Tạp chí Tin học và Điều khiển học, T.27, S.2 (2011), 176–186

NHỮNG THỬ NGHIỆM TẠO CÁC DÃY TỰA-NGẪU-NHIÊN HAI CHIỀU

VŨ HOÀI CHƯƠNG

Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Việt Nam

Abstract. Halton/Hammersley sequences, the most well-known and widely used multi-dimensional quasirandom sequences, are constructed by combining van der Corput sequences with on another or with k/N sequence. We attempt to generate two-dimensional sequences by combining two simple one-dimensional sequences (k/N, mid-point, V, van der Corput, Richtmyer, pseudorandom sequences) and experimentally receive many lowest discrepancy sequences.

Tóm tắt. Các dãy Halton/Hammersley, những dãy tựa-ngẫu-nhiên nhiều chiều quen thuộc nhất và thông dụng nhất, được tạo thành bằng cách kết hợp các dãy van der Corput với nhau hoặc với dãy k/N. Chúng tôi thử tạo ra các dãy hai chiều khác bằng cách kết hợp hai dãy một chiều đơn giản (dãy k/N, dãy trung điểm, các dãy V, các dãy van der Corput, các dãy Richtmyer, các dãy giả-ngẫu-nhiên R) và bằng thực nghiệm đã thu được nhiều dãy tựa-ngẫu-nhiên có độ phân kỳ thấp hơn.

1. MỞ ĐẦU

Bài báo này mở rộng [13] sang trường hợp 2 chiều. Ta sử dụng 24 dãy một chiều bao gồm 5 dãy cách đều, 12 dãy van der Corput (ký hiệu là vdC), 5 dãy Richtmyer và 2 dãy giả-ngẫu-nhiên. Mỗi dãy được kết hợp với 23 dãy khác để tạo thành dãy 2 chiều $ds(i, j)$. Thông thường $ds(i, j) \neq ds(j, i)$. Như vậy có $24 \times 23 = 552$ dãy hai chiều được khảo sát. Việc kết hợp các dãy nào với nhau để đạt được độ phân kỳ thấp nhất là vấn đề có ý nghĩa thực nghiệm hơn là lý thuyết.

Hai dãy vdC khác nhau kết hợp lại thành một dãy Halton 2 chiều; như vậy 12 dãy vdC tạo thành $12 \times 11 = 132$ dãy Halton. Dãy hữu hạn k/N kết hợp với dãy vdC tạo thành một tập điểm Hammersley. Hai dãy hữu hạn $(k-1)/N$ và $(k-0,5)/N$ kết hợp với 12 dãy vdC tạo thành 24 tập điểm tựa-Hammersley. Đó là những dãy tựa-ngẫu-nhiên quen thuộc, sẽ làm cơ sở để so sánh với các dãy thử nghiệm do chúng tôi đề xuất.

Các tính toán cụ thể đã cho phép tìm được một số dãy tựa-ngẫu-nhiên đơn giản mà lại phân bố đều hơn các dãy Halton và các tập điểm tựa-Hammersley nói trên.

Kết quả thử nghiệm cho thấy: các dãy tựa-ngẫu-nhiên hai chiều có độ phân kỳ thấp là sự kết hợp của một (và chỉ một) dãy cách đều với một dãy tựa-ngẫu-nhiên khác. Nếu kết hợp hai dãy cách đều với nhau ta sẽ thu được dãy tựa-ngẫu-nhiên hai chiều rất kém, không thể chấp nhận được.

2. CÁC KHÁI NIỆM CƠ BẢN, CÁC NGUYÊN TẮC CHUNG

2.1. Các dãy tựa-ngẫu-nhiên

Phương pháp Monte Carlo là phương pháp xác suất dựa trên việc sử dụng các số ngẫu nhiên (random numbers) hoặc giả-ngẫu-nhiên (pseudorandom numbers). Bài báo đầu tiên về phương pháp này là của N. Metropolis (1915-1999) và S. Ulam (1909-1984), công bố năm 1949 ([5]).

Sau đó chỉ 2 năm, bản sao tất định của phương pháp Monte Carlo ra đời nhờ các nhà số luận. Tên gọi phương pháp tựa-Monte-Carlo (quasi-Monte Carlo methods) được dùng đến lần đầu tiên trong một báo cáo nghiên cứu vào năm 1951 ([7]) của R. D. Richtmyer (Mỹ, 1910- 2003).

Ý tưởng của phương pháp tựa-Monte-Carlo là thay thế các dãy số ngẫu nhiên hoặc giả-ngẫu-nhiên bằng các dãy số tất định (hoàn toàn không-ngẫu-nhiên) nhưng phân bố rất đều. Đó là các dãy tựa-ngẫu-nhiên (quasirandom) hoặc cận-ngẫu-nhiên (subrandom) hay cũng được gọi là các dãy có độ phân kỳ thấp (low-discrepancy sequences). Khác biệt chủ yếu của phương pháp tựa-Monte-Carlo là dứt khoát không bắt chước tính ngẫu nhiên, mà tăng độ chính xác và giảm thời gian tính toán bằng cách đặc biệt chú trọng tính phân bố đều [4, 10].

Ước lượng sai số của phương pháp Monte Carlo có tính chất xác suất, còn của phương pháp tựa-Monte-Carlo là tất định.

Các dãy số tựa-ngẫu-nhiên đáng kể nhất gắn liền với tên các nhà toán học J. van der Corput (Hà Lan, 1890-1975), J. H. Halton (Mỹ), J. M. Hammersley (Anh, 1920-2004), I. M. Sobol (Nga, 1926-), H. Niederreiter (Áo, 1944-), và H. Faure (Pháp).

Các dãy tựa-ngẫu-nhiên phân bố đều hơn các dãy giả-ngẫu-nhiên cho nên đại diện tốt hơn cho $U[0, 1]$. Trong một số trường hợp chúng cho phép cải tiến hiệu quả tính toán: có thể tăng tốc độ hội tụ đặc trưng của phương pháp Monte Carlo là $O(1/N^{1/2})$ lên đến gần $O(1/N^{1-?})$, trong đó N là số điểm tính toán. Các kỹ thuật giảm phương sai [3] chỉ có thể tác động đến hằng số ẩn trong $O(N^{-1/2})$ chứ không thể đạt tới bước nhảy dài như thế.

Điều đặc biệt là các số giả-ngẫu-nhiên do các nhà thống kê đưa ra, còn các dãy tựa-ngẫu-nhiên lại do các nhà số luận. Các công cụ để phát triển và phân tích các dãy tựa-ngẫu-nhiên khác hẳn với phương pháp Monte Carlo, vì chúng gần với lý thuyết số và đại số trừu tượng hơn là lý thuyết xác suất và thống kê toán học. Việc đổi cơ số, sử dụng các tính chất của số nguyên tố, các hệ số của đa thức nguyên thuỷ, v.v. chính là các khái niệm của lý thuyết số [6].

Các dãy tựa-ngẫu-nhiên là hoàn toàn xác định, cho nên cần phải có một độ đo sự phân tán, tương tự như phương sai của các số ngẫu nhiên. Đó chính là độ phân kỳ (Discrepancy). Các khái niệm, định lý và cách tính độ phân kỳ được trình bày cặn kẽ trong [13] cho nên bài này không nhắc lại.

2.2. Dãy van der Corput (vdC)

Đây là dãy số phân bố đều trong khoảng $(0, 1)$, do nhà toán học Hà Lan Johannes

Gualtherus van der Corput, đề xuất vào năm 1935. Dãy số này lấp đầy đoạn $[0, 1)$ một cách điều hoà với độ phân kỳ nhỏ. Có thể coi nó là dãy tựa-ngẫu-nhiên một chiều quan trọng nhất, đẹp nhất, là yếu tố then chốt của nhiều cấu trúc đa chiều và là cơ sở để đi tới các dãy tựa-ngẫu-nhiên quen thuộc khác như Halton, Faure và Sobol.

Lấy một số nguyên tố $p > 1$ làm cơ số. Mọi số nguyên dương k đều có phép biểu diễn duy nhất theo cơ số p

$$k = \sum_{j \geq 0} a_j(k) p^j,$$

trong đó $a_j(k) \in F_p = \{0, 1, 2, ..., p-1\}$ và $a_j(k) = 0$ khi k khá lớn, tức là tổng nói trên hữu hạn.

Mỗi giá trị k sẽ được ánh xạ vào một điểm trong $[0, 1)$ bằng cách đảo ngược các hệ số $a_j(k)$ và tạo nên phân số $0, a_0 a_1 a_2 ...$ (trong hệ đếm cơ số p). Chính xác hơn

$$\varphi_p(k) = \sum_{j \geq 1} a_j(k) p^{-j-1}, \ k > 0,$$

$\varphi_p(k)$ được gọi là hàm ngược gốc (radical inverse function). Dãy van der Corput cơ số p chính là dãy $X = \{\varphi_p(0) = 0, \ \varphi_p(1), \ \varphi_p(2), \ \varphi_p(3), ...\}$.

Trong [14] chúng tôi trình bày tỉ mỉ hơn về dãy số này.

2.3. Dãy Halton (vô hạn)

Dãy Halton là sự mở rộng dãy vdC ra không gian nhiều chiều. Dãy Halton chuẩn s chiều được ghép lại bởi s dãy vdC một chiều với các cơ số nguyên tố khác nhau (thường là s số nguyên tố đầu tiên). Gọi $p_1, p_2, ..., p_s$ là các số nguyên tố cùng nhau thì dãy Halton s chiều là dãy vô hạn $X_0, X_1, ...$

$$X_k = \{\varphi_{p1}(k), \varphi_{p2}(k), ..., \varphi_{ps}(k)\}, \ \forall k > 0,$$

trong đó $\varphi_p = vdC(p) = vdC_p$ là dãy van der Corput với cơ số p.

Ví dụ, với $p_1 = 2$ và $p_2 = 3$ thì

$$\varphi_2 = vdC_2 = \{1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, 5/16, 13/16, 3/16...\},$$

$$\varphi_3 = vdC_3 = \{1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9, 1/27, 10/27, 19/27, 4/27...\}.$$

Dãy Halton (2,3) sẽ là

$$(\varphi_2, \varphi_3) = \{1/2, 1/3), (1/4, 2/3), (3/4, 1/9), (1/8, 4/9), (5/8, 7/9), (3/8, 2/9), (7/8, 5/9), (1/16, 8/9)...\}.$$

Các giá trị đầu tiên của các dãy thành phần (φ_i) phụ thuộc lẫn nhau. Sự phụ thuộc này sẽ triệt tiêu tại cuối chu kỳ thứ nhất của mỗi dãy. Kết quả quan sát thực tế cho thấy: khi tích phân nhiều chiều, 150 điểm Halton tương đương với 500 điểm giả-ngẫu-nhiên. Tuy vậy số chiều càng cao thì tính đều của dãy Halton càng giảm. Khi đó người ta dùng các dãy pha trộn (scrambled sequences) để giải quyết vấn đề này.

Dãy Halton đạt được tính độc lập tiệm cận thông qua việc sử dụng các số nguyên tố khác nhau trong từng chiều khác nhau.

2.4. Các tập điểm Hammersley (hữu hạn)

Các tập điểm Hammersley là sự cải biên của dãy Halton. Đó là dãy hữu hạn s chiều có N phần tử cố định được định nghĩa gần giống dãy Halton, chỉ khác ở chiều đầu tiên

$$X_k = (k/N, \varphi_{p1}(k), \varphi_{p2}(k), ..., \varphi_{p(s-1)}(k)),\ k = 0, 1, ..., N - 1.$$

Ví dụ, với $p = 3$ và $N = 11$ thì tập điểm Hammmersley sẽ là

$$(k/11, \varphi_3) = \{(0, 1/3), (1/11, 2/3), (2/11, 1/9), (3/11, 4/9), (4/11, 7/9), (5/11, 2/9), (6/11, 5/9),$$
$$(7/11, 8/9), (8/11, 1/27), (9/11, 10/27), (10/11, 19/27)\}.$$

2.5. Dãy Richtmyer $\{k.\alpha\}$

Cho $\alpha = (\alpha_1, \alpha_2, ..., \alpha_s)$ là tập các số vô tỉ và $1, \alpha_1, \alpha_2, ..., \alpha_s$ độc lập tuyến tính. Dãy $\{k.\alpha\}$ được cho dưới dạng

$$X_k = (\{k.\alpha_1\}, \{k.\alpha_2\}, ..., \{k.\alpha_s\}),\ k > 0,$$

trong đó $\{x\}$ biểu thị phần thập phân của số x.

Người ta thường chọn α_i là căn của các số nguyên tố. Trong [13] chúng tôi đã chọn thử nghiệm α_i là các hằng số toán học quen biết như π, Au (tỷ lệ vàng), e, C (hằng sô Euler-Mascheroni), G (hằng số Catalan) và thu được kết quả khá tốt.

2.6. Dãy số V

Dãy số V là dãy tựa-ngẫu-nhiên dựa trên các đẳng thức giữa moment mẫu và moment lý thuyết [12, 13]. Khi cần tạo ra N điểm phân bố đều thì tách N thành tổng của h số hạng

$$N = K_1 + K_2 + ... + K_h$$

sao cho các K_i khác nhau.

Trên đoạn $(0, 1)$ các điểm cách đều nhau một khoảng u; điểm đầu (X_1) cách điểm 0 và điểm cuối (X_K) cách điểm 1 một khoảng T

$$X_1 = T, X_2 = T + u, X_3 = T + 2u, ..., X_K = T + (K - 1).u = 1 - T,$$

trong đó $T = (1 - \sqrt{\dfrac{K-1}{K+1}})/2$ và $u = (1 - 2T)/(K - 1) = 1/\sqrt{(K - 1)(K + 1)}$.

3. CÁC DÃY SỐ SO SÁNH THỬ NGHIỆM

Cơ sở để tiến hành so sánh thử nghiệm là tập dãy thành phần gồm 24 dãy một chiều với 2459 phần tử (2459 là số nguyên tố), trong đó có 5 dãy cách đều (ký hiệu là a), 12 dãy vdC (ký hiệu là b), 5 dãy Richtmyer (ký hiệu là c) và 2 dãy giả-ngẫu-nhiên (ký hiệu là R). Ký hiệu k_a, k_b, k_c, k_R là số dãy trong từng nhóm. Mỗi dãy trong tập này được kết hợp với 23 dãy khác để tạo thành dãy 2 chiều $ds(i,j), 1 \leq i \leq 24,\ 1 \leq j \leq 24,\ i \neq j$. Thông thường $ds(i,j) \neq ds(j,i)$. Như vậy tổng cộng có $24 \times 23 = 552$ dãy hai chiều $ds(i,j)$, trong đó có 132 dãy Halton và 24 tập điểm tựa-Hammersley để so sánh.

a) *Các dãy cách đều* ($k_a = 5$)

$\{1\}$: dãy $(k-1)/N$.

$\{2\}$: dãy $(k-0,5)/N$.

$\{3\}$: dãy V_1 gồm một đoạn duy nhất $N = 2459$.

$\{4\}$: dãy V_2 do 17 đoạn hợp thành với độ dài các đoạn là 141, 59, 265, 35, 89, 79, 32, 38, 46, 264, 33, 83, 279, 50, 28, 841, 97.

$\{5\}$: dãy V_3 do 22 đoạn hợp thành với độ dài các đoạn là 24, 415, 23, 722, 64, 9, 25, 66, 424, 14, 10, 2, 29, 8, 88, 12, 5, 3, 311, 92, 93, 20.

b) *Các dãy số van der Corput* ($k_b = 12$)

Từ $\{6\}$ đến $\{17\}$: các dãy số vdC với cơ số là 12 số nguyên tố đầu tiên: {2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37}.

c) *Các dãy Richtmyer* $\{k.\alpha\}$, ($k_c = 5$)

$\{18\}$: dãy số $\{k.\alpha\}$ với $\alpha =$ hằng số Euler = c = 0,577215665.

$\{19\}$: dãy số $\{k.\alpha\}$ với $\alpha = e = 2,718281828$.

$\{20\}$: dãy số $\{k.\alpha\}$ với $\alpha = \pi = 3,141592654$.

$\{21\}$: dãy số $\{k.\alpha\}$ với $\alpha\sqrt{2} = 1,414213562$.

$\{22\}$: dãy số $\{k.\alpha\}$ với $\alpha = \ln 2 = 0,693147181$.

d) *Các dãy số giả-ngẫu-nhiên* R, ($k_R = 5$)

$\{23\}$,$\{24\}$: các dãy số giả-ngẫu-nhiên R_i do máy tính tạo ra bằng hàm Random của Turbo Pascal. Nhằm tăng tính ngẫu nhiên ta đặt thủ tục Randomize ở đầu chương trình để khởi động bộ tạo số với một giá trị ngẫu nhiên.

4. CÁC TIÊU CHUẨN SO SÁNH THỬ NGHIỆM

Việc tính toán thử nghiệm các dãy tựa-ngẫu-nhiên hai chiều $ds(i,j)$ được tiến hành theo 5 tiêu chuẩn dưới đây. Để tiện so sánh, mỗi chỉ số đều được chia cho giá trị tương ứng của dãy số $(1,6)$, ký hiệu là H_2, với $X = (k-1)/N$ (dãy cách đều) và $Y = \varphi_2(k)$ (dãy vdC nhị phân). Như vậy tất cả các chỉ số của $ds(1,6)$ đều bằng 1.

Trong hệ thức trên, nếu thay $(k-1)/N$ bằng k/N thì ta có dãy Hammersley đầu tiên. Do vậy có thể gọi $ds(1,6)$ là dãy tựa-Hammersley.

a) *Độ phân kỳ D*

Đây là tiêu chí quan trọng nhất đối với các dãy số tựa-ngẫu-nhiên. Việc tính toán dựa vào định nghĩa và các công thức quen biết [6, 13].

b) *Hệ số tương quan tuyến tính R giữa các giá trị x và y*

Hệ số này càng nhỏ thì chất lượng của dãy số càng cao.

c) *Phép thử khi bình phương* $\chi^2(23,7)$

Ký hiệu $\chi^2(23,7)$ có nghĩa là đoạn $[0,1]$ trên trục X được chia thành 23 khoảng bằng nhau, và đoạn $[0,1]$ trên trục Y được chia ra làm 7 khoảng bằng nhau.

d) *Sai số tính tích phân (I)*

Các tích phân đơn giản thường dễ tính toán với độ chính xác cao, cho nên khó so sánh. Ta chọn một tích phân khá phức tạp từ đồng nhất thức

$$1/\Gamma(z) = \frac{e}{\pi}\int_0^{\frac{\pi}{2}} \cos(tgT - zT)\cos^{z-2} T dT,$$

ở trang 151 của [1], trong đó $\Gamma(z)$ là hàm gamma.

Giá trị chính xác của tích phân này là $\pi/(e.\Gamma(z))$, phụ thuộc z. Các trị số gần đúng của tích phân được tính bằng phương pháp Monte Carlo hình học [3, 11] với mỗi dãy số $ds(i,j)$ tại 4 giá trị z sau đây

$$z_1 = 19/11,\ z_2 = 405/91,\ z_3 = 20/9,\ z_4 = 21/4.$$

e) *Sai số tìm cực trị của một hàm số (S)*

Trong hình vuông đơn vị $(0 \le x \le 1,\ 0 \le y \le 1)$ ta xét hàm số

$$f(x,y) = 10(x + 0,2 + |x - 0,8| - |x - 0,2|) + |y - 0,5|.$$

Theo ([9], trang 31), giá trị lớn nhất và nhỏ nhất của hàm số này bằng $\max f(x,y) = 10,5$ và $\min f(x,y) = 4$.

Ta tìm các giá trị cực đại $f\max(i,j)$ và cực tiểu $f\min(i,j)$ của hàm số trên bằng cách tìm kiếm hệ thống [8] với các dãy số $ds(i,j)$ và tính

$$\begin{aligned} S(i,j) &= |f\max(i,j) - \max f(x,y)| + |f\min(i,j) - \min f(x,y)| \\ &= |f\max(i,j) - 10,5| + |f\min(i,j) - 4|. \end{aligned}$$

Sau đó xếp hạng theo $S = S(i,j)/S(1,6)$ tăng dần.

6. KẾT QUẢ SO SÁNH

Bảng 1 ghi 24 dãy có chỉ số tốt nhất và kém nhất theo 2 tiêu chuẩn đầu tiên. Các giá trị $D*$ (độ-phân-kỳ-sao) và R (hệ số tương quan tuyến tính) có tính đối xứng với i và j : $D \times (i,j) = D \times (j,i)$ và $R(i,j) = R(j,i)$, cho nên mỗi chỉ số có $552/2 = 276$ cặp giá trị bằng nhau từng đôi một. Bảng 2 ghi 12 dãy có chỉ số tốt nhất và kém nhất theo 3 tiêu chuẩn cuối.

7. TỔNG HỢP, NHẬN XÉT VÀ KẾT LUẬN

Vì các chỉ số đánh dấu phẩy đều là chỉ số tương đối so với dãy số (1,6) = dãy $H2$, cho nên ta có thể tính chỉ số tổng hợp bằng cách lấy trung bình của chúng, trong đó độ phân kỳ quan trọng nhất có hệ số 2

$$Z = (2D' + R' + 2' + I' + S')/6.$$

Các giá trị tốt nhất và kém nhất của Z được ghi trong hai Bảng 3 và 4.

Bảng 1. Độ phân kỳ $D*$ và hệ số tương quan tuyến tính R
(a = dãy cách đều, b = dãy vdC, bb = dãy Halton),
c = dãy Richtmyer, R = dãy tựa-ngẫu-nhiên)

	D' = D*(i,j)/D*H2	ds(i,j)	Dạng	R=R(i,j)/RH2	ds(i,j)	Dạng
Thứ tự	Các giá trị tốt nhất					
1-2	0,854032	(3,6), (6,3)	ab, ba	0,000507	(6,14), (14,6)	bb, bb
3-4	0,854060	(2,6), (6,2)	ab, ba	0,008410	(7,11), (11,7)	bb, bb
5-6	0,904177	(2,21), (21,2)	ac, ca	0,009271	(16,22), (22,16)	bc, cb
7-8	0,904177	(3,21), (21,3)	ac, ca	0,020868	(7,12). (12,7)	bb, bb
9-10	0,983451	(1,21), (21,1)	ac, ca	0,031526	(7,13). (13,7)	bb, bb
11-12	1	(1,6), (6,1)	ab, ba	0,047548	(11,20). (20,11)	bc, cb
13-14	1,088998	(3,19), (19,3)	ac, ca	0,051301	(10,16), (16,10)	bb, bb
15-16	1,089015	(2,19), (19,2)	ac, ca	0,055610	(8,23), (23,8)	bR, Rb
17-18	1,116360	(2,22), (22,2)	ac, ca	0,071059	(4,21), (21,4)	ac, ca
19-20	1,116366	(3,22), (22,3)	ac, ca	0,086286	(7,9), (9,7)	bb, bb
21-22	1,214437	(1,19), (19,1)	ac, ca	0,098452	(3,18), (18,3)	ac, ca
23-24	1,244991	(1,22), (22,1)	ac, ca	0,098452	(2,18), (18,2)	ac, ca
Thứ tự	Các giá trị kém nhất					
23-24	23,141100	(23,24), (24,23)	RR, RR	19,622492	(19,24), (24,19)	cR, Rc
21-22	26,948091	(19,24), (24,19)	cR, Rc	19,881502	(21,24), (24,21)	cR, Rc
19-20	30,259581	(4,5), (5,4)	aa, aa	23,689223	(17,24), (24,17)	bR, Rb
17-18	57,595554	(2,5), (5,2)	aa, aa	70,658153	(1,4), (4,1)	aa, aa
15-16	57,895560	(3,5), (5,3)	aa, aa	70,658153	(2,4), (4,2)	aa, aa
13-14	57,975965	(1,5), (5,1)	aa, aa	70,658153	(3,4), (4,3)	aa, aa
11-12	67,992778	(3,4), (4,3)	aa, aa	71,934054	(1,5), (5,1)	aa, aa
9-10	67,992787	(2,4), (4,2)	aa, aa	71,934054	(3,5), (5,3)	aa, aa
7-8	68,012616	(1,4), (4,1)	aa, aa	71,934054	(2,5), (5,2)	aa, aa
5-6	195,991911	(2,3), (3,2)	aa, aa	431,982446	(1,3), (3,1)	aa, aa
3-4	196,071551	(1,2), (2,1)	aa, aa	431,982446	(1,2), (2,1)	aa, aa
1-2	196,071551	(1,3), (3,1)	aa, aa	431,982446	(2,3), (3,2)	aa, aa

Bảng 2. Giá trị χ^2, sai số tích phân I' và sai số tìm cực trị S'

	$\chi^{2'} = \chi^2/\chi^2 H2$	ds(i,j)	$I' = I(i,j)/I_{H2}$	ds(i,j)	$S' = S(i,j)/S_{H2}$	ds(i,j)
Thứ tự	Các giá trị tốt nhất					
1	0,275707	(14,9) bb	0,180057	(21,1) ca	0,074284	(8,4) ba
2	0,292952	(2,9) ab	0,313200	(22,1) ca	0,087509	(7,8) bb
3	0,292952	(3,9) ab	0,403425	(21,3) ca	0,123449	(21,19) cc
4	0,292952	(1,9) ab	0,403425	(21,2) ca	0,176031	(8,1) ba
5	0,292952	(14,2) ba	0,461154	(17,18) bc	0,180517	(8,3) ba
6	0,292952	(14,3) ba	0,473574	(9,10) bb	0,180517	(8,2) ba
7	0,292952	(14,1) ba	0,476978	(22,2) ca	0,186562	(17,19) bc
8	0,327442	(2,20) ac	0,476978	(22,3) ca	0,238208	(6,5) ba
9	0,327442	(3,20) ac	0,485855	(17,19) bc	0,249297	(15,19) bc
10	0,327442	(1,20) ac	0,489530	(14,7) bb	0,251370	(21,7) cb
11	0,413668	(18,1) ca	0,510910	(19,17) cb	0,274945	(6,15) bc
12	0,430913	(18,2) ca	0,556497	(19,9) cb	0,300666	(19,12) cb
Thứ tự từ dưới lên	Các giá trị kém nhất					
12	570,501285	(1,4) aa	11,399076	(24,23) RR	2,059233	(17,3) ba
11	573,519173	(2,4) aa	12,162913	(24,5) Ra	2,059534	(17,2) ba
10	573,519173	(3,4) aa	13,948491	(4,1) aa	2,064020	(17,1) ba
9	678,869323	(2,5) aa	14,007590	(4,3) aa	2,376522	(16,3) ba
8	678,869323	(3,5) aa	14,007590	(4,2) aa	2,376522	(16,2) ba
7	686,284704	(1,5) aa	15,217713	(2,4) aa	2,376522	(16,1) ba
6	1717,160668	(2,1) aa	15,217713	(3,4) aa	5,538212	(2,1) aa
5	1717,160668	(3,1) aa	15,532020	(1,4) aa	5,538312	(1,2) aa
4	1717,160668	(1,2) aa	15,911344	(5,3) aa	5,538312	(1,3) aa
3	1717,160668	(1,3) aa	15,911344	(5,2) aa	5,538317	(3,1) aa
2	1719,540488	(2,3) aa	15,989001	(5,1) aa	5,542798	(2,3) aa
1	1719,540488	(3,2) aa	41,581399	(1,3) aa	5,542804	(3,2) aa

Bảng 3 cho thấy cả 24 dãy có chỉ số tổng hợp tốt nhất đều có một (và chỉ một) thành phần thuộc nhóm a. Trong khi đó 20/24 dãy trong bảng 4 có dạng aa. Điều đó có nghĩa là các dãy cách đều $\{a\}$ có tầm quan trọng quyết định. Nếu kết hợp dãy này với một dãy vdC $\{b\}$ hoặc một dãy Richtmyer $\{c\}$ sẽ được dãy tựa-ngẫu-nhiên hai chiều có độ phân kỳ thấp. Nhưng nếu kết hợp hai dãy cách đều với nhau thì hậu quả là rất tồi tệ.

Nhận xét thứ hai là, trong số các dãy tốt nhất, tỷ lệ các dãy Halton (dạng bb, $6 \le i,\ j \le 17$) và các dãy tựa-Hammersley ($1 \le i \le 2,\ 6 \le j \le 17$) khá nhỏ. Như vậy qua thử nghiệm, chúng tôi đã tìm được khá nhiều dãy tựa-ngẫu-nhiên mới tốt hơn và sau này có thể dùng chúng trong các thuật toán tựa-Monte-Carlo.

Trong 4 bảng trên có ghi 108 dãy tốt nhất và 108 dãy kém nhất theo 6 tiêu chuẩn khác nhau: 24-24 dãy theo độ phân kỳ D^* và hệ số tương quan tuyến tính R, 12-12-12 dãy theo giá trị χ^2, sai số tích phân I và sai số tìm cực trị S, 24 dãy theo chỉ số tổng hợp Z. Bảng 5 dưới đây thống kê các dạng theo số tuyệt đối (B, D) và tỷ lệ tương đối (C, E). Các dãy dạng aa, bb, cc, RR không thể có 2 thành phần như nhau, cho nên số dãy 2 chiều dạng này là $k_x(k_x - 1)$, trong đó k_x là số dãy trong nhóm.

Bảng 3. Các chỉ số tổng hợp Z tốt nhất

Thứ tự	Z	(i,j)	Dạng	Dãy(i)	Dãy(j)
1	0,715494	(22,2)	ca	{k.ln2}	(k-0,5)/N
2	0,715496	(22,3)	ca	{k.ln2]	V1
3	0,716177	(21,2)	ca	{k. }	(k-0,5)/N
4	0,716177	(21,3)	ca	{k. }	V1
5	0,718812	(21,1)	ca	{k. }	(k-1)/N
6	0,761364	(22,1)	ca	{k.ln2}	(k-1)/N
7	0,846482	(3,21)	ac	V1	{k. }
8	0,849953	(19,3)	ca	{k.e}	V1
9	0,849960	(19,2)	ca	{k.e}	(k-0,5)/N
10	0,866995	(2,21)	ac	(k-0,5)/N	{k. }
11	0,892417	(1,21)	ac	(k-1)/N	{k. }
12	0,900775	(6,3)	ba	vdC(2)	V1
13	0,900786	(6,2)	ba	vdC(2)	(k-0,5)/N
14	0,914380	(19,1)	ca	{k.e}	(k-1)/N
15	0,948212	(2,6)	ab	(k-0,5)/N	vdC(2)
16	0,952238	(3,19)	ac	V1	{k.e}
17	0,952245	(2,19)	ac	(k-0,5)/N	{k.e}
18	0,961210	(6,1)	ba	vdC(2)	(k-1)/N
19	0,968689	(3,6)	ab	V1	vdC(2)
20	0,973057	(3,22)	ac	V1	{k.ln2}
21	0,973929	(2,22)	ac	(k-0,5)/N	{k.ln2}
22	1	(1,6)	ab	(k-1)/N	vdC(2)
23	1,016718	(1,19)	ac	(k-1)/N	{k.e}
24	1,038467	(1,22)	ac	(k-1)/N	{k.ln2}

Bảng 4. Các chỉ số tổng hợp Z kém nhất

Thứ tự	Z	(i,j)	Dạng	Dãy(i)	Dãy(j)
24	18,026256	(5,17)	ab	V3	vdC(37)
23	19,345818	(11,21)	bc	vdC(13)	{k. }
22	19,772996	(18,17)	cb	{k.c}	vdC(37)
21	19,933728	(24,19)	Rc	R	{k.e}
20	46,157529	(5,4)	aa	V3	V2
19	49,223779	(4,5)	aa	V2	V3
18	69,046864	(5,1)	aa	V3	(k-1)/N
17	69,144894	(5,2)	aa	V3	(k-0,5)/N
16	69,144896	(5,3)	aa	V3	V1
15	78,840910	(4,3)	aa	V2	V1
14	78,840914	(4,2)	aa	V2	(k-0,5)/N
13	78,880731	(4,1)	aa	V2	(k-1)/N
12	157,092889	(1,4)	aa	(k-1)/N	V2
11	157,634993	(3,4)	aa	V1	V2
10	157,634996	(2,4)	aa	(k-0,5)/N	V2
9	173,917215	(2,5)	aa	(k-0,5)/N	V3
8	173,917217	(3,5)	aa	V1	V3
7	175,466489	(1,5)	aa	(k-1)/N	V3
6	512,969214	(1,3)	aa	(k-1)/N	V1
5	512,969214	(1,2)	aa	(k-1)/N	(k-0,5)/N
4	513,008789	(2,1)	aa	(k-0,5)/N	(k-1)/N
3	513,008790	(3,1)	aa	V1	(k-1)/N
2	513,419792	(2,3)	aa	(k-0,5)/N	V1
1	513,419793	(3,2)	aa	V1	(k-0,5)/N

Bảng 5. Thống kê theo dạng a, b, c, R qua 6 thứ tự xếp hạng

(**, * = các giá trị lớn nhất, nhì trong cột)

Dạng	số dãy A	số dãy tốt B	Tỷ lệ % C=100B/6A	số dãy kém D	Tỷ lệ % E=100D/6A
aa	5.(5-1)=20	-	-	86	71,67**
ab	5.12=60	9	2,50	1	0,28
ac	5.5=25	24	16,00*	-	-
aR	5.2=10	-	-	-	-
ba	12.5=60	14	3,89	6	1,67
bb	(dãy Halton)	12.(12-1) =132	16	2,02	-
bc	12.5=60	7	1,94	1	0,28
bR	12.2=24	1	0,69	1	0,69
ca	5.5=25	29	19,33**	-	-
cb	5.12=60	6	1,67	1	0,28
cc	5.(5-1)=20	1	0,83	-	-
cR	5.2=10	-	-	3	5,00
Ra	2.5=10	-	-	1	1,67
Rb	2.12=24	1	0,69	1	0,69
Rc	2.5=10	-	-	4	6,67
RR	2.(2-1)=2	-	-	3	25,00*
Tổng số	552		= 24.(24-1)	108	108

Các tỷ lệ C và E cho ta thấy, đa số các dãy tốt có dạng ca hoặc ac, tức là kết hợp giữa một dãy cách đều và một dãy Richtmyer. Còn đa số các dãy kém có dạng aa hoặc RR (2 dãy cách đều hoặc 2 dãy giả-ngẫu nhiên kết hợp với nhau).

TÀI LIỆU THAM KHẢO

[1] Fazekas Ferenc, Frey Tamás, *Operátorszámítás, Speciális függvények*, Tankönyvkiadó, Budapest, 1957.

[2] Frey Tamás, *Végtelen sorozatok, sorok és szorzatok*, Tankönyvkiadó, Budapest, 1973.

[3] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Methuen & Co Ltd, London, 1964.

[4] R. Hungarowitsch, H. Tatvann, and Vu Toan Thang, "On the use of Low Discrepancy Sequences in Practice". Seminar for Applied Mathematics, Technische Universität Wien, 2007.

[5] N. Metropolis and S. M. Ulam, The monte carlo methods, *J. Amer. Statist. Assoc.* **44** (1949) 335–341.

[6] H. Niederreiter, Random number generation and quasi-monte carlo methods, *Society for Industrial and Applied Marhematics*, Philadelphia, Pennsylvania, 1992.

[7] R. D. Richtmyer, "On the evaluation of definite integrals and a quasi-Monte-Carlo method based on the properties of algebraic numbers", Report LA-1342, Los Alamos Scientific Laboratory, Los Alamos, N. M., 1951.

[8] I. M Sobol, On the systematic search in a hypercube, *SIAM Journal on Numerical Analysis* **16** (5) (October 1979) 790-793.

[9] I. M. Sobol & R. B. Statnikov, *Những giải pháp tốt nhất tìm chúng ở đâu ?*, Nhà xuất bản Znanie, Moskva, 1982 (tiếng Nga).

[10] Uhovai-Hügönc István, "Kvázivéletlen pontok a Monte Carlo algoritmusok-ban". Diplomamunka, BME VIK, Budapest, 2003.

[11] Vígh-Turulcha Ödön, "Kvázi Monte Carlo integrálási technikák", TDK-dolgozat, KLTE, Debrecen, 1999.

[12] Vũ Hoài Chương, Một thuật toán đơn giản tạo dãy số tựa ngẫu nhiên, *Tạp chí Khoa học và Công nghệ* **40** (số ĐB) (2002) 94–99.

[13] Vũ Hoài Chương, Nguyễn Công Điều, So sánh và kiểm định độ phân kỳ của các dãy và tập điểm tựa-ngẫu-nhiên một chiều, *Tạp chí Tin học và Điều khiển học* **23** (3) (2007) 251–259.

[14] Vũ Hoài Chương, Nguyễn Văn Hùng, Hoàng Thị Cẩm Thạch, Dãy tựa-ngẫu-nhiên van der Corput cơ số 2, *Tạp chí Nghiên cứu Khoa học Kỹ thuật và Công nghệ quân sự* (20) (9-2007) 106–113.

Nhận bài ngày 22 - 3 - 2011