Tạp chí Tin học và Điều khiển học, T.23, S.3 (2007), 251–259

SO SÁNH VÀ KIỂM ĐỊNH ĐỘ PHÂN KỲ CỦA CÁC DÃY VÀ TẬP ĐIỂM TỰA-NGẪU-NHIÊN MỘT CHIỀU

VŨ HOÀI CHƯƠNG, NGUYỄN CÔNG ĐIỀU

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

Abstract. Discrepencies of quasi-random sequences depend on bases. The choice of a suitable base has more experimental significance than theoretical. The main object of this paper is one-dimensional sequences of two simple forms: van der Corput with base p and Richtmyer $\{n.\alpha(p)\}$, where p is a prime number. Besides the case $\alpha(p) = \sqrt{p}$ is well-known, we tested following cases $\alpha(p) = \ln(p),\ \alpha(p) = sin(p),\ \alpha(p) = \cos(p),\ \alpha(p) = \arctan(p)$ and α are irrational mathematical constants. We also introduced equidistant (k/N, midpoint, V) point-sets and two pseudo-random sequences R for comparison. Criteria for ranking is a mean discrepancy computed from 21 different values N. Computational results show that modified forms generate numerical sequences with discrepancy in the extreme: very small or very big depending on a choice of concrete base. The ranking table lets us define priority using this sequences in certain quasi-Monte Carlo schemes.

Tóm tắt. Độ phân kỳ của các dãy tựa-ngẫu-nhiên phụ thuộc vào cơ số được dùng đến. Việc chọn cơ số thích hợp là vấn đề có ý nghĩa thực nghiệm hơn là lý thuyết. Đối tượng khảo sát chính trong bài này là các dãy một chiều thuộc hai dạng đơn giản: van der Corput cơ số p và Richtmyer $\{n.\alpha(p)\}$, trong đó p là số nguyên tố. Bên cạnh $\alpha(p) = \sqrt{p}$ đã được thừa nhận, chúng tôi thử kiểm định các dạng cải biên $\alpha(p) = \ln(p),\ \alpha(p) = \sin(p),\ \alpha(p) = \cos(p),\ \alpha(p) = \arctan(p)$ và α là hằng số toán học vô tỷ. Một số tập điểm cách đều (k/N, trung điểm, V) và hai dãy giả-ngẫu-nhiên R được đưa vào để so sánh. Tiêu chí xếp hạng là độ phân kỳ trung bình tính theo 21 giá trị N khác nhau. Kết quả tính toán cho thấy các dạng cải biên nói trên tạo ra những dãy số có độ phân kỳ ở hai thái cực: rất nhỏ và rất lớn, tùy thuộc vào cơ số cụ thể. Bảng xếp hạng cho phép xác định thứ tự ưu tiên sử dụng của các dãy trong một số sơ đồ tựa-Monte-Carlo.

1. MỞ ĐẦU

Phương pháp Monte Carlo có nhiều ưu thế căn bản, do đó được ứng dụng ngày càng rộng rãi. Tuy vậy phương pháp này có một vài điểm bất lợi liên quan đến bản chất ngẫu nhiên của nó, một là chỉ đưa ra được ước lượng sai số xác suất, hai là các kết quả mô phỏng thường hội tụ khá chậm.

Trong sơ đồ Monte Carlo, tính ngẫu nhiên của các số hầu như không có ý nghĩa, mà quan trọng nhất là sự phân bố đều của chúng. Từ đó nảy sinh ra ý tưởng thay thế các số ngẫu nhiên hoặc giả-ngẫu-nhiên (pseudo-random) bằng các dãy số hoàn toàn tất định nhưng có phân bố rất đều. Phương pháp *tựa-Monte-Carlo* (quasi-Monte Carlo, viết tắt là QMC) sử dụng các các dãy *tựa-ngẫu-nhiên* (quasi-random) như thế, nhờ đó có thể tìm được các ước lượng sai số xác định với tốc độ hội tụ lớn. Có thể coi phương pháp tựa-Monte-Carlo là phiên bản tất định của phương pháp Monte Carlo.

Các dãy tựa-ngẫu-nhiên còn được gọi là các dãy có độ phân kỳ thấp (low-discrepancy sequences) hoặc cận-ngẫu-nhiên (sub-random). Các dãy khác nhau đã được nhiều tác giả nghiên cứu theo các quan điểm lý thuyết và ứng dụng thực tế. Hướng thứ hai được quan tâm nhiều hơn, vì việc dùng các cơ số cụ thể nào để đạt được độ phân kỳ thấp nhất là vấn đề có ý nghĩa thực nghiệm hơn là lý thuyết. L. Finschi ([2]) đã tính toán cụ thể độ phân kỳ của một số dãy nhiều chiều như Halton, Sobol, Faure. N. V. Hùng và B. V. Thanh ([9]) so sánh các dãy tựa-ngẫu-nhiên V ([11]) với các dãy ngẫu nhiên và giả-ngẫu-nhiên.

Trong bài báo này chúng tôi so sánh 95 dãy tựa-ngẫu-nhiên (thuộc 7 nhóm khác nhau), 3 tập điểm cách đều (k/N, trung điểm, V) và 2 dãy giả-ngẫu-nhiên R do máy tính tạo ra. Đối tượng khảo sát chính là hai dạng đơn giản: van der Corput với cơ số p và Richtmyer $\{n.\alpha(p)\}$, trong đó p là số nguyên tố. Thông thường người ta dùng $\alpha(p) = \sqrt{p}$. Chúng tôi thử nghiệm thêm các trường hợp $\alpha(p) = \ln(p), \alpha(p) = \sin(p), \alpha(p) = \cos(p), \alpha(p) = \arctan(p)$ và α là hằng số toán học vô tỷ. Tính đều của 100 dãy và tập điểm một chiều nói trên được kiểm định theo tiêu chí quan trọng nhất là độ phân kỳ (discrepancy) và được xếp hạng theo giá trị trung bình tính trên 21 dãy có độ dài N khác nhau.

2. ĐỘ PHÂN KỲ

Độ phân kỳ là đặc trưng phổ biến nhất của tính đều, được nghiên cứu từ những năm 1930. Đúng hơn đó là độ đo tính không đều của phân bố, hay nói cách khác là độ lệch so với phân bố đều. Các khái niệm, định lý và cách tính độ phân kỳ được trình bày cặn kẽ trong ([2, 5 6, 8]). Dưới đây chỉ là những định nghĩa cơ bản nhất.

2.1. Các định nghĩa và định lý

Đặt $I = [0, 1]$ là đoạn thẳng đơn vị và $I^s = [0, 1] \times ... \times [0, 1]$ là siêu lập phương đơn vị s chiều. Cho một hình lập phương con $J \subset I^s$ với một đỉnh là gốc tọa độ và một dãy hữu hạn $X = \{x_k, 0 \leqslant k \leqslant N\}$ của N điểm trong I^s.

Định nghĩa 1. Độ phân kỳ cục bộ (local discrepancy) là

$$D_N(J, X) = |A(J, X)/N - V_s(J)|$$

trong đó $A(J, X)$ là số điểm $x_k \in J$ và $V_s(J)$ là số đo thể tích (độ đo Lebesgue) của J.

Định nghĩa 2. Độ phân kỳ sao (star discrepancy) hay độ phân kỳ gốc của một dãy $X = \{x_k, 0 \leqslant k \leqslant N\}$ được định nghĩa bởi hệ thức

$$D_N^*(X) = \sup_{J \in P} D_N(J, X).$$

Có thể coi độ phân kỳ sao D_N^* như sai phân tuyệt đối lớn nhất giữa xác suất đều liên tục và xác suất đều rời rạc trên tất cả các hình lập phương con của I^s có chứa gốc toạ độ.

Giữa D_N và D_N^* có các hệ thức:

$$D_N^*(X) \leqslant D_N(X) \leqslant 2^s.D_N^*(X).$$

Như vậy với s cố định, cả hai đại lượng này có cùng độ lớn. Trong trường hợp một chiều ($s = 1$), Niederreiter chứng minh được rằng

$$D_N^* \geqslant 1/2N,\ D_N \geqslant 1/N$$

và cả hai đạt được cực tiểu khi $x_k = (2k-1)/2N$. Đây là trung điểm của các đoạn $[(k-1)/N, k/N]$, và được gọi là tập các trung điểm hoặc dãy trung điểm. Chú ý rằng tập này không phải là N điểm đầu tiên của một dãy vô hạn, vì chúng không còn giá trị tương ứng với các tập điểm $(N+1), (N+2)$... phần tử. Thông thường các dãy được coi là có độ phân kỳ thấp nếu D_N^* có độ lớn bậc $O((\log N)^s/N)$. Số hạng logarit có thể bị hút vào một luỹ thừa nào đó của N, dẫn đến cận $O(1/N^{1-\epsilon})$ với mọi $\epsilon > 0$.

Định lý 1. *Dãy vô hạn $X = \{x_k, 0 \leqslant k \leqslant N\}$ phân bố đều trong $[0,1)^s$ nếu*

$$\lim D_N(\{x_k, 0 \leqslant k \leqslant N\}) = 0, \ N \to \infty$$

hoặc tương đương như thế

$$\lim D_N^*(\{x_k, 0 \leqslant k \leqslant N\}) = 0, \ N \to \infty.$$

2.2. Các ước lượng về độ phân kỳ và tốc độ hội tụ

Dãy ngẫu nhiên thực $D_N^* = O(\sqrt{\dfrac{\log\log N}{N}})$ (theo luật logarit lặp).

Dãy tựa ngẫu nhiên (độ dài vô hạn) $D_N^* = O((\log N)^s/N)$.

Tập điểm tựa ngẫu nhiên (độ dài hữu hạn) $D_N^* = O((\log N)^{s-1}/N)$.

Năm 1954 nhà toán học K. F. Roth đã xác định được cận dưới cho N điểm trong không gian s chiều

$$D_N^* \geqslant O((\log N)^{(s-1)/2}/N).$$

Trong báo cáo nghiên cứu thực nghiệm độ phân kỳ của các dãy tựa-ngẫu-nhiên nhiều chiều ([3]) L. Finschi nhận xét rằng đối với mỗi dãy có độ phân kỳ thấp $x_1, x_2, ...$ trong không gian s chiều, tồn tại hằng số C_s sao cho

$$D_N^*(x_1, ..., x_N) \leqslant C_s.(\ln N)^s/N + O((\ln N)^{s-1}/N)$$

và đã tính được các giá trị C_s đối với một số dãy với các chiều khác nhau.

2.3. Độ phân kỳ trong trường hợp một chiều

Trong trường hợp một chiều, độ phân kỳ của dãy số $x_0, x_1, ..., x_{N-1}$ được định nghĩa như sau

$$D(x_0, x_1, ..., x_{N-1}) = D = \sup_{0 \leqslant x \leqslant 1} |S_N(x)/N - x|,$$

trong đó $S_N(x)$ là số điểm thuộc khoảng $[0, x)$.

Nếu x_i đã được sắp $x_1 \leqslant x_2 \leqslant ... \leqslant x_N$ thì hệ thức trên đây được rút gọn thành công thức Niederreiter (xem [10])

$$D = 1/2 + \max_{1 \leqslant i \leqslant N} |i - \frac{1}{2} - N.x_i|.$$

Công thức này cho ta thấy cực tiểu của D là $1/2$ và đạt được khi $x_i = (i-1/2)/N$ trong đó $i = 1, 2, ..., N$ (xem [10]). Trong trường hợp N cố định, dãy số này là tối ưu theo tiêu chuẩn của độ phân kỳ. Nhưng khi chuyển từ N sang $(N+1)$ thì tất cả mọi điểm đều phải thay đổi. Ví dụ dãy số tối ưu với $N = 3$ là (1/6, 3/6 và 5/6), với $N = 4$ lại là (1/8, 3/8, 5/8, 7/8). Như vậy không thể xây dựng được một dãy vô hạn các điểm x_i sao cho đoạn $x_0, x_1, x_2, ..., x_k$ ($k = 2...N$) nào cũng phân bố đều tối ưu.

2.4. Cận Koksma-Hlawka

Độ phân kỳ không chỉ là độ đo quan trọng của phân bố đều mà còn đóng vai trò chính trong việc giới hạn sai số của xấp xỉ tích phân. Kết quả then chốt trong hướng này là bất đẳng thức Koksma-Hlawka do Jurjen Koksma công bố năm 1942 cho trường hợp một chiều và được Edmun Hlawka tổng quát hoá vào năm 1961 ([5]). Theo đó sai số tích phân được giới hạn bởi tích của hai đại lượng: một đại lượng chỉ phụ thuộc vào hàm dưới dấu tích phân f (qua độ đo biến phân toàn phần), đại lượng kia chỉ phụ thuộc vào tập điểm dùng đến (qua D_N^*, tức là độ lệch so với phân bố đều).

Dưới đây là mối quan hệ giữa các dãy tựa-ngẫu-nhiên và phép tích phân tựa Monte-Carlo.

Tích phân s chiều $I(f) = \int\limits_{I^s} f(x)dx$ được xấp xỉ bằng $Q(f) = (1/N) \sum\limits_{k=1}^{N} f(x_k)$, trong đó x_k là dãy tựa ngẫu nhiên s chiều với độ phân kỳ sao D_N^*.

Nếu $f(x)$ có biến phân toàn phần bị chặn $V(f)$ trên [0,1] thì

$$|Q(f) - I(f)| \leqslant V(f).D_N^*.$$

Đáng tiếc rằng trong thực tế rất khó ước lượng sai số từ bất đẳng thức này, và đây chính là vấn đề nan giải của phương pháp tựa-Monte-Carlo (QMC).

3. MỘT SỐ DÃY TỰA NGẪU NHIÊN – DÃY CÓ ĐỘ PHÂN KỲ THẤP

Trong toán học, dãy có độ phân kỳ thấp là dãy mà với mọi N, dãy con $(x_1, x_2, ..., x_N)$ hầu như phân bố đều và dãy $(x_1, x_2, ..., x_N, x_{N+1})$ cũng hầu như phân bố đều. Phải xây dựng tập N phần tử $(x_1, x_2, ..., x_N)$ sao cho khi xây dựng tập $(N + 1)$ phần tử thì không cần tính lại N phần tử trước.

3.1. Các dãy số van der Corput

Đây là dãy số phân bố đều trong khoảng (0,1) và được nghiên cứu khá kỹ trong lĩnh vực lý thuyết số, do nhà toán học Hà Lan Johannes Gualtherus van der Corput (1890-1975) đề xuất vào năm 1935. Dãy số này lấp đầy đoạn [0,1) một cách điều hoà với độ phân kỳ nhỏ mặc dù nó không biến thiên một cách ngẫu nhiên. Cùng với sự phát triển của máy tính, người ta dần dần thấy được khả năng ứng dụng của dãy số này trong tính toán số trị nói chung và trong phương pháp (tựa)-Monte-Carlo nói riêng.

3.1.1. *Định nghĩa cơ bản*

Gọi $p(k)$ là số van der Corput thứ k. Nếu trong hệ nhị phân $k = e_1e_2...e_{m-1}e_m$ trong đó e_i là chữ số nhị phân (tức là chỉ nhận hai giá trị 0 và 1) thì

$$p(k) = 0...e_m.e_{m-1}...e_2e_1.$$

Ví dụ, k lần lượt là 7 số tự nhiên đầu tiên, $k = 1, 2, 3, 4, 5, 6, 7$.

Dạng nhị phân của chúng là $1, 10, 11, 100, 101, 110, 111$.

Các số van der Corput tương ứng là

+ dạng nhị phân: 0,1; 0,01; 0,11; 0,001; 0,101; 0,011; 0,111

+ dạng thập phân: 1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8.

3.1.2. *Tổng quát*

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,

$$k = \sum_{j \geqslant 0} a_j(k).p^j$$

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

Hàm ngược gốc (radical inverse function) sẽ ánh xạ mỗi giá trị k 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

$$\psi_p(k) = \sum_{j \geqslant 0} a_j(k).p^{-j-1}, \ k > 0.$$

Dãy van der Corput cơ số p chính là dãy $X = \{\varphi_p(0) = 0, \varphi_p(1), \varphi_p(2), \varphi_p(3), ...\}$.

Ví dụ, tìm $\varphi_7(1205)$ tức là số hạng van der Corput cơ số 7 tương ứng với số 1205.

Trước tiên ta khai triển theo cơ số 7:

$$1205 = 3.7^3 + 3.7^2 + 4.7^1 + 1.7^0,$$

từ đây chuyển từ hệ đếm thập phân sang hệ đếm thất phân

$$1205_{(10)} = 3341_{(7)}.$$

Sau đó phản chiếu các hệ số qua dấu phẩy, $3341_{(7)} \to 0,1433_{(7)}$. Và thu được kết quả

$$0.1433_{(7)} = 1/7 + 4/7^2 + 3/7^3 + 3/7^4 = 563/2401_{(10)}.$$

Người ta đã chứng minh được rằng tất cả các dãy van der Corput là dãy có độ phân kỳ thấp. Chính xác hơn, D_N^* của N phần tử đầu tiên của dãy có độ lớn $O(\log N/N)$ với một hằng số ẩn phụ thuộc vào cơ số p.

Cơ số p càng lớn thì số điểm cần thiết để đạt được phân bố đều càng lớn. Dãy van der Corput chứa những chu kỳ có độ dài p với các số tăng đơn điệu (trong mỗi chu kỳ).

3.2. Dãy Richtmyer $\{k.\alpha(p)\}$

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 $\{k.\alpha_(p)\}$ 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.

M. Mascagni ([7]) gọi đây là động lực học ergodic (ergodic dynamics).

Người ta thường chọn α_i là căn của các số nguyên tố ([1, 5]). Trong bài này chúng tôi chọn thử nghiệm α_i là các hằng số toán học vô tỷ quen biết như π, Au (tỷ lệ vàng), e, C (hằng số Euler-Mascheroni), G (hằng số Catalan) và các hàm $\ln(p), \sin(p), \cos(p), \arctan(p)$.

3.3. Dãy số V

Có hai loại tiêu chuẩn về phân phối đều của một dãy số

+ Loại 1 dựa trên độ khác biệt của các hàm phân phối mẫu và lý thuyết. Để kiểm định tiêu chuẩn loại này, thường dùng nhất là ba phép thử χ^2, Kolmogorov và ω^2. Đối với các dãy tựa-ngẫu-nhiên còn thêm độ phân kỳ.

+ Loại 2 dựa trên các giá trị của moment mẫu và lý thuyết.

Dãy số V (xem [11]) là dãy tựa-ngẫu-nhiên dựa trên các đẳng thức giữa moment mẫu và moment lý thuyết. 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.

Mỗi dãy số V có K điểm được tạo ra theo các công thức đơn giản. 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

$$x_1 = T,\ x_2 = T + u,\ x_3 = T + 2u, ..., x_K = T + (K-1).u$$

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

Thuật toán trên đây dựa trên các giá trị của moment (tiêu chuẩn phân bố đều loại 2), nhưng các dãy số V được tạo ra cũng đáp ứng đầy đủ cả các tiêu chuẩn phân bố đều loại 1 (về độ phân kỳ của các hàm phân phối). Qua tất cả các phép thử các dãy tựa-ngẫu-nhiên V này đều vượt trội so với các dãy số ngẫu nhiên và giả-ngẫu-nhiên (xem [9]). Kết quả so sánh dãy V với các dãy tựa-ngẫu-nhiên khác sẽ được trình bày trong phần sau.

4. CÁC NGHIÊN CỨU THỰC NGHIỆM SO SÁNH

4.1. Mô tả và ký hiệu các dãy số được khảo sát

Các dãy số và tập điểm với $N = 2459$(N là số nguyên tố), được tạo thành theo các Mục 3.1, 3.2 và 3.3 trên đây với các cơ số khác nhau, và được ký hiệu trong ngoặc đơn từ (1) đến (100). Độ phân kỳ D*, tiêu chí quan trọng nhất đối với các dãy và tập điểm tựa-ngẫu-nhiên, sẽ là tiêu chuẩn so sánh.

4.1.1. *Các dãy số van der Corput*

(1) - (15): các dãy số van der Corput với cơ số là 15 số nguyên tố đầu tiên
$$\{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47\}.$$

4.1.2. *Các dãy Richtmyer* $\{k.\alpha(p)\}$

Trong các dãy (16) - (90) dưới đây, p ký hiệu 15 số nguyên tố đầu tiên.

(16) - (30): các dãy số $\{k.\alpha\}$ với $\alpha = \sqrt{p}$

(31) - (45): các dãy số $\{k.\alpha\}$ với $\alpha = \ln(p)$

(46) - (60): các dãy số $\{k.\alpha\}$ với $\alpha = \sin(p)$

(61) - (75): các dãy số $\{k.\alpha\}$ với $\alpha = \cos(p)$

(76) - (90): các dãy số $\{k.\alpha\}$ với $\alpha = \arctan(p)$.

Các dãy (91) - (95) dựa vào 5 hằng số toán học quen thuộc lấy giá trị từ [12]. Chương trình tính toán với 24 số lẻ, dưới đây chỉ ghi 12 số lẻ đầu tiên.

(91): dãy số $\{k.\alpha\}$ với $\alpha = \pi = 3,141592654358...$

(92): dãy số $\{k.\alpha\}$ với $\alpha =$ tỷ lệ vàng $=$ Au $= (\sqrt{5}+1)/2 = 1,618033988749...$

(93): dãy số $\{k.\alpha\}$ với $\alpha = e = 2,718281828459...$

(94): dãy số $\{k.\alpha\}$ với $\alpha =$ hằng số Euler-Mascheroni,
$$C = \lim(-\ln k + 1 + 1/2 + 1/3 + 1/4 + ... + 1/k) = 0,577215664901...$$

(95): dãy số $\{k.\alpha\}$ với $\alpha =$ hằng số Catalan G,
$$G = 1 - 1/3^2 - 1/5^2 - 1/7^2... = 0,915965594177...$$

4.1.3. *Các dãy số giả ngẫu nhiên R*

(96) - (97): hai 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.

4.1.4. *Các tập điểm cách đều*

(98): tập điểm V

(99): tập điểm $(k-1)/N$

(100): tập các trung điểm $(k-0,5)/N$.

Ba tập điểm cách đều (98), (99), (100) trên đây đều do 21 đoạn hợp thành sao cho số phần tử lần lượt là $N = 128$, 311, 424, 602, 808, 888, 903, 925, 1130, 1205, 1212, 1410, 1603, 1911, 1964, 2003, 2009, 2104, 2307,2404, 2459, có nghĩa là

$k_1 = 128,\ k_2 = 311 - 128 = 183,\ k_3 = 424 - 311 = 113,\ \cdots, k_{21} = 2459 - 2404 = 55.$

4.2. Kết quả so sánh thực nghiệm

Trong 2 bảng dưới đây các dãy và tập điểm mà bài báo đề xuất được ghi bằng chữ đậm. Việc so sánh được tiến hành giữa các nhóm (cùng một dạng) và giữa các dãy với nhau.

4.2.1. *So sánh các nhóm*

Bảng 1 ghi độ phân kỳ trung bình của các nhóm, xếp theo thứ tự tăng dần. Đứng đầu là các dãy Richtmyer nguyên thủy với $\alpha(p) = \sqrt{p}$. Điều đó giải thích vì sao dãy này được mọi người thừa nhận. Đứng thứ hai là các tập điểm cách đều, trong đó có dãy V do tác giả đề xuất ([11]) và dãy $(k-1/2)/N$ được đưa vào thử nghiệm lần đầu. Tiếp theo đó, 2 nhóm α (hằng số vô tỷ) và $\{k.\arctan(p)\}$ đứng trên cả nhóm van der Corput quen thuộc. Còn các nhóm $\{k.\ln(p)\}$ và $\{k.\cos(p)\}$ tuy có kém hơn nhưng vẫn ở trên nhóm giả-ngẫu-nhiên R. Các dãy $\{k.\sin(p)\}$ đứng cuối cùng với độ phân kỳ trung bình quá lớn cho nên nhóm này không chấp nhận được. Tuy vậy, dưới đây sẽ thấy, với một vài giá trị p cụ thể, dãy $\{k.\sin(p)\}$ vẫn có thể là dãy tựa-ngẫu-nhiên.

Bảng 1. Độ phân kỳ trung bình theo nhóm

Thứ tự	D^* trung bình	dãy	số dãy trong nhóm	Nhóm
1	3,19723	(16) - (30)	15	dãy Richtmyer $\{k.\sqrt{p}\}$
2	5,21334	(98) - (100)	3	tập điểm $(k-1)/N$, $\boldsymbol{(k-0,5)/N, V}$
3	6,92485	(91) - (95)	5	dãy Richtmyer cải biên theo hằng số $\boldsymbol{\{k.\pi\}\{k.Au\}\{k.e\},\{k.C\},\{k.G\}},$
4	7,83201	(76) - (90)	15	dãy $\boldsymbol{\{k.\arctan(p)\}}$
5	9,05301	(1) - (15)	15	dãy van der Corput
6	10,76932	(31) - (45)	15	dãy $\boldsymbol{\{k.\ln(p)\}}$
7	10,79202	(61) - (75)	15	dãy $\boldsymbol{\{k.\cos(p)\}}$
8	33,04532	(96) - (97)	2	dãy giả ngẫu nhiên R
9	96,35071	(46) - (60)	15	dãy $\boldsymbol{\{k.\sin(p)\}}$

4.2.2. *So sánh các dãy và tập điểm*

Độ phân kỳ trung bình (bảng 2) được tính theo các giá trị D^* của dãy có độ dài tăng dần: $N = 128$, 311, 424, 602, 808, 888, 903, 925, 1130, 1205, 1212, 1410, 1603, 1911, 1964, 2003, 2009, 2104, 2307,2404, 2459.

Trong 10 dãy đứng đầu có tới 7 dãy do chúng tôi đề xuất. Đặc biệt dãy $\{k.Au\}$, lấy tỷ lệ vàng làm cơ số, có độ phân kỳ trung bình nhỏ nhất. Trong 10 dãy cuối bảng cũng có 8 dãy do chúng tôi thử nghiệm. Điều đó chứng tỏ các dạng cải biên chỉ được chấp nhận với những giá trị p nhất định. Trong hai tốp dẫn đầu và cuối này không có các tập điểm cách đều và các dãy van der Corput.

Các dãy $\{k.\ln(p)\}, \{k.\sin(p)\}\{k.\arctan(p)\}$ xuất hiện ở cả tốp trên và tốp dưới, tùy thuộc vào các giá trị p khác nhau.

Bảng 2. Mười dãy có độ phân kỳ D^* trung bình nhỏ nhất, lớn nhất và thứ tự của chúng tại 3 giá trị N

Thứ tự chung	$D^*_{tr.b.}(i)$	Dãy	Dạng	Thứ tự tại 3 giá trị N		
				$N = 311$	$N = 1205$	$N = 2459$
1.	1,91236	(92)	$\boldsymbol{\{k.Au\}}$	6.	7.	9.
2.	2,08522	(16)	$\{k.\sqrt{2}\}$	19.	14.	7.
3.	2,31358	(90)	$\boldsymbol{\{k.\arctan 47\}}$	4.	8.	14.
4.	2,31592	(85)	$\boldsymbol{\{k.\arctan 29\}}$	15.	9.	10.
5.	2,42076	(17)	$\{k.\sqrt{3}\}$	26.	11.	11.
6.	2,43589	(36)	$\boldsymbol{\{k.\ln 13\}}$	16.	26.	12.
7.	2,45736	(89)	$\boldsymbol{\{k.\arctan 43\}}$	29.	23.	20.
8.	2,46239	(40)	$\boldsymbol{\{k.\ln 29\}}$	20.	21.	23.
9.	2,48654	(57)	$\boldsymbol{\{k.\sin 37\}}$	12.	15.	30.
10.	2,50865	(18)	$\{k.\sqrt{5}\}$	18.	6.	22.
...	...	...	...	...	...	...
91.	15,50845	(68)	$\boldsymbol{\{k.\cos 19\}}$	96.	91.	81.
92.	23,36515	(75)	$\boldsymbol{\{k.\cos 47\}}$	97.	93.	85.
93.	26,16343	(73)	$\boldsymbol{\{k.\cos 41\}}$	73.	94.	91.
94.	32,97025	(96)	$R1$	78.	97.	95.
95.	33,12040	(97)	$R2$	95.	90.	97.
96.	36,50603	(76)	$\boldsymbol{\{k.\arctan 2\}}$	86.	96.	98.
97.	39,16400	(65)	$\boldsymbol{\{k.\cos 11\}}$	99.	98.	93.
98.	44,71712	(37)	$\boldsymbol{\{k.\ln 17\}}$	98.	95.	96.
99.	65,70675	(38)	$\boldsymbol{\{k.\ln 19\}}$	94.	99.	99.
100.	1343,98573	(50)	$\boldsymbol{\{k.\sin 11\}}$	100.	100.	100.

4.2.3. *Kết luận*

Các tính toán kiểm định cho ta một cái nhìn bao quát về ý nghĩa thực tế của các dãy tựa-ngẫu-nhiên. Chẳng hạn như về cơ sở lý thuyết, dãy van der Corput đẹp hơn hẳn dãy Richtmyer. Nhưng khả năng ứng dụng thì hoàn toàn ngược lại.

Các dạng cải biên do chúng tôi thử nghiệm chưa đem lại một dạng mới tổng quát cho các dãy tựa-ngẫu-nhiên, nhưng đã đưa ra được một số dãy mới (7 dãy in đậm trong phần đầu Bảng 2) có độ phân kỳ thấp hơn so với các tập điểm cách đều và các dãy van der Corput. Có thể sử dụng hữu hiệu các dãy này để tăng tốc độ hội tụ trong các tính toán theo sơ đồ tựa-Monte-Carlo. Mặt khác có thể kết hợp các dãy này với các dãy thuộc dạng khác để tạo ra các dãy tựa-ngẫu-nhiên nhiều chiều có tương quan yếu giữa các thành phần.

TÀI LIỆU THAM KHẢO

[1] I. Dupain, T. V. Sós, On the discrepancy of $\{n.\alpha\}$ sequences, *Topics in Classical Number Theory* Vol. 1, North-Holland (1981) 355–387.

[2] S. R. Finch, Discrepancy and Uniformly, Cambridge Univer. Press, 2004.

[3] L. Finschi, Quasi-Monte Carlo: An empirical study on low-discrepancy sequences, *Technical Report, Institute For Operations Research (IFOR)*, ETH, Zurich 1996.

[4] Frey Tamás - Szelezsán János, *Számítástechnika*, Akadémiai Kiadó, Budapest 1973.

[5] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York, 2000.

[6] Hua Loo Keng - Wang Yuan, *Applications of Number Theory to Numerical Analysis*, Springer-Verlag, Berlin-Heidelberg and Science Press, Beijing, 1981.

[7] M. Mascagni, "Quasi-Monte Carlo Methods: Where Randomness and Determinism collide", http://www.cs.fsu.edu/ ∼mascagni

[8] W. J. Morokoff, R. E. Caflisch, Quasi-random sequences and their discrepancies, *SIAM Journal on Scientific Computing* **15** (1994) 1251–1279.

[9] Nguyễn Văn Hùng, Bùi Văn Thanh, Nghiên cứu thực nghiệm tính phân bố đều của các dãy số ngẫu nhiên, giả ngẫu nhiên và tựa ngẫu nhiên, *Tạp chí Tin học và Điều khiển học* **22** (1) (2006) 53–62.

[10] I. M. Sobol, *Các điểm lấp đầy khối lập phương nhiều chiều*, Nhà xuất bản Znanie, Moskva, 1985. (tiếng Nga).

[11] 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.

[12] WolframMathWorld, Constants, http://mathwordl.wolfram.com/topics/constants.html.

Nhận bài ngày 12 - 12 - 2006