論 文

ランダムグラフの統計解析*

高見沢 一 彦** 滝 内 政 昭***
西 関 隆 夫** 斎 藤 伸 自**

Abstract

It is required to generate a great many random graph data by a computer when we try to test the efficiency of some algorithms for a graph or network problem.

In this paper, we give two kind of computer algorithms to generate efficiently random (connected) graphs, and also show some statistic properties of random graphs by means of computer simulation. We find a remarkable structural feature of connected random graphs such that the most of connected random graphs generated consist of a large biconnected component and certain number of biconnected components having only one edge. And it follows that the divide-and-conquer method is not so effective when applied for a random graph. This fact is demonstrated for some numerical examples. Furthermore, we show that the directed random graphs also have the similar feature when they are decomposed into strongly connected components.

1. まえがき

近年，グラフ構造を有する問題を計算機を用いて解くことが多く，そのための効率の良いグラフ解析アルゴリズムの開発が盛んである．それと共に，問題の規模が大きくなるにつれ，計算時間が指数関数的に増大するような解法しか知られていなかった組み合せ問題（彩色数問題，最大カット問題など，NP-完全問題といわれているもの）に対しては，実用的計算時間内で近似最適解を求めるヒューリスティク算法が考案されている[1),2)]．ひとつの問題について考案されたいくつかの算法の有効性の評価ならびに比較をするには，いくつかの例について実際に計算を行う計算機実験に頼らざるを得ない．このようなときに，ランダムなグラフデータ“ランダムグラフ”が必要となる．また，グラフ問題の計算の複雑度は，従来，その最悪値で評価されてきた．しかし，実用的な観点からは，期待値評価の方が望ましい場合も多い．このような，計算の複雑度の期待値評価を行う上でも，ランダムグラフを明確に定義し，その性質を明らかにすることが望まれる．

ランダムグラフは，いくつかの定義が与えられているが，その性質については知られていないことが多く，実際に計算機を用いて，ランダムグラフを発生させ，ランダムグラフ自身の統計的な解析を行った例もないようである．本文では，文献3）に従い，節点数 n，枝数 m のランダムグラフは，可能な $C(n, 2)=n(n-1)/2$ 本の枝の中から，m 本の枝をランダムに選んで得られるグラフであると定義する．本文は，このようなランダムグラフに関する統計的数値解析を試みたものである．なお，ランダムグラフの発生ならびにグラフ処理には，筆者らが開発したグラフ処理プログラム GRAMP[4)] を用いている．

まず，ランダムグラフにおける連結グラフの発生率を測定し，既に示されている理論的な近似値との比較を行う．次に，グラフ問題を計算機で解く場合，与えられたグラフを連結成分あるいは非可分成分に分割して，各部分グラフごとの問題に帰着させて解くという

* Statistical Analysis of Random Graphs by Kazuhiko TAKAMIZAWA, Takao NISHIZEKI, Nobuji SAITO (Faculty of Engineering, Tohoku University), and Masaaki TAKIUCHI (Fujitsu, Ltd.).
** 東北大学工学部通信工学科
*** 富士通(株)

手法，すなわち divide-and-conquer 法がよく用いられるが，ランダムグラフでは，この手法により，どの程度計算時間が改善されるかについて解析を行う．また，この実験結果から，ランダムグラフの連結成分，非可分成分には，ある顕著な性質があることを示す．更に，有向ランダムグラフの強連結成分についても，同様な性質があることを示す[5]．

2. 準　備

本文においては，グラフ G を節点の有限集合 V と枝の有限集合 E の対 $G=(V,E)$ で表わす．E の各元が，2つの節点 u, $v\in V$ の順序対あるいは非順序対であるとき，Gをそれぞれ有向グラフあるいは無向グラフとよぶ．ここで，Gは自己閉路および，Gが有向グラフの場合，同じ向きの並列枝を持たないものと仮定する．

任意の2点間に道が存在する無向グラフを連結グラフとよぶ．極大な連結部分グラフを連結成分という．連結グラフが非可分であるというのは，どの3点 v, w, a をとっても，v とwの間にaを通らない道が存在することである．連結グラフは，極大な非可分部分グラフに一意に分割でき，それらを非可分成分とよぶ．有向グラフの各枝の向きを無視して得られる無向グラフ（すなわち台グラフ）が連結であるとき，元の有向グラフは弱連結であるという．また，任意の点から任意の点へ行く有向道が存在する有向グラフを強連結グラフとよぶ．弱連結グラフの極大な強連結部分グラフを強連結成分という．

3. ランダムグラフの定義と既知の結果

ランダムグラフの定義として，いくつか異なった定義が与えられているが，本文では次の定義を用いることにする．

〔定義 1〕[3]　ラベルの付けられた n 個の節点のいずれか2点を結ぶ総数 $C(n, 2)=n(n-1)/2$ 本の枝から，m本の枝をランダムに選んで得られるグラフを，n 個の節点及びm本の枝を持つランダムグラフとよぶ．但し $C(c(n,2), m)$ 種類の選択を等確率 $1/C(c(n,2),m)$ で行うものとする．（定義終）

有向ランダムグラフは，逆方向の枝も含めた $2\times C(n,2)$ 本の枝から，ランダムに m 本の枝を選んで得られる有向グラフであると定義する．

ここで定義されたランダムグラフ及び有向ランダムグラフは，発生可能なラベル付きグラフをすべて等確率で発生させるものであり，ラベル無しグラフとして同形なものも区別して考えている．従って，ラベル無しグラフとして考えた場合，各グラフは，各々異なった確率で出現することになる．

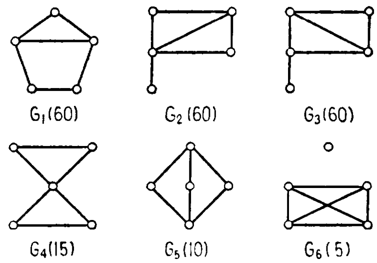

Fig. 1 Undirected labeled graphs with 5 vertices and 6 edges.

〔例〕 定義1から，節点数 5，枝数6のランダムグラフは，ラベル付きグラフとして $C(c(5, 2), 6)=210$ 個の異なったグラフが存在する．これら 210 個のグラフは，**Fig. 1** に示す $G_1\sim G_6$ の6つのラベル無しグラフに分類することができ，それぞれの個数は 60, 60, 60, 15, 10, 5 である．従って，節点数 5，枝数6のランダムグラフでは，$G_1\sim G_3$ がそれぞれ 60/210，G_4 が 15/210，G_5 が 10/210，G_6 が 5/210 の確率で出現することになる．また，連結なランダムグラフの中で，$G_1\sim G_5$ のグラフが出現する確率は 60/205, 60/205, 60/205, 15/205, 10/205 である．（例終）

Erdös, Rényi らは，定義1で示したランダムグラフの連結度あるいは連結成分の個数などについて，いくつかの興味ある結果を明らかにしている[3),6)]．以下にその主なものを示す．

〔定理 1〕[3]　節点数n，枝数mの間に

$$m=\left[\frac{n}{2}\log n+Cn\right]\quad C\text{：実定数}$$

の関係があり，nが十分大きいとする．このとき，ランダムグラフが連結である確率は

$$\exp(-e^{-2c})$$

である．

〔定理 2〕[6]　節点数 n，枝数 m の間に

$$m=\left[\frac{n}{2}\log n+Cn\right]\quad C\text{：実定数}$$

の関係があり，nが十分大きいと仮定する．このとき，ランダムグラフが $k+1$ 個の連結成分を持ち，そのうち，k 個が孤立節点である確率は

$$\exp(-2kC-e^{-2c})/k!$$

である．

上の定理で，$\sum_{k=1}^{\infty}\exp(-2kC-e^{-2c})/k!=1$ に注意すると，ランダムグラフの連結成分のほとんどは孤立節点であり，かつ大部分の節点は1つの大きな連結成分に含まれていることがわかる．これは，ランダムグラフの顕著な特徴の1つである．同様にして，連結ランダムグラフでは，大部分の枝は，1つの非可分成分に含まれて，残りの非可分成分のほとんどは，1本の枝だけからなる非可分成分であることが推測されるが，理論的には確かめられていない．本文では，この推測を実証する実験も行っている．

4. ランダムグラフの発生アルゴリズム

本章では，ランダムグラフならびに連結なランダムグラフの発生アルゴリズムを簡単に紹介する．

ランダムグラフ及び連結ランダムグラフの発生プログラムはともに，発生可能な $C(n, 2)$ 本の各枝に，任意のラベル付けをすることを用いている．連結ランダムグラフは，ランダムグラフの中の連結なグラフだけを取り出すことによっても生成することができるが，枝数の少ないグラフを生成するときには，この方法は効率が悪い．そこで，本文では，始めに節点数 n の木を構成してから，残りの枝を付加して連結ランダムグラフを構成している．節点数 n，枝数 m のランダムグラフ及び連結ランダムグラフの生成アルゴリズムを **Fig. 2** 及び **Fig. 3** に示す．

Fig. 2 のアルゴリズムのブロック 1, 2 でかかる手数は $O(n^2)$ である．ブロック 3, 4, 5 の計算手数は

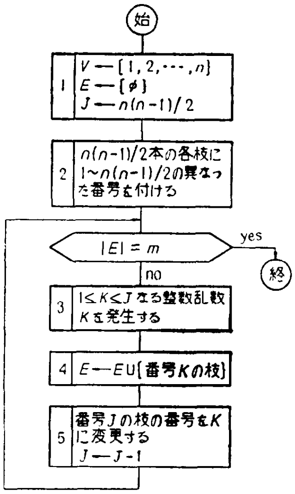

Fig. 2 Algorithm for generating random graphs.

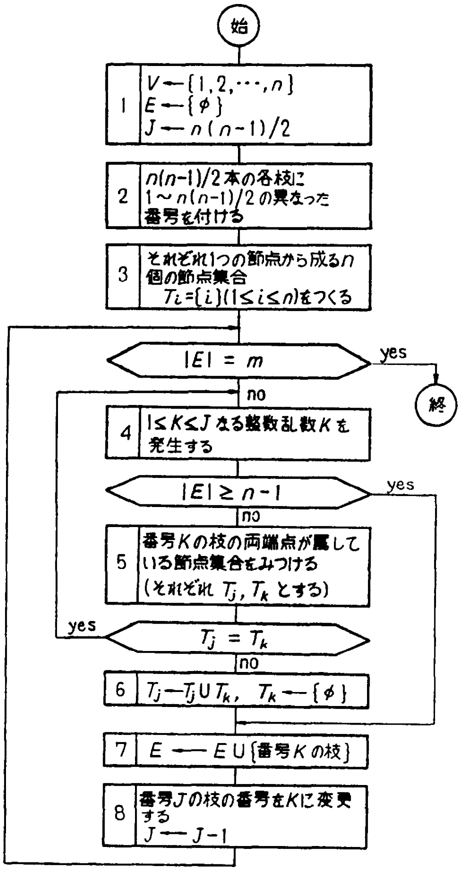

Fig. 3 Algorithm for generating connected random graphs.

$O(1)$ で，ループを回る回数は m 回である．従って，ランダムグラフを発生する計算手数は $O(n^2)$ である．なお，$O(\cdot)$ はいわゆる位数関数である．また，Fig. 3 のアルゴリズムのブロック1から3までの計算手数は $O(n^2)$ である．ブロック4は $O(1)$ である．ブロック5及び6は $T_1, T_2, \cdots, T_n$ に対して適当なデータ構造を考えることにより[7]，それぞれ $O(1)$ 及び $O(\log n)$ の手数で実行できる．また，ブロック7，8は $O(1)$ である．各ブロックの実行回数は，ブロック6が $n-1$ 回，ブロック7と8が m 回である．また，ブロック4と5を実行する回数は期待値として $O(n^2)$ である．従って，連結なランダムグラフを発生するための計算手数は期待値で $O(n^2)$ 以下である．なお，両アルゴリズムとも，Shuffle の手法[8]を用いて，同じ枝が重複して発生することがないようにしている (Fig. 2 のブロック5，Fig. 3 のブロック8)．

5. 実験方法と実験結果

グラフ処理プログラム-GRAMP-には，前章で示したランダムグラフ，連結ランダムグラフの発生プログラムの他，グラフを連結成分あるいは非可分成分に分割するプログラムが用意されている(いずれもFortran Subroutine の形で用意されている)．これらの機能を用いて，以下のような実験を行う．

5.1 連結グラフの発生確率

始めに，多くのランダムグラフを発生して，連結グラフが出現する確率を測定する．連結グラフの発生確率は定理1で与えられるが，これがどの程度の近似となっているのかを調べるのが，この実験の目的である．実際には，同じ節点数，枝数を持つグラフを 100 個発生させ，その中に出現する連結グラフの個数を数えている．**Fig. 4** では，連結グラフの発生確率をパラメータに選び，発生率がそれぞれ 0.1, 0.5, 0.9 となるときの節点数と枝数との関係を示している．実線で示したのは定理1から求められる理論的な近似値であり，実測値とほぼ一致している．定理1は節点数が十分大きいという仮定をおいているが，節点数が 100 個程度でもかなり良い近似となっていると考えることができよう．

5.2 ランダムグラフにおける divide-and-conquer 法の有効性について

グラフ処理問題で，グラフ全体に関する問題が，グラフの個々の連結成分あるいは非可分成分に関する問題に帰着できる場合には（例えば最大カット問題，平面性判定問題など)，グラフをそのようないくつかの部分グラフに分割して，より小規模な問題に帰着させて解く divide-and-conquer 法がしばしば用いられる．

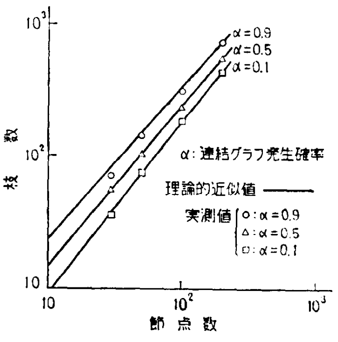

Fig. 4 Measured and theoretical value of probability for random graphs to be connected.

このような手法を用いることにより，対象とするグラフを小さく扱い易いものにすることができる．また，計算時間の短縮も計ることができる．節点数 n のグラフが，節点数 $n_1, n_2, \cdots, n_q$ の連結成分あるいは非可分成分に分割することができるとすれば，もともと $O(n^2)$ の計算時間を必要としたアルゴリズムは，divide-and-conquer 法を用いて，計算速度が $n^2/\sum_{i=1}^{q} n_i^2$ 倍になる．同様に，$O(n\log n)$，$O(n^3)$ のアルゴリズムでは，それぞれ $n\log n/\sum_{i=1}^{q} n_i\log n_i$，$n^3/\sum_{i=1}^{q} n_i^3$ 倍になる．divide-and-conquer 法は，計算手数のオーダを下げるものではないが，計算時間短縮のためによく用いられる手法である．本文では，ランダムグラフに対する divide-and-coquer 法の有効性を調べるためにに，発生したランダムグラフあるいは連結ランダムグラフを連結成分あるいは非可分成分に分割して，それぞれについて $n^2/\sum_i n_i^2$，$n\log n/\sum_i n_i\log n_i$，$n^3/\sum_i n_i^3$ を算出している．**Fig. 5** から **Fig. 10**(次頁参照)にその結果を示す．各値とも，連結成分に分割するときには枝数／節点数≧1.0 のところで，非可分成分に分割するときには枝数／節点数≧2.0 のところで，既に1に近い値となっている．これより，枝数が節点数の倍以上のランダムグラフでは，divide-and-conquer 法の効果がほとんどないことがわかる．これは，前章で述べたように，本文で定義されたランダムグラフでは，大部分の節点が，ある1つの大きな連結成分あるいは非可分成分に含まれてしまうことによるものであろう．この性質は，次節の連結ランダムグラフの非可分成分の個数に関する実験で，更に明確になる．

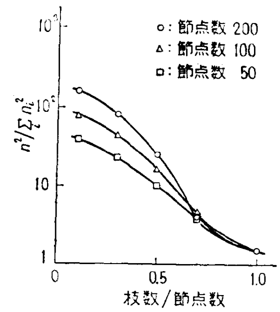

Fig. 5 The computational speed ratio ($n^2/\sum_i n_i^2$ improved by the use of the divide-and-conquer method for a certain kind of random graphs which has originally computational time of $O(n^2)$.

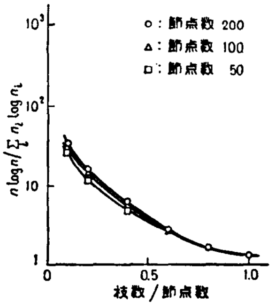

Fig. 6 The computational speed ratio($n\log n/\sum_i n_i\log n_i$) improved by the use of the divide-and-conquer method for a certain kind of process of random graphs which has originally computational time of $O(n\log n)$.

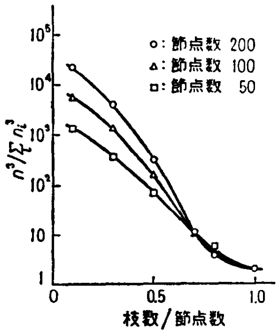

Fig. 7 The computational speed ratio ($n^3/\sum_i n_i^3$) improved by the use of the divide-and-conquer method for a certain kind of process of random graphs which has originally computational time of $O(n^3)$.

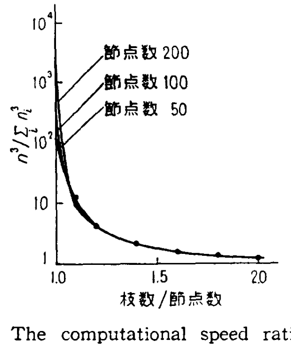

Fig. 8 The computational speed ratio ($n^2/\sum_i n_i^2$) improved by the use of the divide-and-conquer method for a certain kind of process of connected random graphs which has originally computational time of $O(n^2)$.

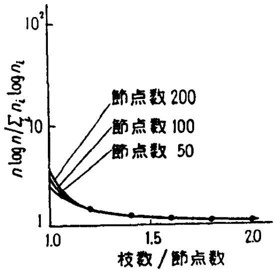

Fig. 9 The computational speed ratio ($n\log n/\sum_i n_i\log n_i$) improved by the use of the divide-and-conquer method for a certain kind of process of connected random graphs which has orignally computational time of $O(n\log n)$.

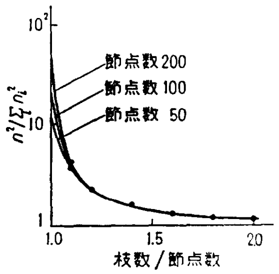

Fig. 10 The computatioinal speed ratio ($n^3/\sum_i n_i^3$) improved by the use of the divide-and-conquer method for a certain kind of process of connected random graphs which has originally computational time of $O(n^3)$.

5.3 ランダムグラフの非可分成分数

定理2及び前節の実験結果から，連結ランダムグラフの非可分成分は，ほとんどが1本の枝だけからなる非可分成分であり，大部分の節点及び枝は，ある1つの大きな非可分成分に含まれていることが推測される．本節では，この推測を実証するために，連結ランダムグラフの平均非可分成分数ならびに4本以上の枝を含む非可分成分の個数と枝数との関係について測定を行っている．**Fig. 11**(次頁参照)及び **Fig. 12**(次頁参照) にその結果を示す．この結果から，ほとんどの連結ランダムグラフは，3本以上の枝を含む非可分成分を1つしか持たないことがわかる．

5.4 有向グラフの強連結成分数

ここでは，有向弱連結ランダムグラフの強連結成分数，ならびに2個以上の節点からなる強連結成分の個数に関する測定を行う．**Fig. 13** と **Fig. 14**(次頁参照)

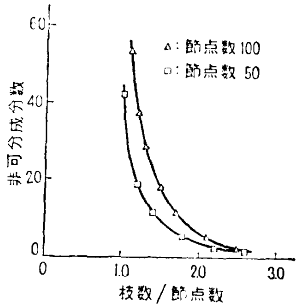

Fig. 11 Number of biconnected components in a connected random graph.

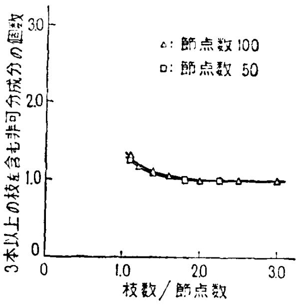

Fig. 12 Number of biconnected components containing three or more edges in a connected random graph.

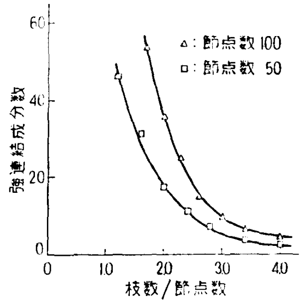

Fig. 13 Number of strongly connected components in a directed random graph.

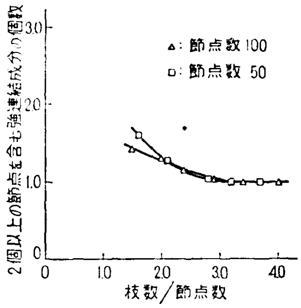

Fig. 14 Number of strongly connected components containing two or more vertices in a directed random graph.

は，100 個のグラフに関するそれぞれの値の平均値と枝数との関係を示したものである．測定結果から，有向弱連結ランダムグラフは，2個以上の節点からなる強連結成分をほぼ1つしか持たないことがわかる．

6. あとがき

筆者らが開発したグラフ処理プログラフ-GRAMP-を用いて，ランダムグラフの統計的解析を行った．本文で取り上げたランダムグラフでは，連結成分の大部分が孤立節点であることが示されていた．本文では，それを確認するとともに，連結なランダムグラフの場合にも「大部分の節点は1つの非可分成分に含まれており，残りの非可分成分は，ほとんど1本の枝だけからなる成分である」という性質があることを明らかにした．従って，グラフを連結成分あるいは非可分成分に分割して計算時間の短縮をはかろうとするdivide-and-conquer 法が，ランダムグラフに関しては，それ程の効果がないことが示された．更に，有向ランダムグラフの強連結成分についても，同様な特徴があることを実験によって明らかにした．

このような実験結果をみる限り，例えば，枝数の少ない非可分グラフの発生確率が非常に小さいなど，ここで取り扱ったランダムグラフと，直感的にランダムな回路図あるいは道路地図などから想起する“ランダムなグラフ”との間には，大きな隔たりがあるように思われる．ランダムな枝の選択が，かえって枝の集中を促す結果となっているような印象を受ける．いずれにしても，本文で取り上げたランダムグラフを用いて計算機実験を行う場合には，ここで明らかにされた性質を充分に考慮する必要があろう．また，そのような計算機実験が，実際のアルゴリズムの有効性の評価という点で，どのような意味をもつのかを明らかにすることは，アルゴリズムの期待値評価などに関係があり，今後の研究課題として残されている．

今後は，本文で明らかにされた性質を考慮に入れた新しいランダムグラフの定義等の研究が待たれる．例えば，節点にラベルの無いグラフ上でランダムグラフを定義することも考えられる．すなわち，ラベル無しグラフとして同形なものは1つのグラフとして，それらが等確率で生起するものをランダムグラフとすること

とである．しかし，同形問題を解く効率の良いアルゴリズムがみつかっていない現在では，このようなランダムグラフを効率良く発生させるのは困難である．また，ラベル無しグラフの場合にも，グラフの数は，節点，枝の増加に伴って，指数関数的に多くなることを考慮に入れると，大部分の節点，枝が1つの非可分成分に含まれるグラフの方が，いくつか同規模な非可成分をもつグラフより多いと思われる．従って，ラベル付きグラフの場合ほど顕著ではないにしても，ラベル無しグラフの場合にも上述の性質があると思われる．

本研究の一部は，文部省科学研究費補助金：総合研究A135017（昭和51年度）「ネットワーク構造を持つシステムに関する基礎研究」及び奨励研究A175174（昭和51年度）「大規模システム及びネットワークの計算機処理に関するグラフ理論的研究」の援助のもとで行われたものである．

参 考 文 献

1) M. R. Garey and D. S. Johnson: The complexity of nearoptimal graph coloring, J. ACM, Vol. 23, 1, pp. 43～46 (Jan. 1976).
2) 西関，小川，斎藤：高次極大カットについて，信学会回路とシステム研資 CST 76-95(1976-10).
3) P. Erdös and A. Rényi: On random graphs I., Publicat. Math. Debrecen, 6, pp. 290～297 (1959).
4) 滝内，高見沢，西関，斎藤：グラフ処理言語 -GRAMP-，信学会回路とシステム研資 CST 76-117 (1976-12).
5) 高見沢，滝内，西関，斎藤：ランダムグラフの統計解析，信学会回路とシステム研資 CST-76-122 (1976-12).
6) P. Erdös and A. Rényi: On the existance of a factor of degree one of a connected random graph, Acta Math. Acad. Sci. Hungaricae 17, pp. 359～368 (1966).
7) A. V. Aho, J. E. Hopcroft and J. D. Ullman: The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading (1974).
8) D. E. Knuth: The Art of Computer Programming, Vol. 21 Seminumerical Algorithms, Addison-Wesley, Reading (1969).

(昭和52年5月18日受付)
(昭和53年1月20日再受付)