

# ランダムデータサーバの開発

3 J-3

二村 良彦 遠藤 貢一 根本 強

早稲田大学 理工学部

## 1. はじめに

アルゴリズムの精密な評価を行うためには、その性能に影響を及ぼす性質を有するランダムデータを用いたシミュレーションを行う必要がある。しかし、指定された性質を持つランダムデータを生成することはその道の専門家でない学生や研究者にとって容易ではない。従って、プログラミングの教科書には、最も簡単な、一様乱数列の作り方さえ載っていないものが多い。また、一様乱順列の作り方にいたってはさらに少ない。しかし、大学におけるアルゴリズムの講義においてプログラムを説明するだけでは、例えば整列法等、何種類も有るアルゴリズムの特性を学生に理解させることは困難である。色々なデータを与えてプログラムの実行結果を比較し、初めてアルゴリズムの特性を理解させることができる。ところが、従来は所望の特性を与えてデータをランダムに生成させるツールが無かった。従来は、言語処理系が備えている乱数発生器を利用して一様乱数列を作る位しか打つ手がなかった。しかし、一様乱数列は非常に偏った構造を持ち、アルゴリズムの公平な評価に適しているとはいえない。ある種のアルゴリズム（例えば、QUICKSORT）は一様乱数列に対して特に速く走り、しかもそのアルゴリズムが現実的に利用される場合のデータが一様乱数である場合が殆どないということがあるからである。

本研究の目的は、指定された性質を有する乱順列を、インターネットを通じて、学生・研究者等ランダムデータを必要とする人たちに供給することである。これにより、例えば、学生が開発した整列アルゴリズム等の精密な評価が可能になり、アルゴリズム教育上有意義であると考えられる。また、アルゴリズムの研究者にと

っても、手軽にランダムデータが入手できれば研究効率の向上が期待できる。

本稿では、ランダムデータサーバ (RDS) の実現方法とユーザインタフェースについて報告する。

## 2. 順列の特性指標と生成方法

順列の特性指標とは、順列から非負の整数上への関数または関数値である[2]。われわれの RDS が制御する事のできる特性指標は、順列の長さ(順列の長さ自身は自明な特性指標である)、葉数(順列において自分よりも小さい隣人を持たない要素の個数)、ランズ及び上昇部分(順列において、左側の要素が右側の要素よりも小さい2つの隣接する要素)数のみである。これらの指標を取り上げた理由は、その指標を持つ乱順列を  $O(n)$  で生成する方法または近似式が知られているからである[2]。

長さ  $n$  の一様乱順列の生成法は文献[3]のシャフル法に基づく。葉数、ランズおよび上昇部分数を指定して乱順列を生成する方法は[2]の近似法に基づく。RDS では比較的精度の良い近似式を用いているが、近似式には違いないので、生成される順列の一様性は保証されない[2]。例えば長さ  $n$  葉数  $m$  の乱順列を生成する際に用いられる確率関数のための近似式は下記の関数  $F2$  である[1, 2] :

$$2m \leq n < 3m-2 \text{ の時 } F2(n, m) = \{0.75^{(n-m)/m}\} * (m-1) / u, \quad 3m-2 \leq n \text{ の時 } F2(n, m) = \{0.75^{n/m}\} * (m-1) / u.$$

この近似式の精度については文献[2]で詳しく議論されている。

一様性の保証が欲しいか、あるいは上記以外の指標を持つ乱順列が欲しい場合には、前講演(3J-02)の RDGG を用いる必要がある。RDGG は計算誤差(計算機  $\epsilon$ )の範囲で、生成される乱順列の一様性を保証する[2]。しかしその生成に要する計算量は、一般的には  $O(n^2)$  時間およびスペースである。従って RDGG を用いては、例えば長さ 10 万以上の大きな順列の生成は不可能である。

### 3. ユーザインタフェース

ランダムデータサーバ(RDS)の目的は、ランダムデータを必要とする学生・研究者等に、指定された性質を有する乱順列を必要なだけ供給することにある。従ってそれは手軽で扱い易い必要がある。これらの点を考慮すると、RDSにはWWW(World Wide Web)上でアクセスできることが望ましい。その理由は次の2点である。

- ・世界中の何処からでも誰にでも手軽にアクセスできる。
- ・ブラウザを用いることにより、誰でも容易に扱うことが可能である。

具体的には、RDSのユーザインタフェースはHTMLで記述し、ユーザ(クライアント)とサーバとのやりとりにはCGIを使用する。ユーザはブラウザを用いてサーバにアクセスする。即ちユーザがRDSを利用する手順は次の通りである：

(1)ブラウザ上で所望の乱順列の各種条件(乱順列の長さ、特性指標および個数)を指定する。ユーザ画面は図1のようになる。

(2)CGIを通して、(1)の条件を満たす乱順列をサーバ上に生成させる。

(3)上記(2)で生成した乱順列をダウンロードする。このとき生成される乱順列は、扱い易いようにテキスト形式で格納される。また、生成される乱順列は、指定された条件によっては膨大なサイズになる。例えば、長さ100万の乱順列を生成すると、その大きさは約6.8MBにもなる。従って、オプションとして、圧縮形式(tar+gzip形式あるいはzip形式)が指定できるようにする。上の長さ100万の順列でもtar+gzip形式で圧縮すれば、約2.2MBになる。一般的な教育において必要なデータサイズは1000位で十分と考えられる。例えば、長さ1000のデータを1000個必要ならば約2.2MBのデータをダウンロードすることになる。これを学生が1人1人ダウンロードすることが通信コストの関係で無理ならば、教員がダウンロードし、適当な方法(FDD, LAN等)で配布しても良い。RDSにおける乱順列生成方式は特許出願済みのものである[1]ので、RDS自身の無料配布は容易ではない。このように、

知的所有権とテクニカルノウハウの凝縮されたソフトウェアを用いて、社会にサービスする機会が今後ますます増大すると考えられる。ネットワークを通じてのサービスでは、時間的あるいはデータサイズの不満足なユーザーにとっては、特許の使用等の困難が生じる可能性がある。ユーザーとしては、交換可能なテクニカルノウハウを所有している必要があろう。

### 4. おわりに

アルゴリズムの性能評価には多種類の特性指標の乱順列の生成が必須である。しかし上で述べたとおり、速度及びスペースの点で実用的な乱順列ジェネレータが扱うことのできる特性指標は、今のところ一様乱数列、葉数およびランズ(または上昇部分数)のみである。その他の特性指標に対する実用的な乱順列の生成法は今後の課題である。

生成する乱順列の条件

長さ:

特性指標:

個数:

圧縮形式:  圧縮しない  tar+gzip形式  zip形式

図1 ユーザインタフェースの例

### 5. 参考文献

- [1] 二村, 高野, 青木: 特性指標を制御した乱順列生成方法およびそのプログラム, 特願平 8-250030, 1996年9月
- [2] 二村, 大谷, 青木, 二村: 単純指標を持つ乱順列の高速生成法, 情報処理学会アルゴリズム研究会, 97年1月.
- [3] Knuth, D.: The Art of Computer Programming, Vol.1-3, Addison-Wesley, 1973.