ISSN: 1810-3022

Донецький національний університет
1937 - 2012

ПРИКЛАДНА СТАТИСТИКА АКТУАРНА ТА ФІНАНСОВА МАТЕМАТИКА

Заснований у 2000 році

видається двічі на рік

2012 рік

№ 2

Донецьк 2012

Прикладна статистика. Актуарна та фінансова математика: Наук. журнал / Донецький нац. ун-т. – 2012. - №2. – 187 с.

Журнал "Прикладна статистика. Актуарна та фінансова математика" приймає оригінальні статті та короткі повідомлення про математичне моделювання та керування різноманітними процесами природознавства, техніки та економіки, страхування, інвестування, фінансування, особливо в тих галузях, що спираються на стохастичні методи. Чекаємо на теоретичні дослідження, а також на статті про практичні засоби та алгоритми розв'язування задач. Усі статті рецензуються.

Свідоцтво про реєстрацію серія КВ № 16146-4618 ПР

Друкується за рішенням Вченої Ради Донецького національного університету

Протокол №4 від 26.04.2013.

Адреса редколегії: кафедра теорії ймовірностей та математичної статистики, ДонНУ, вул. Університетська 24, Донецьк 83001, Україна.

Тел.: +38 (0622) 302-92-36

E-mail: artyombayev@rambler.ru

Applied statistics. Actuarial and financial mathematics: Scientific Journal / Donetsk National University. – 2012. - №2. – 187 p.

Journal "Applied statistics. Actuarial and financial mathematics" will accept for publication original articles and short reports devoted to mathematical modeling, control of different natural, economic, technical, insurance, investment and financial processes especially in the domains based on stochastic methods. You are most welcome to submit results theoretical studies as well as articles on practical methods and algorithms for solution of problems. All articles will be revised.

Registration series КВ № 16146-4618 ПР

Printed by Authority of Academic Council of Donetsk National University

Protocol № 4/ 26.04.2013

Contact Edit. Board at: Department of Probability Theory and Mathematical Statistics, Donetsk National University, Universitetskaya st. 24, Donetsk 83001, Ukraine.

Call us: +38 (0622) 302-92-36

E-mail: artyombayev@rambler.ru

ЗМІСТ

ПРОЦЕДУРА СТОХАСТИЧНОЇ ОПТИМІЗАЦІЇ ДЛЯ ІНДЕКСУ ВЕЛИЧИНИ ПРОГРАМНОГО ПРОДУКТУ ПРИ ТЕСТУВАННІ

УДК 519.21

Хімка У.Т., Чабанюк Я.М.

Резюме. Рассматривается непрерывная процедура стохастической оптимизации для индекса величины программного продукта при тестировании на выявление ошибок в математической модели тестирования программного обеспечения на основании количества ошибок. Получены условия сходимости процедуры к асимптотическим значениям индекса при неполном тестировании с момента выхода количества ошибок на распределение Пуассона.

Резюме. Розглядається неперервна процедура стохастичної оптимізації для індексу величини програмного продукту при тестування на знаходження помилок в математичній моделі тестування програмного забезпечення на основі кількості помилок. Отримано умови збіжності процедури до асимптотичних значень індексу при неповному тестуванні з моменту виходу кількості помилок на розподіл Пуассона.

Abstract. We consider continuous stochastic optimization procedure for the index values of the software under test for finding errors in the mathematical model of software testing based on the number of errors. The conditions of the convergence procedure on the asymptotic values of the index of incomplete testing since the number of errors on the Poisson distribution.

Received/ Надійшла до редакції 01.06.2012.

Ключевые слова: стохастическая оптимизация, марковский процесс, индекс величины програмного продукта, процесс Пуассона.

Ключові слова: стохастична оптимізація, марковський процес, індекс величини програмного продукту, процес Пуассона.

Keywords: stochastic optimization, Markov process, index value of software, Poisson process.

Вступ. Неперервна процедура стохастичної оптимізації

$$du^{\varepsilon}(t) = a(t)\nabla_{b(t)}C(u^{\varepsilon}(t))dt + \frac{a(t)}{b(t)}\sum_{r=1}^{k}\sigma_r(t,u)d\xi_r(t) \quad (1)$$

пошуку точки $u = u^*$ екстремуму функції регресії $C(u), u \in R,$ коли її значення спостерігаються з помилками типу гаусівського білого шуму $\dot{\xi}_r(t) = \dot{\xi}_r(t,\varpi), r = \overline{1,k},$ розглянуто в [1, 2]. Доцільність побудови псевдоградієнта

$$\nabla_{b(t)}C(u) = \frac{C(u+b(t)) - C(u-b(t))}{2b(t)}$$

обґрунтовано в [3]. Збіжність процедури (1) формулюється в термінах властивостей гладкості функцій Ляпунова, та при виконанні умов

$$\int_0^{\infty} a(t)dt = \infty, \int_0^{\infty} \frac{a^2(t)}{b^2(t)}dt < \infty, \quad (2)$$

на керуючі функції $a(t) > 0, b(t) > 0.$

Відзначимо, що перша з умов (2) "примушує" рухатись випадкову еволюцію $u(t)$ до точки $u = u^*$, а друга умова сповільнює цей рух таким чином, щоб встигнути досягнути точку $u = u^*$ при $t \to \infty.$

Процедура стохастичної оптимізації (1) використовуются як для непараметричних так і для параметричних задач математичної статистики [4, 5].

При безпосередній залежності функції регресії від зовнішнього середовища, що описується марковськими переключеннями, розглянуто процедуру стохастичної оптимізації [6]

$$\frac{du(t)}{dt} = a(t)\nabla_{b(t)}C(u(t);x(t)), u \in R^d, \tag{3}$$

де $\nabla_{b(t)}C(u(t);x(t)) := \dfrac{C(u(t)+b(t);x(t)) - C(u(t)-b(t);x(t))}{2b(t)}$, а $x(t), t \geq 0$, рівномірно ергодичний марковський процес у вимірному фазовому просторі станів $(X,\mathbf{X})$ з стаціонарним розподілом $\pi(B), B \in \mathbf{X}$ [7].

В цьому випадку експериментатор має можливість отримати асимптотично оптимальний результат навідь при відомій функції регресії , та при всіх значеннях марковського процесу $x(t), t \geq 0$ [1, 7]. Доведення збіжності процедури стохастичної оптимізації (3) з ймовірністю 1, а саме

$$P\{\lim_{t\to\infty} u(t) = u^*\} = 1$$

проводиться з використанням асимптотичних властивостей генератора $\boldsymbol{L}_t^\varepsilon$ двокомпонентного марковського процесу $u^\varepsilon(t), x(t/\varepsilon), t \geq 0, \varepsilon > 0$ - малий параметр серій[6], розв'язку проблеми сингулярного збурення та застосування теореми Невельсона-Хасьмінського ([1], Теорема 2.8.1, стор. 100).

Отже важливм кроком застосування процедури стохастичної оптимізації (3) є встановлення природи стохастичності , що входить в функцію регресії $C(u;x)$.

Основною метою роботи є встановлення можливості використання процедури стохастичної оптимізації (3) для отримання асимптотичних значень індекса величини програмного проекту з функцією інтенсивності виявлених помилок, що запропонована авторами в [6]. В першій частині данної статті досліджуються стохастичні властивості точкових оцінок запропонованої моделі опису процесу тестування програмного продукту, а в другій - проводиться аналіз щодо застосування процедури стохастичної оптимізації (3) для індексу величини проекту. Отже ця статті є продовженням робіт [6].

1. Математична модель тестування ПЗ з індексом величини проекту.

Розглядається модель дослідження тестування програмного забезпечення на основі кількості помилок [8, 9, 10]. Предметом такої моделі є кількість помилок на певному скінченному інтервалі часу, а не час між виявленими помилками. При цьому, як правило, припускають, що кількість помилок на декількох сусідніх інтервалах є стохастичним процесом з дискретною або неперервною інтенсивністю виявлення помилок залежною від часу. Отже параметри функції інтенсивності оцінюються за спостереженнями процесу тестування ПЗ через отриману статистику помилок [11].

Основним припущенням класичних моделей на основі кількості помилок є те, що з певного моменту кількість помилок на декількох сусідніх часових інтервалах тестування мають розподіл Пуассона.

Розглянемо запропоновану в [12] функцію інтенсивності з індексом величини програмного проекту

$$\lambda(t) = \alpha\beta^{s+1}t^s \exp(-\beta t) \tag{4}$$

де α – коефіцієнт, що характеризує загальну кількість помилок в ПЗ, β – коефіцієнт, що характеризує загальну тривалість процесу виявлення помилок, s – індекс величину проекту.

Згідно з [12] для (4) визначена функція кумулятивної кількості помилок до моменту t

$$\mu(t)=\int_0^t \lambda(\tau)d\tau=\alpha\left[s\Gamma_{\beta t}(s)-\beta^s t^s e^{-\beta t}\right] \tag{5}$$

де $\Gamma_z(p)=\int_0^z t^{p-1}e^{-t}dt$, $\operatorname{Re} p>0$, – неповна гама-функція.

Зауваження 1. При $s=1$ маємо з (4) та з (5) наступні представлення

$$\lambda(t)=\alpha\beta^2 t\exp(-\beta t),\ \mu(t)=\alpha(1-(1+\beta t)\exp(-\beta)),$$

що відповідає S - подібній моделі [13].

Загальна кількість помилок, що міститься в програмному продукті визначається з (5) при $t\to\infty$

$$\mu(\infty)=\alpha s\Gamma(s),$$

де $\Gamma(s)$ – гама-функція, що також співпадає з S - подібною моделлю.

Таким чином, згідно з основним припущенням та властивостями кумулятивної функції (5), для знаходження параметрів α, β та s будуємо функцію максимальної правдоподібності

$$L(\alpha,\beta,s)=\prod_{i=1}^{n}\frac{\left[\mu(t_i)-\mu(t_{i-1})\right]^{m_i}}{m_i!}\exp\left(\mu(t_{i-1})-\mu(t_i)\right), \tag{6}$$

де m, $(\sum_{i=1}^{n} m_i=k)$, - кількість виявлених помилок на інтервалі $(t_i;t_{i+1}], i=\overline{0,n}.$

Отже, використовуючи (6), отримуємо точкові оцінки $\hat{\alpha}$, $\hat{\beta}$ та $\hat{s}$ параметрів α, β, s з системи рівнянь

$$\begin{cases}\dfrac{\partial\ln L(\alpha,\beta,s)}{\partial\alpha}=0,\\[2mm] \dfrac{\partial\ln L(\alpha,\beta,s)}{\partial\beta}=0,\\[2mm] \dfrac{\partial\ln L(\alpha,\beta,s)}{\partial s}=0.\end{cases}$$

З останньої системи маємо систему

$$\begin{cases}\alpha=\dfrac{k}{s\Gamma_{\beta t_n}(s)-\beta^s t_n^s e^{-\beta t_n}};\\[3mm] \displaystyle\sum_{i=1}^{n}\frac{m_i\left(t_i^{s+1}e^{-\beta t_i}-t_{i-1}^{s+1}e^{-\beta t_{i-1}}\right)}{s\Phi_{\beta t_i t_{i-1}}(s)+\beta^s\left(t_{i-1}^s e^{-\beta t_{i-1}}-t_i^s e^{-\beta t_i}\right)}-\frac{k t_n^{s+1}e^{-\beta t_n}}{s\Gamma_{\beta t_n}(s)-\beta^s t_n^s e^{-\beta t_n}}=0;\\[3mm] \displaystyle\sum_{i=1}^{n}\frac{m_i}{s\Phi_{\beta t_i t_{i-1}}(s)+\beta^s\left(t_{i-1}^s e^{-\beta t_{i-1}}-t_i^s e^{-\beta t_i}\right)}\left[\Phi_{\beta t_i t_{i-1}}(s)+s\mathrm{F}_{\beta t_i t_{i-1}}(s)+\beta^s(t_{i-1}^s e^{-\beta t_{i-1}}\ln(\beta t_{i-1})-\right.\\ \displaystyle\left.-t_i^s e^{-\beta t_i}\ln(\beta t_i))\right]+\frac{k}{s\Gamma_{\beta t_n}(s)-\beta^s t_n^s e^{-\beta t_n}}\left[\beta^s t_n^s e^{-\beta t_n}\ln(\beta t_n)-\Gamma_{\beta t_n}(s)-s\int_0^{\beta t_n}\tau^{s-1}\ln\tau e^{-\tau}d\tau\right]=0;\end{cases} \tag{7}$$

де

$$\Phi_{\beta t_i t_{i-1}}\left(s\right)=\int\limits_{\beta t_{i-1}}^{\beta t_i}\tau^{s-1}\exp\left(-\tau\right)d\tau\,,\ \mathrm{F}_{\beta t_i t_{i-1}}\left(s\right)=\int\limits_{\beta t_{i-1}}^{\beta t_i}\tau^{s-1}\ln\tau\exp\left(-\tau\right)d\tau\,.$$

Приклад 1. Для тестування приведеної схеми пошуку $\hat{\alpha}$, $\hat{\beta}$ та $\hat{s}$ розглянуто емпіричні данні першого експерименту роботи [10]. В данному експерименті програмний продукт мав 38 помилок, що тестувався 40 разів з різною кількістю серій тестувань (до 1200). Згідно з (7) для експерименту отримали $\hat{s}=0.265$, що дозволяє розглядати програмний продукт як невеликий (9564 рядків програмного коду мовою С, з яких 6218 виконувані). За отриманих значень $\hat{\alpha}=38.879$, $\hat{\beta}=0.005217$ та $\hat{s}$ з (5) отримуємо 35, 1 помилок в ПЗ.

Зауваження 2. Запропонована математична модель не дозволяє отримати оцінку кількості залишкових помилок в програмному продукті.

2. Процедура стохастичної оптимізації для індексу величини проекту.

Оскільки α, β та s в (4) визначається під дією процесу Пуассона кількості помилок на сусідніх інтервалах, визначимо можливість побудови процедури стохастичної оптимізації для них. Лінійність функції інтенсивності $\lambda(t)$ відносно α не дозволяє отримати таку процедуру. Тому $\hat{\alpha}$, отримане з (7) є постійним для процедури стохастичної оптимізації відносно параметрів β та s. З іншої сторони з природи параметру β в (4) слідує, що він змінюється мало при наборі достатньої статистики помилок при тестуванні (перехід кількості виявлених помилок на сусідніх інтервалах $(t_i;t_{i+1}],(t_{i+1};t_{i+2}]$до розподілу Пуассона), що підтверджують експериментальні данні (Приклад 1.).

Таким чаном отримуємо єдино можливу процедуру стохастичної оптимізації для параметра s з функцією регресії

$$\lambda(s,t)=\hat{\alpha}\hat{\beta}^{s+1}t^{s}\exp\left(-\hat{\beta}t\right). \tag{8}$$

Зауваження 3. Згідно властивостей точкових оцінок $\hat{\alpha}$, $\hat{\beta}$ (при великих n та k) отримуємо їх незміщеність відносно α та β. З іншої сторони α та β змінюються під впливом марковського процесу виявлення помилок [14].

Отже отримуємо процедуру стохастичної оптимізації згідно (3) параметра s

$$\frac{ds(t)}{dt}=\frac{a(t)}{2b(t)}\times\left[\hat{\alpha}\hat{\beta}^{s(t)+b(t)+1}t^{s(t)+b(t)}e^{-\hat{\beta}t}-\hat{\alpha}\hat{\beta}^{s(t)-b(t)+1}t^{s(t)-b(t)}e^{-\hat{\beta}t}\right]. \tag{9}$$

Зауваження 4. Функція інтенсивності виявлення помилок $\lambda(s,t)$ згідно (8) має єдиний максимум.

Теорема. Якщо функції $a(t)>0, b(t)>0$ задовольняють умовам

$$\int\limits_{t_0}^{\infty}a(t)dt=\infty,\int\limits_{t_0}^{\infty}a(t)b(t)dt<\infty, t_0>0, \tag{10}$$

то для процедур стохастичної оптимізації (9) має місце збіжність

$$P\{\lim_{t\to\infty}s(t)=s^*\}=1, \tag{11}$$

де s^* таке, що $\max\lambda(s,t)=\lambda(s^*,t)$.

Доведення. Збіжність з ймовірністю 1 (11) слідує з гладкості функції $\lambda(s,t)$, тобто виконання умови Ліпшиця в формі

$$\left|\nabla_{b(t)}\lambda(s,t)-\lambda_s'(s,t)\right|\le cb(t),$$

умов (10) та твердження теореми 1 з [6].

Приклад 2. Для розглянутого тестового експерименту з прикладу 1 побудовано процедуру стохастичної оптимізації (9) керуючими функціями

$$a(t)=\frac{a}{t}, b(t)=\frac{b}{\sqrt[4]{t}}$$

де a,b – деякі додатні константи, наприклад $a=1, b=1$. З врахуванням початкової умови значення $\hat{s}$ останньої тестової ітерації отримуємо задачу Коші $s(t)$. Розв'язком цієї задачі буде граничне значення індексу величини програмного продукту. Використання асимптотичного значення такого розв'язку $s(\infty)=0{,}272$ дозволяє оцінити загальну кількість помилок через значення $\mu(\infty)=36.7$ за вище вказаною формулою при визначених в прикладі 1 $\hat{\alpha}$ та $\hat{\beta}$.

Висновки. Побудовано процедуру стохастичної оптимізації для індексу величини програмного продукту як критерій достатності процесу тестування програмного забезпечення з врахуванням стохастичності процесу тестування. Створена процедура дозволяє оцінити кількість залишкових помилок, що в свою чергу дає можливість передбачити матеріальні витрати на тестування та його доцільність.

ЛІТЕРАТУРА

1. Невельсон М. Б. Стохастическая аппроксимация и рекуррентное оценивание / М. Б. Невельсон, Хасьминский Р. З. // М.:Наука, – 1972. – 304с.
2. Ljung L. Stochastic Approximation and Optimization of Random Systems / L. Ljung, G. Pflug, H. Walk // Birkhauser Verlag, Basel, Boston, Berlin, – 1992. – 113P.
3. Kiefer E. Stochastic estimation of the maximum of regression function./ Kiefer E., Wolfowitz J.// Ann. Math. Statist., - 1952. - **23**, №3,- P. 462 - 466.
4. Стратович Р.Л. Условные марковские процессы и их применение в теории оптимального управления. - М.: МГУ - 1966. -324с.
5. Цыпкин Я. З. Основы теории обучающихся систем. - М.: Наука. - 1970.- 251с.
6. Khimka U.T. Stochastic optimization procedure convergence with Markov switching in the averaging scheme / Khimka U.T., Chabanyuk Ya.M. // Matematychni studii. Vol 34. № 1. 2010.- P. 126-134.
7. Korolyuk V. S. Stochastic Systems in Merging Phase Space. / V S Korolyuk, Limnios // World Scientific Singapore. – 2005. – 330 p.
8. Липаев В.В. Надежность программных средств. – М.: СИНТЕГ, 1998. – 232 с.
9. Goel, A.L. Software reliability models: assumptions, limitations, and applicability. // IEEE Transactions on software engineering. 1985, Vol. SE-11, No 12, pp. 1411-1423.
10. K.-Y. Cai, D.-B. Hu, Ch.-G. Bai, H. Hu, T. Jing Does software reliability growth behavior follow a non-homogeneous Poisson process // Information and Software Technology. – Vol. 50. – 2008. – P. 1232–1247.
11. Половко А.М., Гуров С.В. Основы теории надежности. – СПб.: БХВ-Петербург, 2006. – 704 с.
12. Хімка У.Т. Критерій достатності процесу тестування програмного продукту./Чабанюк Я. М., Яковина В.С., Федасюк Д.В., Сенів М.М., Хімка У.Т.//

Вісник НУ «ЛП».Комп'ютерні науки та інформаційні технології. №672. 2010.- С. 346-358.

13. S. Yamada, M. Ohba, S. Osaki S-shaped reliability growth modeling for software error detection // IEEE Transactions on Reliability. – Vol. R-32. – No.5. – 1983. – P. 475–478.
14. Королюк В. С. Стохастичні моделі систем /В. С. Королюк // Киев: Либідь, – 1993. – 136с.