УДК 519.21

С.Я. Гончарова, доц., канд. фіз.-мат. наук
Кіровоградський національний технічний університет

Дослідження задачі про розладнання бінарної послідовності та її застосування у схемах захисту інформації

Досліджується задача про розладнання та її застосування в схемах захисту інформації. Побудований алгоритм для визначення моменту «переключення» ймовірностей для бінарної послідовності.
момент розладнання, бінарна послідовність

Вступ. Двійкові випадкові і псевдовипадкові послідовності широко використовуються в різних галузях науки і техніки: системах керування, контролю, захисту інформації тощо. На ймовірнісні характеристики таких послідовностей накладаються, як правило, жорсткі вимоги; особливо це стосується випадкових послідовностей, що використовуються в сучасних системах захисту інформації. Залишається актуальною задача побудови ефективних і зручних у реалізації алгоритмів визначення моментів зміни ймовірнісних характеристик різних типів випадкових послідовностей. У класичній постановці задачі про розладнання [1] існує не більше одного випадкового моменту зміни ймовірнісних характеристик випадкового процесу, для якого потрібно побудувати оцінки. При цьому часто використовується байєсівський підхід, що вимагає знань апріорного ймовірнісного розподілу моменту появи розладнання і призводить нерідко до трудомістких обчислень.

Задача про розладнання полягає у знаходженні моменту зміни ймовірностей потоку бітів. Джерело передає повідомлення з певною ймовірністю, яка в деякий момент часу змінюється. Потрібно визначити момент зміни ймовірностей. Цей момент називають моментом розладнання.

1. Постановка задачі та обґрунтування алгоритму її розв'язання. Дослідимо задачу про розладнання для потоку нулів та одиниць. Джерело повідомлень передає нулі й одиниці із заданою ймовірністю, у деякий момент часу, ймовірність передачі повідомлення змінюється. Потрібно визначити момент розладнання, тобто момент «переключення» ймовірностей.

Нехай до моменту «переключення» нулі надходять з імовірністю p_0, а одиниці – з ймовірністю $1-p_0$, після «переключення» ці ймовірності змінюються на p_1 і $1-p_1$ відповідно.

Відомо, що якщо при великих N відносна частота $\nu_N(A)$ появи події A мало відрізняється від деякого фіксованого значення p, то подія A стохастично стійка, а число p є ймовірністю події A.

Природно, враховуючи частотну інтерпретацію ймовірності, прийняти в якості оцінки ймовірності, тобто її наближеного значення, частоту настання події A в проведених експериментах. При цьому частота $\nu_N(A)$ є незміщеною і обґрунтованою оцінкою ймовірності, так як її математичне сподівання як випадкової величини рівне істинному значенню параметра і при $N\to\infty$ $\nu_N(A)$ збігається в середньоквадратичному до ймовірності.

Розглядаємо послідовність нулів та одиниць довжиною M. Для визначення моменту

«переключення» підраховуємо кількість нулів K на інтервалі від 0 до N, $N < M$, потім підраховуємо кількість нулів на інтервалах від 1 до N+1, від 2 до N+2, …, від $M - N$ до M.

Спостережною подією A в серії цих стохастичних експериментів є поява нуля. Частота $v_N(A)$ появи події A рівна

$$v_N(A) = k_N(A)/N,$$

де N – кількість експериментів,

$k_N(A)$ – кількість експериментів, в яких подія A відбулась.

Так як до моменту «переключення» нулі надходять з імовірністю p_0, то частота появи нулів $v_N(A)$ повинна зберігати майже стале значення, тобто $v_N(A) \approx p_0$. Тоді кількість нулів K на інтервалах довжиною N теж повинна зберігати майже стале значення:

$$K = v_N(A)N \approx p_0 N.$$

Частота $v_N(A)$ зміниться при «переключенні» ймовірності, тобто $v_N(A) \approx p_1$, а отже, зміниться і кількість нулів K на інтервалах довжиною N: $K \approx p_1 N$.

Саме ця властивість стійкості відносної частоти появи випадкової події в серії експериментів, що дає можливість розглядати емпіричну частоту як ймовірність, і лежить в основі алгоритму розв'язання задачі про розладнання.

При достатньо великому N і ймовірності p, не дуже близькій до нуля та одиниці, можна вважати, що відносна частота розподілена приблизно нормально, тому розподіл статистики K близький до нормального з параметрами Np, Npq, де $q = 1 - p$. Для малих вибірок ($N \le 30$) заміна розподілу нормальним приводить до грубих помилок.

Критичну область обираємо у вигляді $U_{\text{кр}} = \{K > C\}$.

При заданій помилці першого роду α для визначення порогу C використаємо інтегральну теорему Лапласа:

$$P\left(0 < \frac{K - Np_1}{\sqrt{Np_1q_1}} < \frac{C - Np_1}{\sqrt{Np_1q_1}}\right) = \Phi\left(\frac{C - Np_1}{\sqrt{Np_1q_1}}\right) = 1 - \alpha, \quad (1)$$

де p_1 – імовірність надходження нулів після «переключення»;

$q_1 = 1 - p_1$ – імовірність надходження одиниць після «переключення»;

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ – функція Лапласа.

Із співвідношення (1) отримаємо:

$$C = Np_1 + t_{1-\alpha}\sqrt{Np_1q_1}, \quad (2)$$

де $t_{1-\alpha}$ – квантиль нормального розподілу,

$\Phi(t_{1-\alpha}) = 1 - \alpha$.

Аналогічно при заданій помилці другого роду β за теоремою Лапласа

$$P\left(0 < \frac{K - Np_0}{\sqrt{Np_0q_0}} < \frac{C - Np_0}{\sqrt{Np_0q_0}}\right) = \Phi\left(\frac{C - Np_0}{\sqrt{Np_0q_0}}\right) = \beta.$$

Відповідний даному N і C, визначеному в (2), квантиль $t_{1-\beta}$ ймовірності помилки другого роду β рівний

$$t_{1-\beta} = -t_\beta = \frac{Np_0 - C}{\sqrt{Np_0q_0}}. \quad (3)$$

Із (3) знаходимо

$$Np_0 - C = t_{1-\beta}\sqrt{Np_0q_0}. \quad (4)$$

Виражаючи C із (2) і підставляючи в (4), отримаємо

$$N(p_0 - p_1) = t_{1-\beta}\sqrt{Np_0q_0} + t_{1-\alpha}\sqrt{Np_1q_1},$$

звідки довжина найменшого інтервалу, на якому підраховуємо кількість нулів, буде рівною

$$N = \left(\frac{t_{1-\beta}\sqrt{p_0q_0} + t_{1-\alpha}\sqrt{p_1q_1}}{p_0 - p_1}\right)^2. \qquad (5)$$

2. Розв'язання задачі. Приклади. Для розв'язання задачі застосовуємо програму. Програма зчитує послідовність із файлу data.txt. Потім підраховує кількість нулів на інтервалах від 0 до N, від 1 до N+1, від 2 до N+2 і т.д. і виводить у файл.

За отриманими даними можна побудувати графік залежності частоти появи нулів від позиції в Excel або MathCAD.

Приклад 1. Нехай $p_0 = 0{,}8$; $p_1 = 0{,}4$; $\alpha = 0{,}01$; $\beta = 0{,}001$.

За формулою (5) знайдемо мінімальну довжину спостережного інтервалу: $N = 44$.

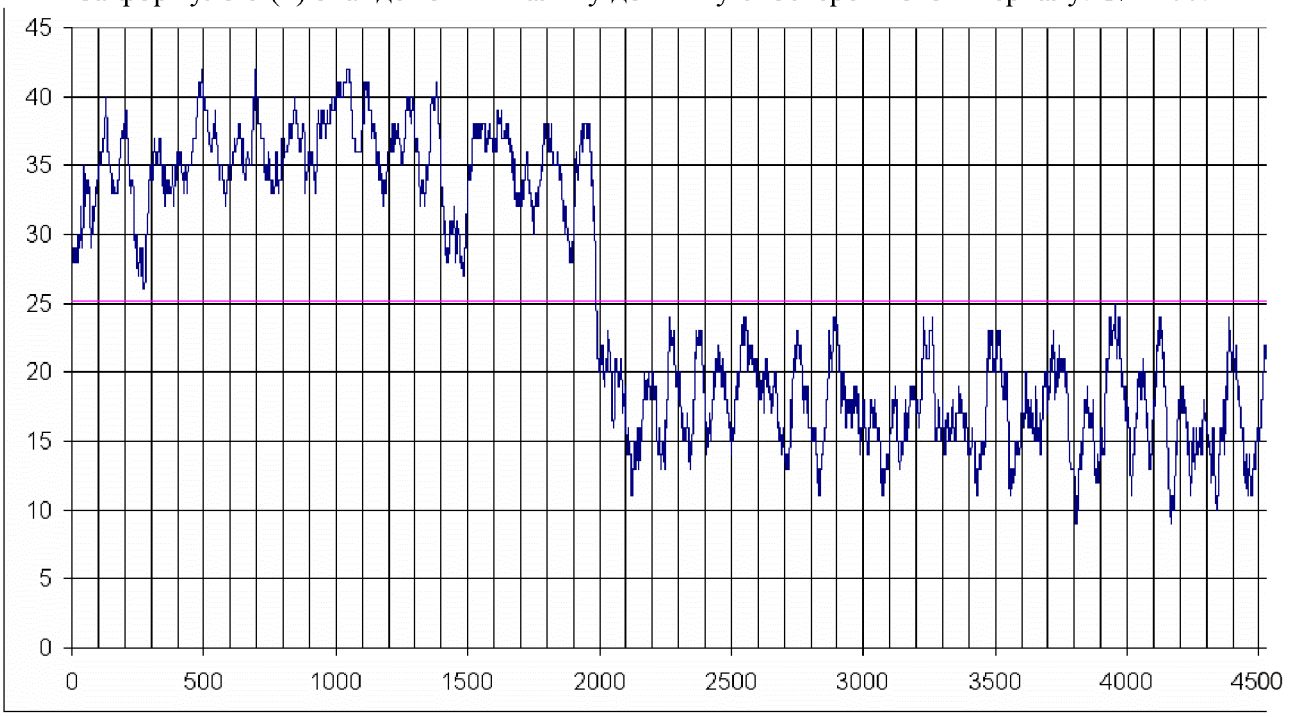

Рисунок 1- Графік, який показує кількість K нулів у кожний момент часу. Лінією позначено поріг C

Далі обчислюємо поріг C за формулою (2):

$$C = Np_1 + t_{1-\alpha}\sqrt{Np_1q_1} = 44 \cdot 0{,}4 + t_{0{,}99} \cdot \sqrt{44 \cdot 0{,}4 \cdot 0{,}6} \approx 25{,}1 \quad (t_{0{,}99} = 2{,}326).$$

Отже, для значень $K < 26$ можна вважати з імовірністю висновку 0,99, що момент «переключення» відбувся. Із графіка (рис. 1) видно, що момент розладнання знаходиться приблизно в позиції 2000.

Приклад 2. Нехай $p_0 = 0{,}7$; $p_1 = 0{,}8$; $\alpha = 0{,}01$; $\beta = 0{,}001$.

За формулою (5) знайдемо мінімальну довжину інтервалу: $N = 530$. Далі обчислюємо поріг C за формулою (2): $C \approx 395{,}5$.

Оскільки ймовірність p_0 менша, ніж ймовірність p_1, то беремо критичну область $U_{\text{кр}} = \{K < C\}$, тобто для значень $K > 396$ можна вважати з імовірністю 0,99, що «переключення» відбулося.

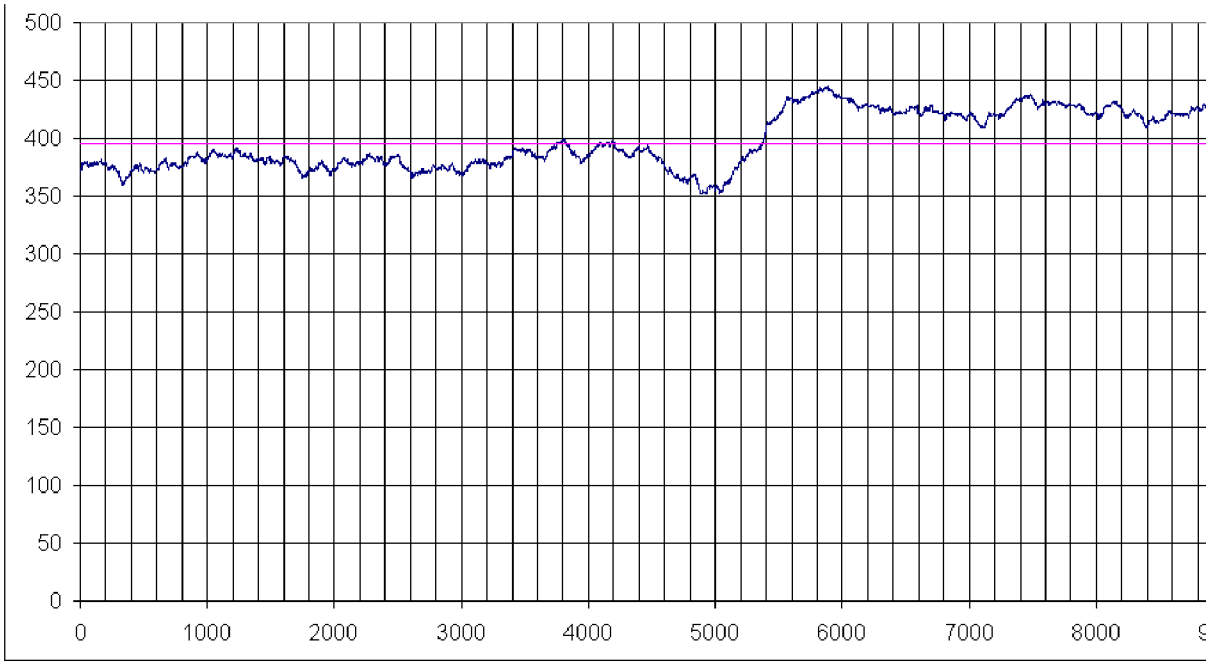

Рисунок 2- Графік, який показує кількість нулів *K*. Лінією позначена межа критичної області *C*

Як видно з графіка (рис. 2), момент «переключення» відбувся приблизно в позиції 5400.

Висновки. Для знаходження моменту розладнання у випадку потоку нулів та одиниць запропонований алгоритм, в основі якого лежить властивість стійкості відносної частоти появи випадкової події в серії експериментів. Момент розладнання співпадає з моментом зміни кількості нулів на спостережних інтервалах.

Практична цінність результатів дослідження визначається можливістю їх використання в схемах захисту інформації.

Список літератури

1. Ширяев А.Н. Статистический последовательный анализ. – М.: Наука, 1976. – 272 с.
2. Савчук М.Н., Синявский В.Ф. Об алгоритме определения моментов изменения параметров бернуллиевской последовательности // Проблемы управления и информатики. – 1999. – №1. – С.84–89.
3. Дарховский *Б.С.* Ретроспективное обнаружение «разладки» в некоторых моделях регрессивного типа // Теория вероятностей и ее применение. – 1995. – 40, №4. – С.898–903.

С Гончарова

Исследование задачи о разладке бинарной последовательности и ее применение в схемах защиты информации

Исследуется задача о разладке и ее применение в схемах защиты информации. Построен алгоритм для определения момента «переключения» вероятностей для бинарной последовательности.

S. Goncharova

Investigation of problem of the change moment of binary sequence and its application in schemes of privacy

Disorder problem and its application in schemes of privacy are investigated. The algorithm for determination of the «change» moment of probabilities for binary sequence is constructed.

Одержано 23.11.09