

В. Яковина, О. Нитребич, Д. Федасюк  
Національний університет “Львівська політехніка”,  
кафедра програмного забезпечення

## ВИКОРИСТАННЯ ІНФОРМАЦІЙНОГО КРИТЕРІЮ АКАЙКЕ В ЗАДАЧАХ МОДЕЛЮВАННЯ НАДІЙНОСТІ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

© Яковина В., Нитребич О., Федасюк Д., 2012

У багатьох моделях прогнозування надійності програмного забезпечення (ПЗ) припущення про незалежність виконання компонент є спрощенням реального характеру програмного продукту. Врахувати залежності виконання ПЗ у прогнозуванні його надійності можна за допомогою марковських ланцюгів вищих порядків. Запропоновано використовувати інформаційний критерій Акаїке (AIC) для визначення порядку марковського ланцюга, що є головною проблемою під час роботи з ними.

**Ключові слова:** надійність ПЗ, марковські ланцюги вищих порядків, критерій AIC.

In lots of models for prediction software reliability the assumption of independent software runs is just a simplification of real software. High-order Markov chains are used to count these dependencies between software runs and prediction of software reliability. In this article it is proposed to use Akaike information criterion (AIC) for Markov chain's order calculation, which is the main problem of its usage.

**Key words:** software reliability, high order Markov chain, criterion AIC.

### Вступ

Сьогодні значна кількість робіт присвячена дослідженню проблеми прогнозування надійності ПЗ, запропоновано багато моделей та методів прогнозування та оцінювання надійності програмного продукту. Залежно від використання інформації про архітектуру ПЗ, дані моделі поділяють на моделі «білої скриньки» та моделі «чорної скриньки» [1].

В останнє десятиліття все більше уваги приділяється моделям, що враховують архітектуру ПЗ, тобто моделям «білої скриньки». Переважно дані моделі використовують теорію марковських ланцюгів першого порядку, припускаючи, що виконання компонент є незалежним [2–4]. У роботах [5, 6] показано неефективність використання таких моделей і запропоновано для моделювання залежності перебування у компонентах ПЗ використовувати марковські моделі вищих порядків.

Однак використання марковських ланцюгів вищих порядків вимагає розв'язання таких задач:

- визначення порядку марковського ланцюга;
- визначення ймовірностей переходів між компонентами;
- врахування перетворення стохастичної моделі на статичну;
- врахування швидкого зростання матриці ймовірностей переходів зі збільшенням порядку марковського ланцюга.

Розглянуто задачу визначення порядку марковського ланцюга для адекватнішого прогнозування надійності програмного забезпечення.

Існують різні підходи до вирішення цього питання, зокрема, у [7] розглянуто метод  $\chi^2$  та метод, що використовує інформацію про ентропію для визначення порядку марковського ланцюга. Ми пропонуємо використовувати критерій AIC, адже він не є тестом перевірки гіпотези, не використовує рівень значущості. Окрім того, AIC дає стабільні результати і не залежить від

порядку, в якому обчислюються моделі. Цей критерій був ефективно використаний для прогнозування погоди, вибору адекватної екологічної моделі. Однак застосування цього критерію в задачах моделювання надійності програмного продукту залишається недослідженим. Тобто метою цієї роботи є дослідження можливості використання інформаційного критерію Акаїке для формалізації процесу визначення оптимального порядку марковського ланцюга, яким моделюють надійність програмної системи.

### Критерій АІС для визначення порядку марковського ланцюга в теорії надійності ПЗ

Критерій АІС, що ґрунтується на розширеному методі максимальної правдоподібності, вперше опублікував Акаїке у 1974 [8].

У загальному випадку:

$$AIC = 2k - 2 \ln(L), \quad (1)$$

де  $k$  – число параметрів в статистичній моделі,  $L$  – значення функції максимальної правдоподібності моделі. Цей критерій був запропонований як міра відхилення моделі, що розглядається, від реального об'єкта [9].

Розглянемо програмний продукт, що складається з  $M$  компонент (функціональні одиниці, які можна тестувати незалежно одна від однієї). Вважаємо, що архітектура ПЗ моделюється Марковським процесом, який містить  $M$  станів.

Нехай  $P_{ij...kl}$  ймовірність переходів для ланцюга, де суфікс містить  $r+1$  характеристику.

Визначимо гіпотезу  $H_{r-1}$ :  $P_{ij...kl} = P_{j...kl}, i = \overline{1, S}$ .

Через  $n_{ij...kl}$  позначимо кількість переходів  $i \rightarrow j \rightarrow \dots \rightarrow k \rightarrow l$  у спостережуваній послідовності. Тоді функція правдоподібності дорівнює:

$$L = \prod_{i,j,\dots,k,l} P_{ij...kl}^{n_{ij...kl}}, \quad (2)$$

де  $i, j, \dots, k, l = \overline{1, S}$ .

Оскільки  $P_{ij...kl}$  невідомі, використаємо метод максимальної правдоподібності для їхнього оцінювання:

$$\hat{P}_{ij...kl} = n_{ij...kl} / n_{ij...k}, \quad (3)$$

де  $n_{ij...k} = \sum_{l=1}^M n_{ij...kl}$ .

Позначимо оцінки гіпотези  $H_{r-1}$  штрихом:

$$\hat{P}'_{ij...kl} = n_{j...kl} / n_{j...k} \quad (4)$$

Отже, відношення правдоподібності для порівняння гіпотези  $H_{r-1}$  та  $H_r$ :

$$\lambda_{r-1,r} = L(\hat{P}'_{ij...kl}) / L(\hat{P}_{ij...kl}) = L_0(\max) / L_1(\max) \quad (5)$$

Відношення правдоподібності відповідності  $H_k$  та  $H_r$  позначимо як:

$$\lambda_{k,r} = \lambda_{k,k+1} \lambda_{k+1,k+2} \dots \lambda_{r-1,r}. \quad (6)$$

У контексті моделювання марковських ланцюгів Тонг [10] запропонував наступну функцію ризику, що ґрунтується на АІС підході:

$$R(k) = -2 \ln \lambda_{k,M} - 2(S^{M+1} - S^M - (S^{k+1} - S^k)), \quad (7)$$

де  $M$  – найвищий порядок моделі,  $k$  – порядок моделі, що розглядається. Значення змінної  $k$ , при якому функція  $R(k)$  набуває свого найменшого значення, є оптимальним порядком марковського ланцюга. Визначивши порядок марковського ланцюга, можна використовувати відомий математичний апарат класичних марковських ланцюгів, оскільки будь-який марковський ланцюг вищого порядку можна подати через марковський ланцюг першого порядку.

Очевидно, що для визначення оптимального порядку марковського ланцюга необхідно попередньо обчислити найвищий можливий порядок моделі. У роботі [11] запропоновано обчислювати максимально можливий порядок марковської моделі як найбільшу спільну довжину підпоследовності експериментальних последовностей перебування в компонентах. На нашу думку, у випадку використання архітектурного підходу до моделювання надійності ПЗ [1, 3, 4], в якому станами системи будуть стани виконання методів певного компонента (чи класу), як максимальний порядок ланцюга слід використати значення довжини максимального шляху на графі шляхів виконання програмної системи. Після цього з рівняння (7) можна визначити оптимальний порядок марковського ланцюга, який відобразатиме ту підмножину (чи підмножини) на шляху виконання програми, в яких перебування в наступному стані залежить від усіх попередніх станів цієї підмножини. Далі за допомогою виразу (1) можна оцінити ступінь адекватності отриманої моделі надійності програмної системи. Отже, використання інформаційного критерію Акаїке дає змогу оптимізувати та формалізувати процес визначення оптимального порядку марковського ланцюга для побудови моделі надійності програмного продукту, а також оцінити ступінь адекватності отриманої моделі. Подальші дослідження ведуться в напрямку експериментальних досліджень ефективності такого підходу, та в напрямку розроблення засобів автоматизації визначення оптимального порядку моделі на основі використання UML діаграм.

### Висновки

Проаналізовано використання інформаційного критерію Акаїке для задачі моделювання надійності програмних систем у випадку взаємозалежності виконання компонент. Показано, що у випадку моделювання надійності програмного забезпечення засобами марковських ланцюгів вищих порядків, використання АІС дає змогу оптимізувати та формалізувати процес визначення оптимального порядку марковського ланцюга для побудови моделі надійності програмної системи, а також оцінити ступінь адекватності отриманої моделі. Подальші дослідження будуть присвячені практичному підтвердженню результату та порівнянню із існуючими альтернативними підходами.

1. K. Goseva-Popstojanova, A.P. Mathur, K.S. Trivedi "Comparison of architecture-based software reliability models" // *12th International Symposium on Software Reliability Engineering (2001)*, 22-31.
2. H. Pham "System Software Reliability" // *Springer series in reliability engineering, Springer-Verlag London Limited(2006)*.
3. K. Goševa-Popstojanova, S. Trivedi "Architecture-based approach to reliability assessment of software systems" // *Performance Evaluation 4 (2001)*,179-204.
4. S. Krishnamurthy, A. Mathur "On the estimation of reliability of a software system using reliabilities of its components" // *Proceedings of the Eighth International Symposium on Software Reliability Engineering (1997)*, 146–155.
5. T. Takagi, Z. Furukawa, T. Yamasaki "Accurate Usage Model Construction Using High-Order Markov Chains" // *Supplementary Proceedings of 17th International Symposium on Software Reliability Engineering (2006)*, 1-2.
6. W. Burkhart, Z. Fatiha "Testing Software and Systems" // *23rd Ifip Wg 6.1 International Conference (2011)*, 236.
7. K. Goseva-Popstojanova, M. Hamill "Estimating the Probability of Failure when Software Runs are Dependent: An Empirical Study"// *20th IEEE International Symposium on Software Reliability Engineering (2009)*, 21-30.
8. H. Akaike "A new look at the statistical model identification" // *IEEE Trans. Auto. Control.* (1974), 716-723.
9. Gates, P., H. Tong "On Markov Chain Modeling to Some Weather Data" // *J. Appl. Meteor.* 15 (1976), 1145–1151.
10. H. Tong, "Determination of the order of a Markov chain by Akaike's information criterion", *Journal of applied probability* 12 (1975), 488-497.
11. S. Herbold, J. Grabowski, S. Waack "A Model for Usage-Based Testing of Event-Driven Software" // *Secure Software Integration & Reliability Improvement Companion (2011)*, 172-178.