乱数取得条件）　サンプル数１　　電圧３．３Ｖ　　温度２５℃　　乱数生成クロック２５０ＫＨｚ

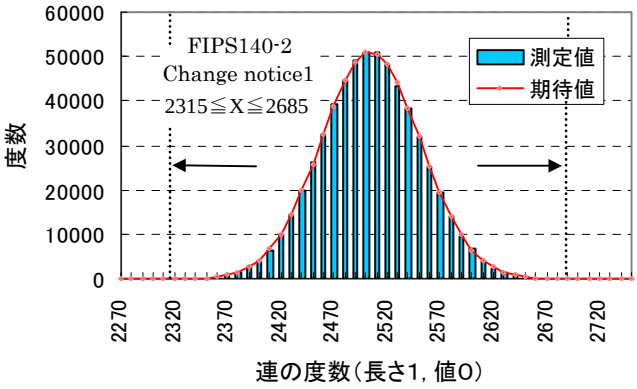
The monobit test

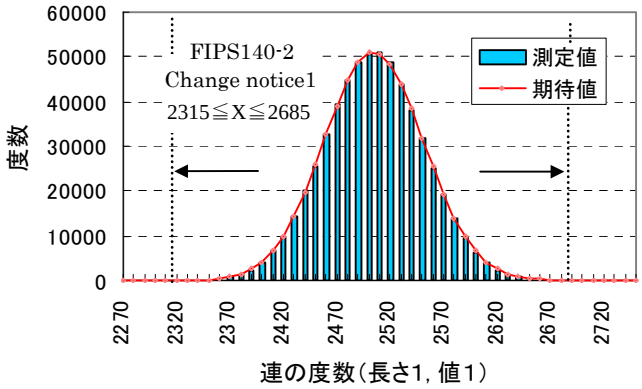


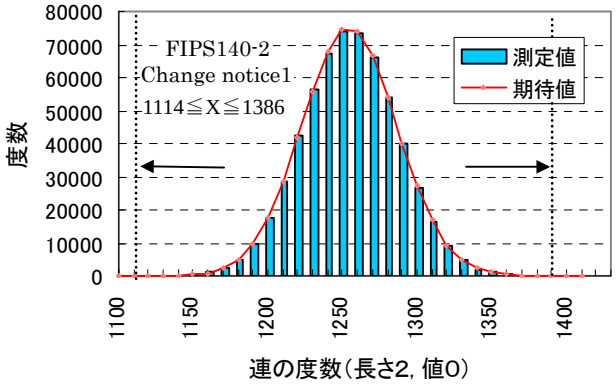
The poker test

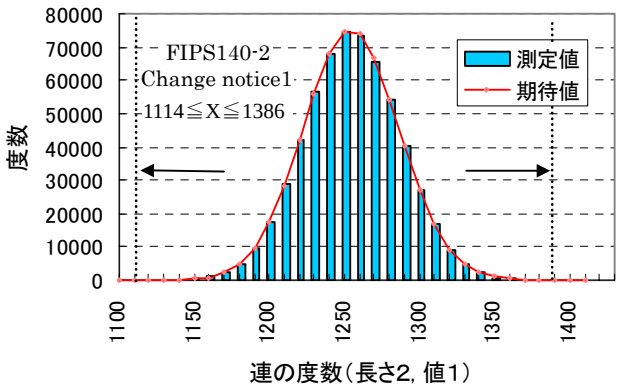


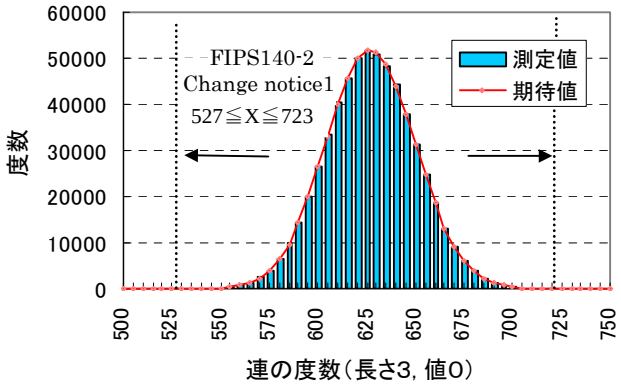
The runs test

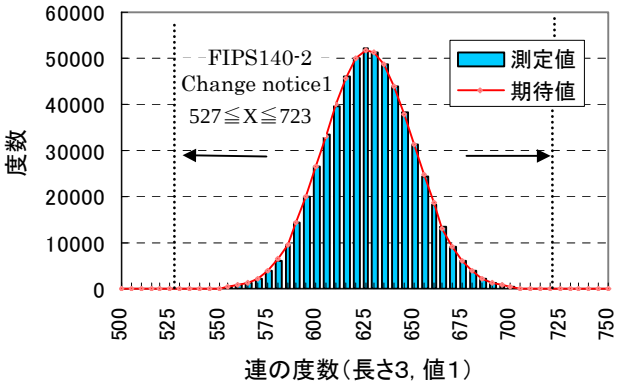


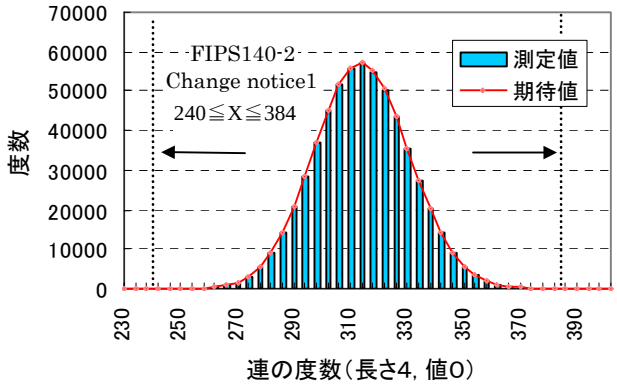


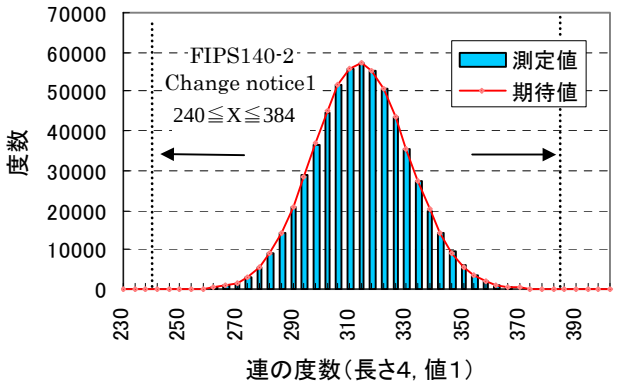


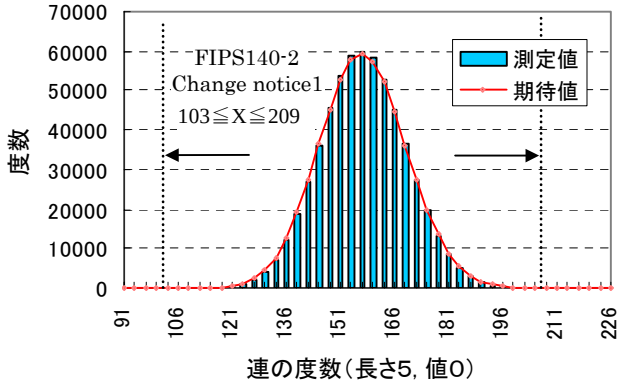


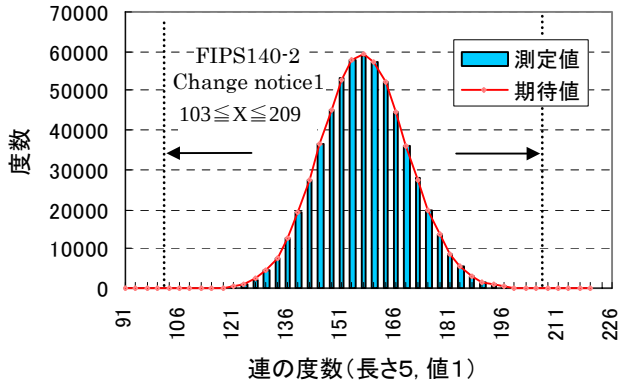


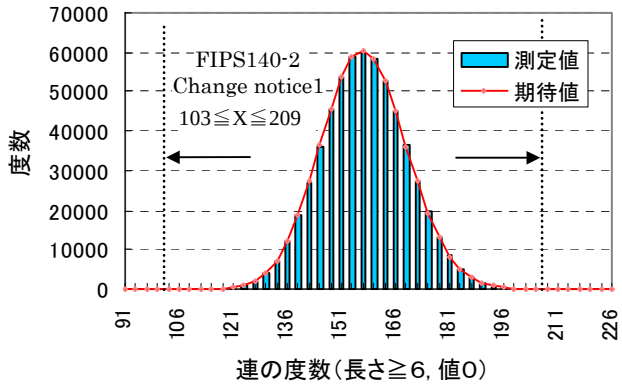


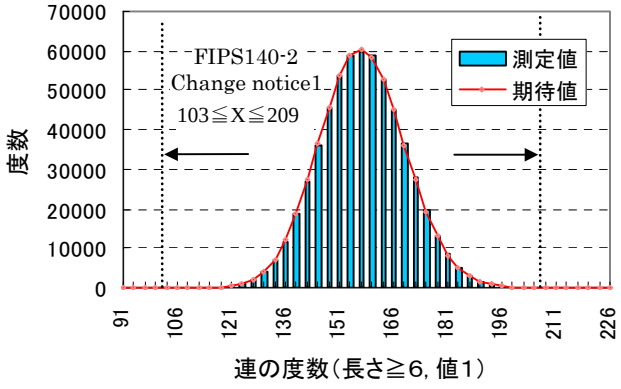




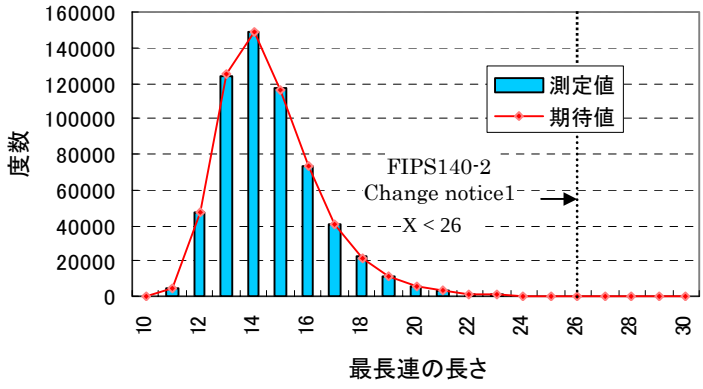








The longruns test



6 ＦＩＰＳ乱数検定不合格数の分布

乱数検定を 4,000 回行い, その中に現れた不合格数の分布を示す。検定 4,000 回を１セットとし、これを 1500 セット行った。

（乱数取得条件）　サンプル数１　　電圧３．３Ｖ　　温度２５℃　　乱数生成クロック２５０ＫＨｚ