Tạp chí Tin học và Điều khiển học, T. 19, S. 1 (2003), 72–77

VỀ TIÊU CHUẨN ĐẠO HÀM NHỊ PHÂN KIỂM TRA TÍNH NGẪU NHIÊN CỦA MỘT DÃY NHỊ PHÂN HỮU HẠN

NGUYỄN THỊ HẢI YẾN

Số 162, ngõ 205, đường Giải phóng, Hai Bà Trưng, Hà Nội

Abstract. In this paper, we would then present a demonstrate soundness of the standard and word out the throat value in order to test randomness of finite binary sequences in various cases.

Tóm tắt. Trong bài này, chúng tôi đưa ra phương pháp xác định ngưỡng của tiêu chuẩn độ phức tạp đạo hàm nhị phân để kiểm tra tính ngẫu nhiên của dãy nhị phân trong nhiều trường hợp khác nhau và đồng thời chứng minh tính đúng đắn của tiêu chuẩn này.

1. GIỚI THIỆU

Việc kiểm tra tính ngẫu nhiên của dãy nhị phân đã được quan tâm nhiều (chẳng hạn như năm tiêu chuẩn cơ bản [2]), test sử dụng tiêu chuẩn độ phức tạp Ziv Lempet [3],...). Đặc biệt trong [1] các tác giả đưa ra tiêu chuẩn độ phức tạp đạo hàm nhị phân và vận dụng chúng để kiểm tra khóa yếu nhưng đã không đưa ra các cơ sở khoa học của tiêu chuẩn. Trong bài này, chúng tôi sẽ chứng minh tính đứng đắn của các tiêu chuẩn này và tính toán giá trị ngưỡng để kiểm tra tính ngẫu nhiên của một dãy nhị phân.

2. ĐỘ PHỨC TẠP ĐẠO HÀM CỦA DÃY NHỊ PHÂN

Định nghĩa 2.1. Giả sử ta có dãy $s^0 = s_1^0, s_2^0 ... s_n^0 \;\; s_i^0 \in \{0,\ 1\}, \;\; i = 1, 2, ..., n$. Ta lập dãy mới $s^1 = s_1^1 s_2^1 ... s_n^1$ như sau:

$$s_1^1 = s_1^0 \oplus s_2^0, \quad s_2^1 = s_2^0 \oplus s_3^0, ..., s_{n-1}^1 = s_{n-1}^0 \oplus s_n^0,$$

trong đó $\oplus$ là phép cộng loại trừ bít.

Từ dãy s^1 ta lại tạo ra dãy s^2 theo cách tương tự và cứ như thế cho đến khi được dãy $s^m,\ m < n$. Sau đó tiến hành tính:

$$p_k = \sum_{i=1}^{n-k} \frac{s_i^k}{n-k} \;\; ; \;\; k = 0, 1, 2, ..., m.$$

Độ phức tạp đạo hàm nhị phân của dãy s^0 là đại lượng $r = p_{\max} - p_{\min}$, trong đó $p_{\max} = \max\{p_0, ..., p_m\}$; $\;p_{\min} = \min\{p_0, ..., p_m\}$.

Sau đây là ví dụ tìm độ phức tạp đạo hàm nhị phân của một dãy có quy luật (lặp lại):

Ví dụ 1. Dãy có quy luật

Dãy nhị phân	**Giá trị p**
S(0) = 10001000100010001000100010001000100011	0,2857142
S(1) = 10011001100110011001100110011001100110010	0,4878048
S(2) = 10101010101010101010101010101010101011	0,525
S(3) = 11111111111111111111111111111111111110	0,9743589
S(4) = 00000000000000000000000000000000000001	0,0263157
S(5) = 00000000000000000000000000000000000001	0,027027

$$r = p_{\max} - p_{\min} = p_3 - p_4 = 0,9743589 - 0,0263157 = 0,9480432.$$

Ví dụ 2. Đối với độ phức tạp đạo hàm nhị phân của một dãy ngẫu nhiên:

Dãy ngẫu nhiên

Dãy nhị phân	**Giá trị p**
S(0) = 01000111011001110010101111000001100010110 1	0,5
S(1) = 11001001101010010111110001000010100111011	0,5121951
S(2) = 01011010111110111000010011000111101001 10	0,55
S(3) = 11101111000011001000110101001000111010 1	0,5128205
S(4) = 00110001000101011001011111011001001111	0,5263158
S(5) = 01010011001111101011100001101011010 00	0,5135135

$$r = p_{\max} - p_{\min} = p_2 - p_0 = 0,55 - 0,5 = 0,05.$$

Dãy ở Ví dụ 1 có tính quy luật rõ ràng và r tiến tới gần giá trị 1, trong khi đó dãy ở Ví dụ 2 "ngẫu nhiên" hơn và có r tiến tới gần giá trị 0. Ta nhận thấy rằng dãy có giá trị r càng gần giá trị 0 thì dãy đó càng ngẫu nhiên hơn.

3. TIÊU CHUẨN ĐỘ PHỨC TẠP ĐẠO HÀM NHỊ PHÂN

3.1. Xác định ngưỡng

Độ phức tạp đạo hàm nhị phân của dãy s^0 có tính chất là: r nhận giá trị lớn nếu dãy s^0 có "quy luật rõ rệt", và nhận giá trị nhỏ nếu s^0 là dãy ngẫu nhiên thật sự. Ta cần tìm ngưỡng để phân biệt dãy nào là ngẫu nhiên thực sự và dãy nào là có quy luật.

Trong [1] các tác giả đã giới thiệu khái niệm độ phức tạp đạo hàm nhị phân và vận dụng để kiểm tra khóa yếu. Tuy nhiên, các tác giả này đã không dẫn giải việc xác định ngưỡng của tiêu chuẩn trong các trường hợp khác nhau.

Mệnh đề 1. *Nếu $a_1, a_2, ..., a_n$ là các biến ngẫu nhiên độc lập, có cùng phân phối [2], nhận giá trị trong tập $\{0, 1\}$:*

$$P\{a_k = 0\} = 1/2, \quad k = 1, 2, ..., n$$

thì $b_1 = a_1 \oplus a_2$; $b_2 = a_2 \oplus a_3$; ...; $b_{n-1} = a_{n-1} \oplus a_n$ cũng độc lập, có cùng phân phối.

$$P\{b_k = 0\} = 1/2, \quad k = 1, 2, ..., n-1.$$

Chứng minh. Trước hết, xét

$$\begin{aligned}
P(b_1 = 0) &= P\{a_1 \oplus a_2 = 0\} \\
&= P\{a_1 = a_2\} \\
&= P\{[(a_1 = 0) \text{ và } (a_2 = 0)] \text{ hoặc } [(a_1 = 1) \text{ và } (a_2 = 1)]\} \\
&= P\{\sum_{i=0}^{1}[(a_1 = i) \text{ và } (a_2 = i)]\} \\
&= \sum_{i=0}^{1}[P\{(a_1 = i) \text{ và } (a_2 = i)\}] \\
&= \sum_{i=0}^{1}[P\{a_1 = i\}P\{a_2 = i\}] \\
&= \sum_{i=0}^{1}[1/2 \times 1/2] \\
&= 0,5.
\end{aligned}$$

Kết quả này đúng cho mọi $b_i,\ i = 1, 2, ..., n$. Bây giờ ta chứng minh rằng b_1, b_2 độc lập. Xét:

$$\begin{aligned}
P\{b_1 = 0,\ b_2 = 0\} &= P\{(a_1 \oplus a_2 = 0) \text{ và } (a_2 \oplus a_3 = 0)\} \\
&= P\{(a_1 = a_2) \text{ và } (a_2 = a_3)\} \\
&= P\{(a_1 = 0 \text{ và } a_2 = 0) \text{ và } (a_2 = 0 \text{ và } a_3 = 0)\} \text{ hoặc} \\
&\quad P\{(a_1 = 0 \text{ và } a_2 = 0) \text{ và } (a_2 = 1 \text{ và } a_3 = 1)\} \text{ hoặc} \\
&\quad P\{(a_1 = 1 \text{ và } a_2 = 1) \text{ và } (a_2 = 0 \text{ và } a_3 = 0)\} \text{ hoặc} \\
&\quad P\{(a_1 = 1 \text{ và } a_2 = 1) \text{ và } (a_2 = 1 \text{ và } a_3 = 1)\} \\
&= P\{\sum_{i=0}^{1}[(a_1 = i) \text{ và } (a_2 = i)] \text{ và } \sum_{j=0}^{1}[(a_2 = j) \text{ và } (a_3 = j)]\} \\
&= P\{\sum_{i=0}^{1}\sum_{j=0}^{1}[(a_1 = i \text{ và } a_2 = i) \text{ và } (a_2 = j \text{ và } a_3 = j)]\} \\
&= \sum_{i=0}^{1}\sum_{j=0}^{1}[P\{(a_1 = i \text{ và } a_2 = i) \text{ và } (a_2 = j \text{ và } a_3 = j)\}] \\
&= \sum_{i=0}^{1}\sum_{j=0}^{1}[P\{(a_1 = i \text{ và } a_2 = i)\}P\{(a_2 = j \text{ và } a_3 = j)\}] \\
&= \sum_{i=0}^{1}\sum_{j=0}^{1}[P\{a_1 = i\}P\{a_2 = i)\}P\{(a_2 = j\}P\{a_3 = j)\}] \\
&= \sum_{i=0}^{1}\sum_{j=0}^{1}(1/2 \times 1/2 \times 1/2 \times 1/2) \\
&= 0,25
\end{aligned}$$

$$\begin{aligned}
P\{b_1 = 0,\ b_2 = 1\} =& P\{(a_1 \oplus a_2 = 0) \text{ và } (a_2 \oplus a_3 = 1)\} \\
&= P\{(a_1 = a_2) \text{ và } (a_2 \neq a_3)\} \\
&= P\{[(a_1 = 0) \text{ và } (a_2 = 0)] \text{ hoặc } [(a_1 = 1) \text{ và } (a_2 = 1)] \text{ và} \\
&\quad [(a_2 = 0) \text{ và } (a_3 = 1)] \text{ hoặc } [(a_2 = 1) \text{ và } (a_3 = 0)]\} \\
&= P\{[(a_1 = 0) \text{ và } (a_2 = 0) \text{ và } (a_3 = 1)] \text{ hoặc} \\
&\quad [(a_1 = 1) \text{ và } (a_2 = 1)] \text{ và } (a_3 = 0)]\} \\
&= 1/2 \times 1/2 \times 1/2 + 1/2 \times 1/2 \times 1/2 \\
&= 0,25
\end{aligned}$$

Tương tự, ta có:

$$P\{b_1 = 1,\ b_2 = 0\} = 0,25; \quad P\{b_1 = 1,\ b_2 = 1\} = 0,25.$$

Tóm lại $P = \{b_1 = \xi,\ b_2 = \eta\} = 0,25$ với mọi $\xi, \eta \in \{0, 1\}$ nghĩa là b_1, b_2 độc lập. Có thể mở rộng phương pháp trên bằng quy lạp để chứng minh $n-1$ biến ngẫu nhiên $b_1, b_2, ..., b_n$ là độc lập trong toàn bộ.

Mệnh đề 2. Giả sử $s^0 = s_1^0 s_2^0 ... s_n^0$, trong đó $s_i^0,\ \ i = 1, 2, ..., n$ là các biến ngẫu nhiên độc lập, cùng phân phối đều, nhận giá trị trong tập $\{0, 1\}$. Khi đó với xác suất $(1-\alpha)^m$ ta có:

$$|P_{\max} - P_{\min}| \leq \frac{u(\alpha/2)}{\sqrt{n-m}}. \qquad (*)$$

Chứng minh. Ta xét trường hợp sau đây: $a_1, a_2, ..., a_n;\ b_1, b_2, ..., b_k$ là hai dãy trong đó mỗi dãy gồm các biến ngẫu nhiên độc lập, phân phối đều và hai dãy có độc lập với nhau.

Gọi x là số các số 1 trong dãy $a_1, a_2, ..., a_n,\ x = \sum_{i=1}^{n} a_i$, tần suất $P_0 = (1/n)\sum_{i=1}^{n} a_i = x/n$.

Theo định lý giới hạn trung tâm dạng tích phân [4], ta có:

$$\begin{aligned}
&P\{|p_0 - p| < \varepsilon\} = P\Big\{\Big|\frac{x}{n} - p\Big| < \varepsilon\Big\} = P\Big\{-\varepsilon < \frac{x}{n} - p < \varepsilon\Big\} \\
&= P\Big\{-\varepsilon < \frac{x - np}{n} < \varepsilon\Big\} = P\Big\{\varepsilon\sqrt{\frac{n}{pq}} < \frac{x - np}{\sqrt{npq}} < \varepsilon\sqrt{\frac{n}{pq}}\Big\} \\
&= \varnothing\varepsilon\sqrt{\frac{n}{pq}} - \varnothing\Big(-\varepsilon\sqrt{\frac{n}{pq}}\Big) = \varnothing\varepsilon\sqrt{\frac{n}{pq}} - \Big[1 - \varnothing\varepsilon\sqrt{\frac{n}{pq}}\Big] \\
&= 2\varnothing\varepsilon\sqrt{\frac{n}{pq}} - 1 = (1-\alpha) \\
\Longrightarrow\ &\varnothing\varepsilon\sqrt{\frac{n}{pq}} = 1 - \frac{\alpha}{2} \ \Longrightarrow\ \varepsilon\sqrt{\frac{n}{pq}} = u(\alpha/2) \ \Longrightarrow\ \varepsilon = u(\alpha/2)\sqrt{\frac{pq}{n}}.
\end{aligned}$$

Vậy với xác suất $(1-\alpha)$ trong đó $p = 1/2$, ta có:

$$\Big|\frac{1}{n}\sum_{i=1}^{n} a_i - \frac{1}{2}\Big| \leq u(\alpha/2)\sqrt{\frac{pq}{n}} = \frac{u(\alpha/2)}{2\sqrt{n}}.$$

Tương tự, với xác suất $(1-\alpha)$ ta có:

$$\Big|\frac{1}{k}\sum_{i=1}^{k} b_i - \frac{1}{2}\Big| \leq u(\alpha/2)\sqrt{\frac{pq}{k}} = \frac{u(\alpha/2)}{2\sqrt{k}}.$$

Kết hợp hai bất đẳng thức trên đây, với xác suất $(1-\alpha)^2$, ta được:

$$\Big|\frac{1}{n}\sum_{i=1}^{n}a_i-\frac{1}{k}\sum_{i=1}^{k}b_i\Big|\le\Big|\frac{1}{n}\sum_{i=1}^{n}a_i-\frac{1}{2}\Big|+\Big|\frac{1}{k}\sum_{i=1}^{k}b_i-\frac{1}{2}\Big|\le$$
$$\le\frac{u(\alpha/2)}{2}\Big(\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{k}}\Big)\le u(\alpha/2)\max\Big\{\frac{1}{\sqrt{n}};\ \frac{1}{\sqrt{k}}\Big\}$$

Vậy với xác suất $(1-\alpha)^m$ ta có:

$$\Big|\frac{1}{n-j}\sum_{i=1}^{n-j}a_i-\frac{1}{n-k}\sum_{i=1}^{n-k}a_i\Big|\le u(\alpha/2)\Big(\frac{1}{\sqrt{n-m}}\Big)\ ;\ j,k=0,1,...,m.$$

Do đó

$$|p_{\max}-p_{\min}|\le\frac{u(\alpha/2)}{\sqrt{n-m}}).$$

■

Ví dụ 3. Ta xét một vài giá trị cụ thể của α như sau: $\alpha=0,05,\ m=6$ thì $(1-\alpha)^6\approx 0,74$; $\alpha=0,025$ thì $(1-\alpha)^6\approx 0,86$; $\alpha=0,01$ thì $(1-\alpha)^6\approx 0,94$.

Với $n=500$ và $m=6,\alpha=0,01$ theo công thức (*) ta có:

$$\frac{u(\alpha/2)}{\sqrt{n-m+1}}\approx 0,116.$$

3.2. Kết quả tính toán

Sau đây là giá trị $u(\alpha/2)/\sqrt{n-m}$ của độ phức tạp đạo hàm nhị phân trong một số trường hợp đặc biệt: $m=6$, xác suất $(1-\alpha)^6$.

α / Độ dài n	0,02	0,15	0,1	0,05	0,025	0,01
256	0,082	0,091	0,104	0,124	0,142	0,063
500	0,058	0,065	0,074	0,088	0,101	0,116
1000	0,041	0,046	0,052	0,062	0,071	0,082
1500	0,033	0,037	0,043	0,051	0,058	0,067
2000	0,029	0,032	0,037	0,044	0,050	0,058
2500	0,026	0,029	0,033	0,039	0,045	0,052
3000	0,024	0,026	0,030	0,036	0,041	0,047
3500	0,022	0,024	0,028	0,033	0,038	0,044
4000	0,020	0,023	0,026	0,031	0,035	0,041
4500	0,019	0,021	0,025	0,029	0,033	0,038
5000	0,018	0,020	0,023	0,028	0,032	0,037
10000	0,013	0,014	0,017	0,020	0,022	0,026
50000	0,006	0,006	0,007	0,009	0,010	0,012
100000	0,004	0,005	0,005	0,006	0,007	0,008

3.3. Tiêu chuẩn độ phức tạp đạo hàm nhị phân

Theo mục 3.1 với giả thiết dãy nhị phân là các biến ngẫu nhiên độc lập, cùng phân phối đều trên tập $\{0,1\}$, khi $m=6$ với xác suất $(1-\alpha)^6$, độ phức tạp đạo hàm nhị phân r của dãy

bé hơn vế phải của (*). Vì vậy ta sẽ lấy vế phải của (*) để xác định ngưỡng của tiêu chuẩn đạo hàm nhị phân.

Rõ ràng là khi $n \geq 50.000$ bít thì ngưỡng của độ phức tạp đạo hàm nhị phân xấp xỉ dưới 0,01. Các kết quả (không chứng minh) trong [1] phù hợp với kết quả ở đây trong trường hợp mức ý nghĩa α=0,1 (do đó $(1-\alpha)^6 \approx$0,53). Với các ví dụ ở mục 2.2 và căn cứ vào bảng tiêu chuẩn ở mục 3.2, ta thấy với dãy đầu tiên hoàn toàn bị loại bỏ vì r quá lớn so với các giá trị trong bảng, còn ở dãy thứ hai $n \leq 256$ và r=0,05 thực sự nhỏ hơn giá trị 0,104 trong bảng, Như vậy dãy này được chấp nhận.

Chú ý: Nên kiểm tra cả dãy bù của dãy ban đầu theo tiêu chuẩn này (dãy bù là dãy thu được bằng cách thay mỗi phần tử của dãy ban đầu bằng một phần bù của nó).

4. KẾT LUẬN

Để kiểm tra tính ngẫu nhiên của một dãy nhị phân, người ta đã đưa ra tiêu chuẩn để xác định độ phức tạp đạo hàm nhị phân. Phương pháp này có ưu điểm là tính toán đơn giản nhưng rất có hiệu quả, có thể ứng dụng để kiểm tra tính ngẫu nhiên của dãy nhị phân.

TÀI LIỆU THAM KHẢO

[1] John M. Carroll and Sri Nurdiati, Weak keys and Weak data foiling the two nemeses, *Cryptolagia,* **XVIII** (3) (1994).

[2] Benezes A., Van Oorschot P.C., Van stole S., *Handbook of Applied Cryptograpfy*, CRC Press, 1997.

[3] Nguyễn Thị Hải Yến, Một phương pháp kiểm tra tính ngẫu nhiên của dãy nhị phân, *Tạp chí Tin học và Điều khiển học,* **18** (2) (2002)

[4] Đào Hữu Hồ, *Xác xuất thống kê*, NXB Đại học Quốc gia Hà Nội, 1999.

Nhận bài ngày 10 - 1 - 2002

Nhận lại sau khi sửa 15 - 3 - 2002