

連による 0,1 乱数列の無作為性の考察

清水 道夫*

(昭和51年12月16日受理)

A Study of Randomness of 0,1 Random Number by Runs

Michio SHIMIZU

Several tests of randomness for two kinds of elements' arrangement have already been proposed, and these distributions are provided on the assumption that hypothesis of randomness should be true from the point of view of controlling the type I error. This paper is concerned with the test of randomness of 0,1 random number sequence by the total number of runs. The model of nonrandom 0,1 random number is set up by using a binomial-beta distribution, and the distribution of total number of runs for this model is established. And then, the power of the randomness test is considered for the type II error.

1 ま え が き

0,1 乱数は1の生起確率が p であるベルヌーイ試行の結果として現われる二種要素の数列である。

0,1 乱数としての条件は主に等確率性(等出現性)と無作為性(独立性)が保証されることであるが、独立性を検定する方法として連の検定がある。

連の理論に基づいて0,1乱数の無作為性の仮説を検定する方法として次のものがある。

[1]

- 1) 要素1(又は0)の連の数による方法
- 2) 連の総数による方法
- 3) 要素1(又は0)の連の長さの最大のものによる方法
- 4) 連の長さの最大のものによる方法
- 5) 連の種々の長さの数による方法

以上の検定方法に用いられている公布は、何れも第一種の誤りを制御するという見地から、無作為性の仮説が真であるという仮説の下に定められたものである。

本論では、0,1乱数列から引き続き長さ n のランダムサンプリングを k 回行なったものに対する2)のサンプル内の連の総数による検定方法を扱う。2章で既に求められている

* 情報工学教室 助手

連の総数の分布と期待値を示す。

第二種の誤りについて考察するために、3章で非独立な0,1乱数のモデルを要素1のベータ二項分布で表現し、4章でそのモデルに対する連の総数の分布を求める。さらに5章で検定方法と検出力について述べる。

2 連の総数の分布

引き続き長さ n の0,1数列 $b_1 b_2 \cdots b_n$ に対し、要素1の総数を x 、 $(x = \sum_{i=1}^n b_i)$ とし、連の総数 r を次のように表現する。

$$r = \sum_{i=1}^{n-1} (b_i \oplus b_{i+1}) + 1$$

Wishert と Hirschfeld [2] は、二項母集団からのサンプルにおける連の総数 r の期待値を、二要素の起こる確率をそれぞれ p, q 、($q=1-p$) として、次のように導いた。

$$E(r) = 2pq(n-1) + 1 \quad (1)$$

ただし、無作為性が保証されている場合である。

特に $p = q = 1/2$ のとき、 $E(r) = (n+1)/2$ で等出現性が保証された場合である。

A. Wald と J. Wolfowitz [3] は要素1の総数 x と要素0の総数 $n-x$ に対する連の総数の分布を次のように得ている。

$r : \text{even}$

$$p_r(r=2s) = 2 \binom{x-1}{s-1} \binom{n-x-1}{s-1} / \binom{n}{x} \quad (2)$$

$r : \text{odd}$

$$p_r(r=2s+1) = \left\{ \binom{x-1}{s} \binom{n-x-1}{s} + \binom{x-1}{s-1} \binom{n-x-1}{s-1} \right\} / \binom{n}{x}$$

また、 x に対する連の期待値 $E(r|x)$ と分数 $\sigma^2(r|x)$ をそれぞれ次のように求めた。

$$E(r|x) = \frac{2x(n-x)}{n} + 1 \quad (3)$$

$$\sigma^2(r|x) = \frac{2\{2x^4 - 4nx^3 + (2n^2 + n)x^2 - n^2x\}}{n^2(n-1)} \quad (4)$$

3 非独立0,1乱数の表現

非独立な0,1乱数は、要素1の全体の母比率 p' を中心に部分的な母集団の母比率がバラツいたものと解釈する。

ベータ分布の確率密度関数を次式で示す。[4]

$$f(p) = \frac{1}{B(ap', aq')} p^{ap'-1} (1-p)^{aq'-1} \quad (5)$$

$p', q' = (1-p')$ をそれぞれ要素 1, 要素 0 の母比率とする. パラメータ $a \rightarrow 0, a \rightarrow \infty$ に対するベータ分布をそれぞれ Fig. 1, (A), (B)に示す.

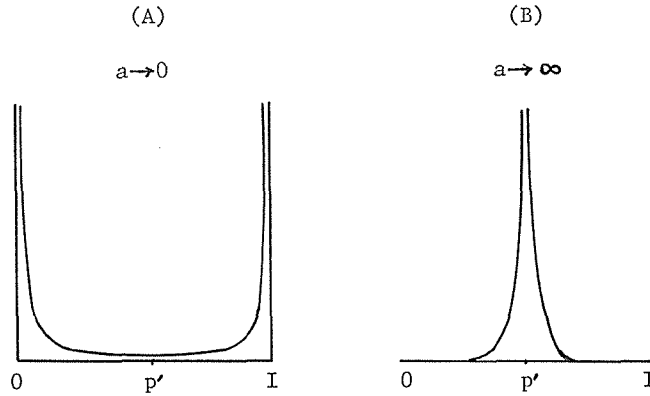


Fig. 1 Beta Distribution for $a \rightarrow 0, a \rightarrow \infty$

ベータ分布を非独立な 0,1 乱数の部分的な母比率の分布とすれば, 引き続き長さ n のランダムサンプリングした x の分布は, Fig. 1, (A)の場合ほとんど 0 か n に分離し, (Fig. 2, (A')). Fig. 1, (B)の場合は二項分布 $B(n, p')$ に近い分布となる. (Fig. 2, (B'))

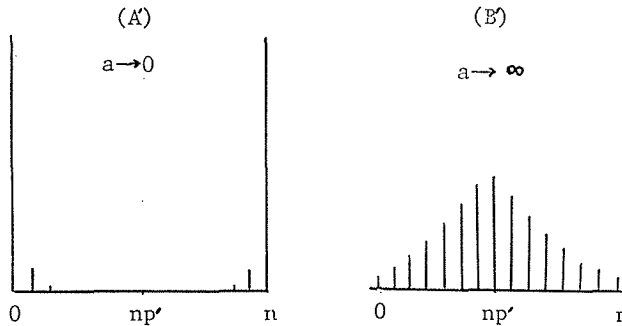


Fig. 2 Binomial-Beta Distribution for $a \rightarrow 0, a \rightarrow \infty$

ところで同一母比率の部分母集団からのサンプルに対する x の分布は二項分布に従うから, ベータ分布と二項分布の複合分布 (ベータ・二項分布) は, 母比率に部分的なバラツキのある非独立な 0,1 乱数列からのサンプルに対する x の分布と考えられる.

二項分布は次式

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

で表わされるから(5)式との複合分布（ベータ・二項分布）は次式で示される。

$$f(x) = \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \frac{p^{ap'-1}(1-p)^{aq'-1}}{B(ap', aq')} dp \quad (6)$$

ベータ・二項分布に従う x の確率変数 \mathbf{x} の期待値 $E^*(\mathbf{x})$ と分散 $V^*(\mathbf{x})$ を求めるとそれぞれ(7), (8)式のようになる。以後ベータ・二項分布 ($BB(n, p', a)$ で表わす) の場合は*を付ける。

$$E^*(\mathbf{x}) = np' \quad (7)$$

$$V^*(\mathbf{x}) = np'q'(n+a)/(a+1) \quad (8)$$

パラメータ $a \rightarrow \infty$, $a \rightarrow 0$ に対する分散 $V^*(\mathbf{x})$ は、それぞれ次のようになる。

$$V_\infty(\mathbf{x}) = \lim_{a \rightarrow \infty} V^*(\mathbf{x}) = np'q' \quad (9)$$

$$V_0(\mathbf{x}) = \lim_{a \rightarrow 0} V^*(\mathbf{x}) = n^2p'q' \quad (10)$$

(9)式は二項分布の分散そのものであり、(10)式は 0 と n に完全分離した分布の分散である。また、(9), (10)式を用いて(8)式の分散は

$$V^*(\mathbf{x}) = \frac{1}{a+1} V_0(\mathbf{x}) + \frac{a}{a+1} V_\infty(\mathbf{x}) \quad (11)$$

のように表わすことができる。したがって

$$M = 1/(a+1) \quad (12)$$

を全体の母比率が p' である非独立 0,1 乱数の母比率のバラツキを表わす尺度とする。パラメータ a ($0 \leq a \leq \infty$) に対して $1 \geq M \geq 0$ である。

尺度 M に対するベータ・二項分布の分散は、 $np'q' \leq V^*(\mathbf{x}) \leq n^2p'q'$ で単調に増加をするから、ベータ・二項分布で表現した非独立性のモデルは、完全独立である場合の二項分布からの偏りを連続的に表現することができる。

4 非独立 0,1 乱数の連の総数の分布

引き続き長さ n のサンプルに対する要素 1 の総数 x の分布がベータ・二項分布をするとき、連の総数 r をとる確率変数 \mathbf{r} の期待値 $E^*(\mathbf{r})$ と分散 $V^*(\mathbf{r})$ を求める。

独立な 0,1 乱数において x の分布 $p_r(\mathbf{x} = x)$ は二項分布で示されるが、非独立な 0,1 乱数として $p_r(\mathbf{x} = x)$ がベータ・二項分布をするのであるから、これを $p_r^*(\mathbf{x} = x)$ とする。

連の総数 r をとる確率を $p_r^*(\mathbf{r} = r)$ と、おくと確率変数 \mathbf{r} の期待値 $E^*(\mathbf{r})$ と分散 $V^*(\mathbf{r})$ はそれぞれ次の定義式で示される。

$$E^*(\mathbf{r}) = \sum_{r=1}^n r p_{r^*}(\mathbf{r} = r)$$

$$V^*(\mathbf{r}) = \sum_{r=1}^n (r - E^*(\mathbf{r}))^2 p_{r^*}(\mathbf{r} = r)$$

これを具体的に計算するとそれぞれ(13), (14)式のようにになる。(付録参照)

$$E^*(\mathbf{r}) = \frac{2ap'q'}{a+1}(n-1) + 1 \quad (13)$$

$$V^*(\mathbf{r}) = \frac{2}{n(n-1)} \{2E^*(\mathbf{x}^4) - 4nE^*(\mathbf{x}^2) + (2n^2 - 2n + 3)E^*(\mathbf{x}^2) + (2n^2 - 3n)E^*(\mathbf{x})\} + 1 - \{E^*(\mathbf{r})\}^2 \quad (14)$$

($E^*(\mathbf{x}^i)$ はベータ・二項分布の i 次モーメント)

パラメータ $a \rightarrow \infty$, $a \rightarrow 0$ に対する \mathbf{r} の期待値 $E^*(\mathbf{r})$ はそれぞれ次のようになる。

$$E_{\infty}(\mathbf{r}) = \lim_{a \rightarrow \infty} E^*(\mathbf{r}) = 2p'q'(n-1) + 1 \quad (15)$$

$$E_0(\mathbf{r}) = \lim_{a \rightarrow 0} E^*(\mathbf{r}) = 1 \quad (16)$$

(15)式は無作為性が保証された(1)式にほかならない。

(8)式と(13)式を a について微分するとそれぞれ

$$\frac{dV^*(\mathbf{x})}{da} = \frac{np'q'(1-n)}{(a+1)^2} \leq 0 \quad (\text{単調減少})$$

$$\frac{dE^*(\mathbf{r})}{da} = \frac{2p'q'(n-1)}{(a+1)^2} \geq 0 \quad (\text{単調増加})$$

となり、次の比例関係があることがわかる。

$$\frac{dV^*(\mathbf{x})}{da} = \left(-\frac{n}{2}\right) \frac{dE^*(\mathbf{r})}{da}$$

この比例定数 $h = (-n/2)$ を用いて次式が導びかれる。

$$V^*(\mathbf{x}) = h \{E^*(\mathbf{r}) - 1\} + V_0(\mathbf{x}) \quad (17)$$

5 無作為性の検定と検出力

非独立な 0, 1 乱数として x の分布がベータ・二項分布をするとき、(17)式で示される一次関係から任意のパラメータ a に対する確率変数 \mathbf{x} の分散 $V^*(\mathbf{x})$ は、確率変数 \mathbf{r} の期待値 $E^*(\mathbf{r})$ に比例定数 h と定数項で変換できることがわかった。ゆえに \mathbf{x} の分散の偏りで示さ

れる非独立性に対し、連の総数 r の平均値に関する独立性の検定を行なえばよいことがいえる。

Wishert と Hirschfeld [2]は、連の総数 r の期待値を $2(n-1)p'q' + 1$ 、分散の近似値を $2\sqrt{(n-1)p'q'(1-3p'q')}$ として n が大きいとき

$$u_0 = \frac{r - \{2(n-1)p'q' + 1\}}{2\sqrt{(n-1)p'q'(1-3p'q')}}$$

が正規分布に近似することを示した。

したがってサンプルの長さ n が比較的大きく、サンプル数を k とすると $\bar{r} = \Sigma r/k$ として

$$u_0 = \frac{\bar{r} - \{2(n-1)p'q' + 1\}}{2\sqrt{(n-1)p'q'(1-3p'q')/k}} \quad (13)$$

は正規分布近似をするとみなしうるから正規分布による検定方法を用いる。

r の母平均を μ とすると、完全独立 0,1 乱数では $\mu = 2(n-1)p'q' + 1$ であるが、 x の分布がベータ・二項分布をする非独立 0,1 乱数では、(13)式より $1 \leq \mu \leq 2(n-1)p'q' + 1$ であるから片側検定とする。実際に検定する場合母比率 p' は推定量 \hat{p}' を用いなければならぬが、ここでは p' がわかっているものと仮定する。

0,1 乱数列の無作為性の検定を次のように設定する。有意水準を α とする。

仮説： $H_0: \mu = \mu_0, \mu_0 = 2(n-1)p'q' + 1$

$H_1: \mu = \mu_1, 1 \leq \mu_1 < 2(n-1)p'q' + 1$

統計量： $u_0 = \frac{\bar{r} - \{2(n-1)p'q' + 1\}}{\sqrt{(n-1)p'q'(1-3p'q')/k}} \quad \left(\bar{r} = \frac{\Sigma r}{k}\right)$

棄却域 $R: u_0 \leq -u(2\alpha)$

検出力 $1-\beta$ は次式で示される。

$$1 - \beta = p_r \{u_0 \leq -u(2\alpha)\}$$

ここで r の母標準偏差を σ 、($\sigma = \sqrt{V^*(r)}$) として

$$u_0 = \frac{\bar{r} - \mu}{\sigma/\sqrt{k}} \cdot \frac{\sigma}{2\sqrt{(n-1)p'q'(1-3p'q')}} + \frac{\mu - \{2(n-1)p'q' + 1\}}{2\sqrt{(n-1)p'q'(1-3p'q')/k}}$$

と変形し、さらに

$$u = \frac{\bar{r} - \mu}{\sigma/\sqrt{k}}$$

とおくと検出力 $1-\beta$ は次式で示される。

$$1 - \beta = p_r \left\{ u \leq \left(-u(2\alpha) - \frac{\mu - \{2(n-1)p'q' + 1\}}{2\sqrt{(n-1)p'q'(1-3p'q')/k}} \right) \right\}$$

$$\frac{2\sqrt{(n-1)p'q'(1-3p'q')}}{\sigma} \tag{19}$$

したがって母標準偏差 σ が統計量 $\hat{\sigma}$ で近似される場合には、検出力が導出される。厳密には $\hat{\sigma}$ を用いる必要がある。

例として要素 1 の母比率 $p' = 1/2$ が保証されているとき、サンプルの長さを $n = 10$ として、尺度 M に対する検出力を求め Fig. 3 に示す。サンプル数 k は $2^i, (i = 1 \sim 5)$ とする、理論の妥当性を検証するためコンピュータシミュレーションを行ない、 $M = 0.1$ (0.1) 0.9 に対し離散値で示す。

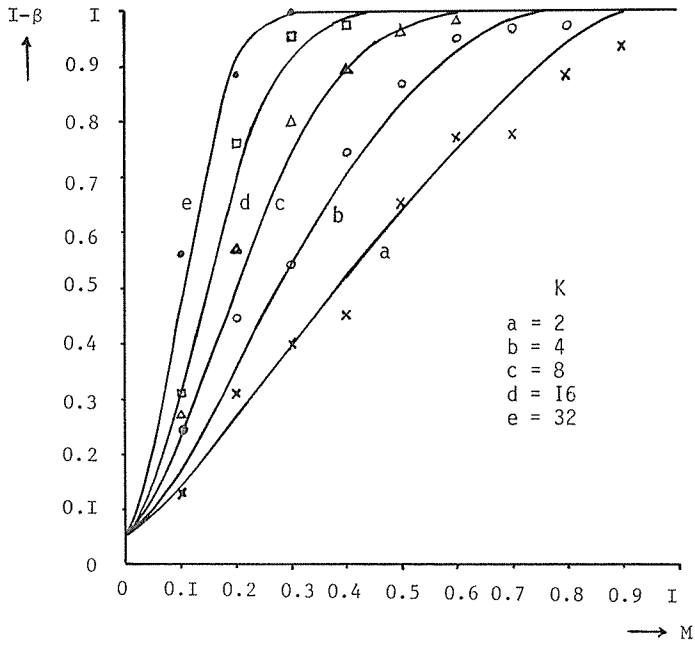


Fig. 3 Power Curve ($\alpha = 0.05$)

逆に $\Delta\mu = \mu_0 - \mu$ がある値であるとき、これに対する所望の検出力を確保するためには、サンプルの長さ n とサンプル数 k をいくらにとればよいかということを検討する必要がある。サンプルの長さ n が決まっているとき、必要なサンプル数 k は次式で求められる。

$$\left(\frac{\sigma \cdot u(2\beta) + \sigma_0 \cdot u(2\alpha)}{\mu_0 - \mu}\right)^2 \leq k \tag{20}$$

ただし母平均 μ と母標準偏差 σ を既知とする。

6 あとがき

ベータ・二項分布の導入によって非独立な 0,1 乱数列を表現する方法は、非独立な 0,1

乱数の 1 モデルにすぎないが、理想的なランダムネスからの偏りを連続的に表現できる上で適した方法であると考えられる。

非独立な 0, 1 乱数を具体的に設定することによって、検出力が求められることを示したことは、乱数検定研究の新たな方法論を提起したといえよう。

今後の研究として次の如きものがある。

- 1 多種の要素の場合への拡張
- 2 連の他の分布に対する非独立性の設定
- 3 マルコフ連鎖による非独立性の表現
- 4 他の検定方法との検出力の比較

最後に本論文作成にあたり、御指導、御示唆を下された山梨大学工学部 赤尾洋二教授、吉沢正助教授、新藤久和助手に謝意を表したい。

付 録

x がベータ・二項分布をするときの \mathbf{r} の期待値 $E^*(\mathbf{r})$ と分散 $V^*(\mathbf{r})$ の導出

$$\begin{aligned}
 1 \quad E^*(\mathbf{r}) &= \sum_{r=1}^n r \cdot \dot{p}_r^*(\mathbf{r} = r) \\
 &= \sum_{x=0}^n \sum_{r=1}^n r \cdot \dot{p}_r^*(\mathbf{x} = x, \mathbf{r} = r) \\
 &= \sum_{x=0}^n \sum_{r=1}^n r \cdot \dot{p}_r^*(\mathbf{x} = x) \dot{p}_r(\mathbf{r} = r \mid \mathbf{x} = x) \\
 &= \sum_{x=0}^n \dot{p}_r^*(\mathbf{x} = x) \sum_{r=1}^n r \cdot \dot{p}_r(\mathbf{r} = r \mid \mathbf{x} = x)
 \end{aligned}$$

ここで $\sum_{r=1}^n r \cdot \dot{p}_r(\mathbf{r} = r \mid \mathbf{x} = x)$ は(3)式の x に対する連の期待値 $E(r \mid x)$ にほかならない。

ゆえに

$$\begin{aligned}
 E^*(\mathbf{r}) &= \sum_{x=0}^n \dot{p}_r^*(\mathbf{x} = x) \left\{ \frac{2x(n-x)}{n} + 1 \right\} \\
 &= 2 \sum_{x=0}^n x \cdot \dot{p}_r^*(\mathbf{x} = x) - \frac{2}{n} \sum_{x=0}^n x^2 \dot{p}_r^*(\mathbf{x} = x) + 1 \\
 &= 2E^*(\mathbf{x}) - \frac{2}{n} E^*(\mathbf{x}^2) + 1 \\
 &= 2np' - \frac{2}{n} \left\{ n(n-1) \frac{(ap' + 1)p'}{a+1} + np' \right\} + 1 \\
 &= \frac{2ap'q'(n-1)}{a+1} + 1
 \end{aligned}$$

$$\begin{aligned}
2 \quad V^*(\mathbf{r}) &= \sum_{r=1}^n \{r - E^*(\mathbf{r})\}^2 p_{r^*}(\mathbf{r} = r) \\
&= \sum_{r=1}^n r^2 p_{r^*}(\mathbf{r} = r) - \{E^*(\mathbf{r})\}^2 \\
&= \sum_{x=0}^n p_{r^*}(\mathbf{x} = x) \sum_{r=1}^n r^2 p_r(\mathbf{r} = r \mid \mathbf{x} = x) - \{E^*(\mathbf{r})\}^2
\end{aligned}$$

ここで

$$\sigma^2(r \mid x) = \sum_{r=1}^n r^2 p_r(\mathbf{r} = r \mid \mathbf{x} = x) - \{E(r \mid x)\}^2$$

から

$$\sum_{r=1}^n r^2 p_r(\mathbf{r} = r \mid \mathbf{x} = x) = \sigma^2(r \mid x) + \{E(r \mid x)\}^2$$

である。

したがって(3), (4)式を代入して

$$\begin{aligned}
V^*(\mathbf{r}) &= \sum_{x=0}^n p_{r^*}(\mathbf{x} = x) \left[\frac{2\{2x^4 - 4nx^3 + \{(2n^2 + n)x^2 - n^2x\}}{n^2(n-1)} \right. \\
&\quad \left. + \left\{ \frac{2x(n-x)}{n} + 1 \right\}^2 \right] - \{E^*(\mathbf{r})\}^2 \\
&= \sum_{x=0}^n p_{r^*}(\mathbf{x} = x) \left[\frac{2}{(n-1)n} \{2x^4 - 4nx^3 + (2n^2 - 2n + 3)x^2 \right. \\
&\quad \left. + (2n^2 - 3n)x + 1 \right] - \{E^*(\mathbf{r})\}^2 \\
&= \frac{2}{n(n-1)} \{2E^*(\mathbf{x}^4) - 4nE^*(\mathbf{x}^3) + (2n^2 - 2n + 3)E^*(\mathbf{x}^2) \\
&\quad + (2n^2 - 3n)E^*(\mathbf{x}) + 1\} - \{E^*(\mathbf{r})\}^2
\end{aligned}$$

ベータ・二項分布のモーメント

$$E^*(\mathbf{x}) = np'$$

$$E^*(\mathbf{x}^2) = n(n-1) \frac{(ap' + 1)p'}{a+1} + E^*(\mathbf{x})$$

$$\begin{aligned}
E^*(\mathbf{x}^3) &= n(n-1)(n-2) \frac{(ap' + 2)(ap' + 1)p'}{(a+2)(a+1)} + 3E^*(\mathbf{x}^2) \\
&\quad - 2E^*(\mathbf{x})
\end{aligned}$$

$$E^*(\mathbf{x}^4) = n(n-1)(n-2)(n-3) \frac{(ap' + 3)(ap' + 2)(ap' + 1)p'}{(a+3)(a+2)(a+1)}$$

$$+6E^*(x^3) - 11E^*(x^2) + 6E^*(x)$$

参 考 文 献

- [1] ウィルクス/小河原正己訳, “数理統計学”, p328, 育文社 (1967).
- [2] J. Wishert, H. O. Hirschfeld, “A Theorem Concerning the Distribution of Joins Between Line Segments”, Journal of the London Math. Soc., VoL. XI, p227 (1936).
- [3] A. Wald, J. Wolfowitz, “On a Test of Whether Two Samples are from the Same Population”, Annals of Math. Stat., VoL. XI, p147 (1940).
- [4] Norman L. Johnson, Samuel Kotz, “Continuous Univariate Distributions-2,” p34, Houghton Mifflin Company (1970).
- [5] A. M. Mood, “The Distribution Theory of Runs”, Annals of Math. Stat., VoL. XI, p367 (1940).
- [6] Fieda S. Swed, C. Eisenhart, “Tables for Testing Randomness of Grouping in a Sequence of Alternatives”, Annals of Math. Stat., VoL. XIV, p66 (1943).
- [7] S.S. Wilks, “Mathematical Statistics”, p144, p452 John Wiley & Sons, INC., U. S. A. (1962).
- [8] 清水, 新藤, 赤尾, “接触数による乱数検定”, 情報処理学会第16回大会, 243 (1975).