連続系アルゴリズム レポート課題10

- 宿題(レポート)
  - 提出状況に応じて，期末試験の点に最大 50 点を加算します
    - まったく提出しなくても，期末試験の点で成績が出ます
    - ただし，試験を受けなければ単位は出ません
  - プログラムだけはメールの添付ファイルで
    reiji@is.s.u-tokyo.ac.jp
    に送付してください
    - 所属学科，学年，学籍番号，氏名をメール本文に明記すること
    - メールの題名(Subject)は「連続系レポート課題第10回 提出者氏名」とすること
  - 手計算をするものや，計算結果と考察など，プログラム以外は A4 の紙 1 枚(裏もつかってよい)にまとめてください
    - 所属学科，学年，学籍番号，氏名をレポートの最初に明記すること
    - 次回の講義の前に集めます
    - あるいは PDF で A4 サイズ(ページ数任意)でもよい
- 問題1
  - u_i を [0,1] の一様乱数として，$y_i = u_i^2$ とする．
  - y_i の期待値と分散を求めよ．
  - 以下の Y の期待値と分散を求めよ．

$$Y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- 問題2
  - [0,1) の一様乱数 u_i (i = 1, 2, ..., n)を発生させ，$y_i = u_i^2$ の標本平均と標本分散を求めよ．
  - 余力があれば，n を 2 倍ずつにして，(1) n に対して標本平均と母平均との差，(2) n に対して標本分散と母分散との差，を両対数グラフにプロットせよ．

期末試験 1/23 持込不可

線形合同法の最長周期

- 線形合同法の最長周期は m
  - c と m が互いに素
  - $a-1$ が m のすべての素因数で割り切れる
  - m が 4 の倍数なら，$a-1$ も 4 の倍数
- 乗算合同法: $m = 2^e$ ($e \geqq 4$)の最長周期は $m/4$
  - 周期が $m/4 \Leftrightarrow a \bmod 8$ が 3 or 5 で，x_0 が奇数
- 乗算合同法: m が 2 より大きい素数なら $m-1$
  - 周期が $m-1 \Leftrightarrow x_0 \neq 0$ で，a が p の原始根
    - つまり $m-1$ の任意の素因数 q に対し $a^{(m-1)/q} \neq 1 \bmod m$

38

擬似乱数の危険性

A. M. Ferrenberg, D. Landau, and Y. J. Wong, Phys. Rev. Lett. 69, 3382 (1992).

- Ising model のシミュレーションで，統計的に正しくない(と思われる)結果が得られた
  - 上位ビットに弱点のある乱数だったという説がある

TABLE I. Values of the internal energy (top) and specific heat (bottom) for ten independent runs with $L=16$ at K_c obtained using the Wolff algorithm. The last number in each column, labeled "dev." gives the difference between the simulation value and the exact value, measured in terms of the standard deviation σ of the simulation.

	CONG	R250	R1279	SWC	SWCW
$-<E>$	1.453055	1.455017	1.453056	1.452320	1.453153
error	0.000030	0.000046	0.000032	0.000044	0.000046
dev.	-0.31σ	42.09σ	-0.27σ	-16.95σ	1.94σ
$<C>$	1.498860	1.448627	1.497926	1.514237	1.497398
error	0.000182	0.000467	0.000250	0.000473	0.000356
dev.	0.82σ	-107.16σ	-3.14σ	32.81σ	-3.68σ
	合同法	M 系列(2 種類)		Subtract with carry	

$k(v)$ 均等分布

- 乱数列の上位 v ビットを k 個ずつまとめた kv ビットのパターンが，一周期に同じ回数ずつ表れる
  - ただし，「すべて 0」というパターンだけは 1 回少なくてもよい

Table II. Parameters and k-distribution of Mersenne Twisters

ID	Generator Parameters	The order of equidistribution $k(1)$	$k(2)$	$k(3)$	$k(4)$	$k(5)$	$k(6)$
		$k(7)$	$k(8)$	$k(9)$	$k(10)$	$k(11)$	$k(12)$
(the number of terms in the characteristic polynomial)		$k(13)$	$k(14)$	$k(15)$	$k(16)$	$k(17)$	$k(18)$
		$k(19)$	$k(20)$	$k(21)$	$k(22)$	$k(23)$	$k(24)$
		$k(25)$	$k(26)$	$k(27)$	$k(28)$	$k(29)$	$k(30)$
		$k(31)$	$k(32)$				
MT19937	$(w, n, m, r) = (32, 624, 397, 31)$	19937	9968	6240	4984	3738	3115
	$a = 9908B0DF$	2493	2492	1869	1869	1248	1246
	$u = 11$	1246	1246	1246	1246	623	623
	$s = 7$, $b = 9D2C5680$	624	623	624	623	624	623
(135)	$t = 15$, $c = EFC60000$	623	623	623	623	624	623
	$l = 18$	623	623				

すごく周期が長いので，一周期分の分布だけを考えていてはいかんのでは？

M 系列乱数

- xor を使った lagged Fibonacci 乱数

$$x_n = x_{n-q} \text{ xor } x_{n-p} \qquad q<p$$

- 最大周期は $2^p - 1$
  - ビットごとの計算なので，1 ビットと考えてもよい
    - x_{n-1} から x_{n-p} までがすべて同じになると，以降は同じ
    - すべて 0 だと 0 しか発生しない
- 最大周期のとき M 系列（最長周期系列）乱数
  - 例：$p = 89, q = 38$
    - 他，伏見正則：乱数，東京大学出版会に多くの例がある

5

ワードごとの M 系列乱数

- 各ビットが同一の系列で，開始位置が違うだけ
  - 初期値が特殊だと「乱数らしくない」
- k 次均等分布
  - l ビットの M 系列乱数 1 周期につき，k 次元ベクトル

$$(x_n, x_{n-1}, \cdots, x_{n-k+1})$$

が零ベクトルを $2^{p-kl} - 1$ 回，その他のすべての l ビットベクトルを 2^{p-kl} 回取る

  - 均等分布の次元 k は $k \leq p/l$ をみたす
  - Tausworthe 法という初期値の定め方が知られている

6

Mersenne Twister

- l ビットの乱数を l 元横ベクトルとみなし，サイズが $l \times l$ のビット行列 A を使い

$$x_n = x_{n-q} \text{ xor } \underbrace{(x^u_{n-p} \mid x^l_{n-p+1})}_{\substack{x_{n-p} \text{ の上 } r \text{ ビットと} \\ x_{n-p+1} \text{ の下 } l-r \text{ ビットを} \\ \text{つなげたもの}}} A$$

- 周期は最大で $2^{pl-r} - 1$ となる

M. Matsumoto and T. Nishimura, "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator", ACM Trans. on Modeling and Computer Simulation Vol. 8, No. 1, pp.3-30 (1998)

http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html

7

Mersenne Twister の特徴

- いまのところ，地上最強の擬似乱数生成法（？）
  - 標準版である MT19937 の周期は $2^{19937} - 1$
  - 623 次元まで均等分布
  - 生成もかなり高速（624 ワードのメモリが必要）
- 弱点
  - 0 が沢山続いてしまう傾向（M 系列の一種なので）
  - 624 次元以上の分布は急に悪くなる（らしい）
    - 緩やかに悪くなるように改良を検討中というウワサ
  - Ising model で悪い結果が出たという人もいる

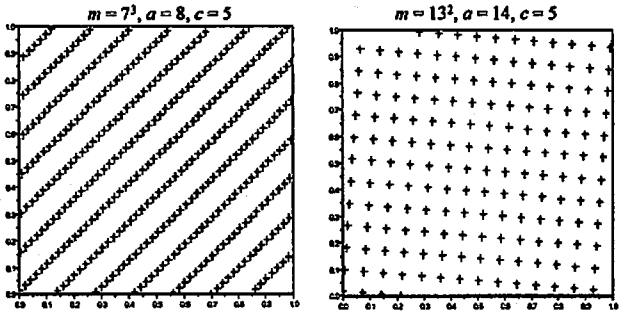
8

線形合同法の弱点

- 下位ビットはランダム性が悪い
  - 0 から $t-1$ までの整数の乱数を得るのに $x_n \bmod t$ としてはだめ
    - $t * x_n / (\mathrm{RAND\_MAX} + 1.0)$
  - ルーチンによっては下位ビットをシフトアウトして返す
- ブロック構造
  - ある t と b に対して $x_{k+t} = x_k + b$ という構造を持つ
  - 実質的な周期は t になってしまう

疎結晶構造

- (x_n, x_{n-1}) をプロット



小さくなる 収容ロジック

多次元疎結晶構造

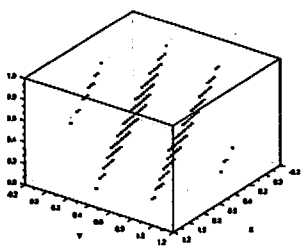
- 疎結晶構造は高次元になると極端になる
  - k 次元では $(k!\, m)^{1/k}$ 枚の等間隔の超平面に乗る

m \ k	3	5	10
2^{16}	73	23	13
2^{32}	2,953	220	41
2^{48}	119,086	2021	126

伏見正則:乱数, 東京大学出版会

    - これは上限で, 実際の乱数はもっと悪い

(x_n, x_{n-1}, x_{n-2}) をプロット



$m = 13^2, a = 14, c = 5$

Additive Lagged Fibonacci 乱数

- 生成法 $(j < k)$

$$x_n = x_{n-j} + x_{n-k} \bmod 2^m$$

  - 最大周期 $(2^k - 1)\, 2^{m-1}$
  - この周期を持つ独立な系列 $2^{(k-1)(m-1)}$ 個
    - 並列乱数に向いている
- 報告されている弱点
  - 連続する乱数の和が正規分布からずれる

M. Matsumoto and T. Nishimura "Sum-discrepancy test on pseudorandom number generators" Mathematics and Computers in Simulation, Vol. 62 (2003), pp 431-442

逐次乱アルゴリズム 1/16

$[0,1)$ の一様乱数 u を整数から
↑ つくる。

0 ~ RAND_MAX
整数の一様乱数 x

$u = x / ((\text{double})\text{RAND\_MAX} + 1.0)$

下位ビットの ランダム性が悪い。

線型合同法

$$x_n = a x_{n-1} + c \mod m$$

a: 乗数　c: 加数　m: 法

cf. $c=0$ のとき、乗算合同法という。

1回用いるものが出ると、あとは すべて同じ。

周期

m 回 以内で もとの数字に もどる。

確率密度 $p(x)$ に 従う乱数が ほしい。

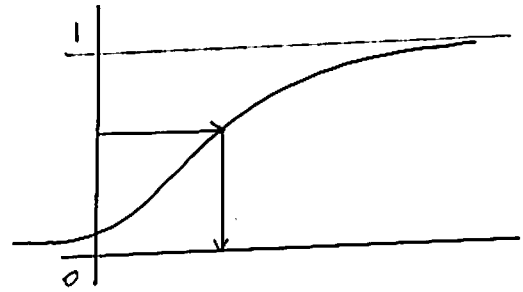
逆関数法

累積密度関数 $f(u) = \int_{-\infty}^{x} p(\xi)\, d\xi$

$f^{-1}(\xi) = \inf \{ x \mid f(x) \geq \xi \}$

計算できれば、$u \in (0,1)$

$x = f^{-1}(u)$



Box-Muller 法.

一様乱数 u_1, u_2 から

$x_1 = \sqrt{-2\log u_1}\cos(2\pi u_2)$

$x_2 = \sqrt{-2\log u_1}\sin(2\pi u_2)$

負もほしい.

$u_1 \in (0,1]$, $u_2 \in [0,1)$

これで独立な標準正規乱数となる.

$x = u_1 + u_2 + \cdots + u_{12} - 6$.

ほぼ正規乱数.

棄却法

ほしい密度関数 $p(x)$

もとの密度関数 $q(x)$.

c: 定数 : $p(x) \le cq(x)$.

アルゴリズム.

1. $q(x)$ に従う乱数 x. $[0,1)$ の一様乱数 u
2. $u \cdot c \cdot q(x) \le p(x)$ なら. x を採用.

   そうでなければ棄却. 1 にもどってやりなおす.

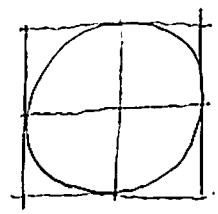
例.

(円内に分布するようなものがほしい)

半径円内で一様分布が良い.

$[-1,1] \times [-1,1]$ 上の一様分布する乱数 (x,y)

を取り $x^2+y^2 \le 1$ なら採用.



Monte Carlo 法.

確率分布 $p(x)$ に従う乱数 x_j を発生させて,

$$I_N = \frac{1}{N}\sum_{j=1}^{N} f(x_j).$$

多次元の積分でつかう

$$E(I_N) = \int f(x)\,p(x)\,dx \quad \cdots \text{積分の近似}.$$

$$\sigma_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(f(x_i) - I_N\right)^2}$$

← 標準偏差.

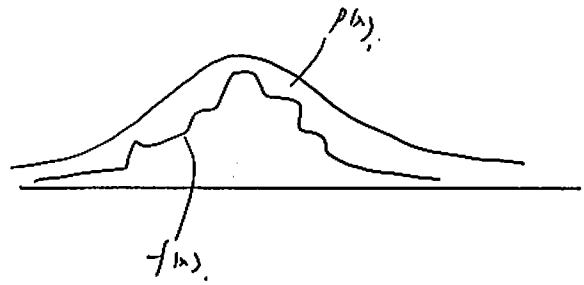
積分値.

$$I_N \pm a\frac{\sigma_N}{\sqrt{N}}$$

N 個の乱数で誤差 $O\left(\frac{1}{\sqrt{N}}\right)$

$p(x)$ に従う乱数.

$$\frac{1}{N}\sum\frac{f(x_i)}{p(x_i)} \approx \int \frac{f(x)}{p(x)}\,p(x)\,dx.$$



$p(x) \propto |f(x)|$ のとき

分散が最小になる」

層別サンプリング.

分散はできるだけ小さくしたい.

いろいろ.