Література
1. Гантмахер Ф.Р. Теория матриц / Ф. Р. Гантмахер - М. : Наука, 1966. - 576 с.
2. Псевдообернена матриця [Електронний ресурс]./ – Режим доступу: http://uk.wikipedia.org/wiki/Псевдообернена_матриця
3. Стренг Г. Линейная алгебра и ее применения / Г. Стренг. - М. : Мир, 1980. - 454 с.

УДК 004.94(075.8)

АНАЛІЗ ВИХІДНИХ ДАНИХ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ

В.М. Томашевський
Національний технічний університет України «КПІ»

Імітаційне моделювання традиційно використовується у тих випадках, коли застосування аналітичних методів неможливе. Розвиток потужних і недорогих комп'ютерів, а також досягнення у розробці програмного забезпечення сприяє збільшенню популярності імітаційного моделювання. Різні пакети імітаційного моделювання з дружнім до користувача інтерфейсом пропонують візуальні інтерактивні можливості, що спрощують процес створення імітаційних моделей. Незалежно від того, як розвиваються методології програмування для реалізації імітаційного моделювання, події, що моделюються, в більшості випадків керуються випадковими числами. Це призводить до того, що отримані результати – не що інше як статистичні вибіркові послідовності – часові ряди.

Процес побудови, валідації, перевірки та використання імітаційної моделі для прийняття рішення може бути важким. Витрачається багато часу і зусиль для декількох відокремлених задач: обговорення моделі з особами, що приймають рішення, які є кінцевими користувачами результатів моделювання; визначення необхідних даних і їх збирання для отримання прийнятного розподілу для різноманітних компонентів моделі; кодування та перевірка імітаційної моделі; валідація моделі та оцінка її функціонування.

Після того як приділено стільки уваги на ранніх кроках процесу моделювання, необхідно вірно проаналізувати вихідні дані. Якщо це не зробити, то зводиться нанівець уся проведена робота. На щастя, етап аналізу вихідних даних звичайно є набагато менш трудомісткий ніж раніше згадані етапи моделювання та кодування. Ситуація покращується ще й тим, що розробники програм імітаційного моделювання збільшують можливості аналізу в своїх програмних продуктах. Аналіз вихідних даних дозволяє кінцевому користувачеві краще впливати на ефективність моделі, і таким чином приймати кращі рішення.

Існує декілька причин того, що аналіз вихідних даних не завжди виконується належним чином. По-перше, у користувачів нерідко створюється хибне враження, що імітаційне моделювання – всього лише вправа в комп'ютерному програмуванні, нехай і досить складна. В результаті чого часто вивчення систем за допомогою імітаційного моделювання починається з розробки концептуальної моделі з подальшим „програмуванням" і закінчується єдиним прогоном для видачі „відповідей". Тоді як імітаційне моделювання – це експеримент зі статистичними вибірками з використанням обчислювальної машини. Щоб результати імітаційного моделювання мали яке-небудь значення, потрібно використовувати статистичні методи для розробки і аналізу моделюючих експериментів. По-друге, вихідні процеси практично всіх прогонів імітаційних моделей є нестаціонарними і корельованими. Таким чином, для їх аналізу класичні статистичні методи, що основані на незалежних і однаково розподілених спостереженнях, не можуть застосовуватись безпосередньо. На сьогодні є декілька проблем, що пов'язані з аналізом вихідних даних, для яких не існує повністю прийнятного рішення; при цьому доступні методи виявляються дуже складними в застосуванні.

Статистичний висновок є абсолютно необхідним у будь-якій ситуації, коли одна й та ж сама (правильна) програма створює різні (але правильні) вихідні дані від кожного прогону. Будь-яка послідовність $x_1, x_2, ..., x_n$ таких вихідних даних просто складається з реалізації випадкових величин $X_1, X_2, ..., X_n$.

Найпростіша мета застосування імітаційного моделювання – оцінка середнього μ_X, проаналізованого випадкового процесу $x_1, x_2, ..., x_n$, що можна зробити обчислюючи середнє число послідовності зібраних спостережень як

$$\overline{X}(n) = \sum_{i=1}^{n} \frac{xi}{n} \quad . \tag{1}$$

Таке середнє число приймає випадкове значення, що залежить від послідовності спостережень. Точність з якою $\overline{X}(n)$ оцінює невідомий параметр μ_X може бути оцінена ймовірністю

$$P(\left|\overline{X}(n) - \mu_x\right| < \Delta_x) = 1 - \alpha \tag{2}$$

або

$$P(\overline{X}(n) - \Delta_x < \mu_x < \overline{X}(n) + \Delta_x) = 1 - \alpha\,, \tag{3}$$

де Δ_x – напівширина довірчого інтервалу для оцінюваного значення; ($1-\alpha$) – довірчий рівень, $0 < \alpha < 1$.

Таким чином, якщо ширина $2\Delta_x$ довірчого інтервалу знайдена для прийнятого довірчого рівня $1-\alpha$, та імітаційне моделювання було повторене декілька разів, то інтервал $(\overline{X}(n) - \Delta_x, \overline{X}(n) + \Delta_x)$ буде містити невідоме середнє число μ_X у 100 ($1-\alpha$) % випадках і не буде містити його в $100\,\alpha$% випадках. Відомо що, якщо спостереження $x_1, x_2, \ldots, x_n$ розглядаються як реалізація незалежних і нормально розподілених випадкових величин $X_1, X_2, \ldots, X_n$, то

$$\Delta_x = t_{n-1,1-\alpha/2}\hat{\sigma}[\overline{X}(n)]\,, \tag{4}$$

де

$$\hat{\sigma}^2[\overline{X}(n)] = \sum_{i=1}^{n} \frac{(x_i - \overline{X}(n))^2}{n(n-1)} \tag{5}$$

є незміщеною оцінкою дисперсії $\overline{X}(n)$, і $t_{n-1,1-\alpha/2}$ є верхньою ($1-\alpha/2$) критичною межею $t-$розподілу Стьюдента з (*n* - 1) ступенями вільності. Інакше кажучи, для даної величини ($1-\alpha/2$) приймається $t-$розподіл для випадкової змінної

$$T_{n-1} = \frac{\overline{X}(n) - \mu_x}{\sigma^2[\overline{X}(n)]}\,, \tag{6}$$

і виходить, що

$$P(T_{n-1} \le t) = 1 - \alpha/2 \tag{7}$$

для $t = t_{n-1,1-\alpha/2}$.

Із цієї причини $t_{n-1,1-\alpha/2}$ називається ($1-\alpha/2$) квантіллю. Для значення *n* > 30, $t-$розподіл може бути замінений нормальним розподілом.

Рівняння (4) може також бути застосоване, якщо спостереження $x_1, x_2, ..., x_n$ представляють собою випадкові величини, що мають розподіл відмінний від нормального. Тобто якщо ці спостереження – реалізація незалежних і однаково розподілених випадкових величин $X_1, X_2, ..., X_n$, тоді відповідно до центральної граничної теореми, розподіл змінної $\overline{X}(n)$ наближається до нормального розподілу зі збільшенням кількості спостережень *n*. На практиці рівняння (4) дає хороше наближення для $n > 100$. Результати, отримані з рівнянь (1) і (4) називаються точковими та інтервальними оцінками відповідно. Обидві ці характеристики є важливі: перша характеризує проаналізовану систему, а друга показує точність отриманих характеристик.

Якщо спостереження $x_1, x_2, ..., x_n$ не можуть бути розцінені як реалізація незалежних і нормально розподілених випадкових величин, повинні розглядатися деякі модифікації для вищезгаданих оцінок. Це піднімає проблему якості оцінок. Є три загальних міри ефективності оціночної функції:

Зміщення, що вимірює систематичне відхилення оціночної функції від справжнього значення параметра, що оцінюється; наприклад, у випадку $\overline{X}(n)$

$$Bias[\overline{X}(n)] = E[\overline{X}(n) - \mu_x] \,. \tag{8}$$

Дисперсія, що вимірює середнє відхилення оціночної функції від його середнього значення

$$\sigma^2[\overline{X}(n)] = E[(\overline{X}(n) - E[\overline{X}(n)])^2] \,. \tag{9}$$

Середня квадратична помилка (MSE) оціночної функції

$$MSE[\overline{X}(n)] = E\{[\overline{X}(n) - \mu_x]^2\} \,. \tag{10}$$

Потрібно зауважити, що із цих визначень випливає наступне:

$$MSE[\overline{X}(n)] = \{Bias[\overline{X}(n)]\}^2 + \sigma^2[\overline{X}(n)] \tag{11}$$

Основна аналітична проблема, як виникає при аналізі результатів стохастичного імітаційного моделювання полягає в тому, що отримані результати моделювання корельовані та не задовольняють передумові статистичної незалежності. Якщо спостереження $x_1, x_2, ..., x_n$ представляють собою автокоррельовану і стаціонарну послідовність випадкових величин $X_1, X_2, ..., X_n$, тоді дисперсія $\overline{X}(n)$ задається формулою

$$\sigma^2[\overline{X}(n)] = \frac{[R(0) + 2\sum_{k=1}^{n-1}(1 - \frac{k}{n})R(k)]}{n} , \qquad (12)$$

де

$$R(k) = E[(X_i - \mu_x)(X_{i-k} - \mu_x)],\ 0 \le k \le n-1 \qquad (13)$$

є автоковаріацією порядку k (лаг k функції автокореляції $R(k)$). Автоковаріації, визначені в рівнянні (13) не залежать від індексу i завдяки стаціонарності процесу, що аналізується. Потрібно звернути увагу, що дисперсія $\sigma^2[\overline{X}(n)]$ може бути зменшена до $\frac{R(0)}{n}$, і отже може бути оцінена рівнянням (5), тоді і тільки тоді, коли спостереження є не корельованими. Нехтування існуючою статистичною кореляцією еквівалентно видаленню всіх компонентів, окрім $R(0)$ з рівняння (12). Таке наближення часто не є припустимим на практиці. Наприклад, у статті [1] для системи масового обслуговування M/M/1 з 90 % коефіцієнтом використання показано, що дисперсія середнього довжини черги, розрахованої відповідно до рівняння (12), в 367 разів більше ніж, якщо розрахувати її за рівнянням (5). Будь-який аналіз дисперсії з ігноруванням кореляції між спостереженнями веде або до надмірно оптимістичного довірчого інтервалу для μ_X, у випадку позитивно корельованих спостережень, або до надмірно песимістичного довірчого інтервалу для μ_X, у випадку негативно корельованих спостережень (див. рівняння (4)). Позитивна кореляція між спостереженнями типова в простих системах масового обслуговування без зворотного зв'язку, і вона тим більш сильна, чим більший коефіцієнт використання пристрою в системі. Результати дослідження СМО за допомогою регенеративного методу, реалізованого в GPSS, надані у книзі [2].

Взагалі, дисперсійний аналіз корельованих процесів та аналіз їхніх функцій автокореляції зокрема є складною статистичною проблемою й тому створює великі труднощі в статистичному аналізі вихідних даних імітаційного моделювання. При скінченному імітаційному моделюванні, коли відбувається «природна» подія, яка дає змогу визначити тривалість прогону моделі, вищезгадана проблема може бути вирішена шляхом багатократного запуску незалежних повторень експерименту з різними послідовностями випадкових чисел. У цьому випадку середні індивідуальних спостережень, зібраних протягом різних прогонів імітаційного моделювання, можуть бути розцінені як послідовність незалежних вихідних даних (у силу незалежності випадкових чисел), і може бути застосовано рівняння (5).

Розглянемо імітаційне моделювання в стаціонарному режимі на прикладі функціонування систем масового обслуговування (СМО) за тривалий період часу. Після запуску моделювання система масового обслуговування перебуває у нестаціонарному, перехідному режимі, а потім спостерігається перехід до сталого режиму (режиму статистичної рівноваги). Тривалість перехідного режиму залежить від параметрів СМО. Оскільки спостереження, зібрані протягом початкових перехідних періодів не характеризують стаціонарний стан, головна ідея полягає у відмові від усіх таких спостережень перед подальшим аналізом. Це вимагає оцінки ефективної довжини початкового перехідного періоду. Ігнорування існування цього періоду може призвести до істотного зсуву кінцевих результатів. З іншого боку, видалення будь-яких спостережень збільшує дисперсію оцінок, які у свою чергу можуть збільшити значення середньої-квадратичної помилки. Таким чином, рішення, видаляти або не видаляти початкові спостереження залежить від прийнятого критерію якості оцінок.

Існують кілька методів збирання даних й аналізу для вирішення теоретичних проблем, пов'язаних з кореляцією спостережень для сталого режиму імітаційного моделювання, описаних у книзі [3], які відрізняються оцінюванням дисперсії аналізованих процесів. Ці оцінки є необхідними для визначення ширини довірчого інтервалу. Найбільш використовуваними є наступні методи: повторних експериментів (реплікацій); спільних середніх, що не перетинаються; спільних середніх, що перетинаються; отримання некорельованої вибірки; регенеративних циклів; метод, що базується на спектральному аналізі; метод, що базується на авторегресивному поданні даних; використання стаціонарних часових рядів.

Будь-який метод аналізу повинен бути застосований для статистично значущої кількості повторних експериментів імітаційного моделювання (зазвичай 200 або більше реплікацій), щоб визначити частину експериментів з довірчими інтервалами, які покривають справжнє середнє значення оцінюємого параметра. Помилка охоплення та її основні джерела були теоретично проаналізовані в роботах, посилання на які містяться в статті [4]. Взагалі, даний метод збирання даних й аналізу можна розглянути як створення дійсних $(1-\alpha)100$ % довірчих інтервалів, якщо верхня межа покриття довірчих інтервалів не менша ніж ($1-\alpha$). У всіх інших випадках, довірчі інтервали для параметра, що оцінюється, повинні бути розцінені як неправильні а метод як неточний.

Створення автоматизованої процедури, яка могла б використовуватися для стохастичного імітаційного моделювання широкого класу систем – проблема для майбутніх досліджень.

Література

1. Blomqvist N.. The covariance function of the M/G/1 queueing system.//Skandinavisk Aktuarietidskrift, p. 157-174, 1967.
2. Томашевский В.Н., Жданова Е.Г. Имитационное моделирование в бреде GPSS. – М.: Бестселлер, 2003.
3. Аверилл М. Лоу, Келтон В.Д. Имитационное моделирование. 3-е изд. – СПб.: Питер; Киев: Издательская группа BHV, 2004.
4. Susan M. Sanches "ABC's of output analysis" // Proc. 1999 Winter Simulation Conf., p.24-30, IEEE, 1999.

УДК 004.42; 004.588

ОСОБЕННОСТИ СИСТЕМ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Е.В. Рындич
Черниговский национальный технологический университет

Постоянное развитие современных информационных технологий привело к тому, что такая сфера жизнедеятельности, как образование, требует проведения модернизации, создания и внедрения новых средств и инструментов в обучение для поддержания требуемого уровня. Наиболее перспективной является концепция дистанционного образования, поскольку именно данное направление предполагает использование практически всех доступных достижений современных информационных технологий. Она охватывает широкие слои общества и становится важнейшим фактором его развития.

Дистанционное образование – это система образования, которая предусматривает активное общение студента с преподавателем при помощи современных информационных технологий и дает свободу выбора места, времени и темпа обучения [1].

На данный момент существует уже не одна система дистанционного обучения, существуют инструменты и технологии, которые позволяют упростить процесс создания систем такого класса. Кроме стандартных CMS(Content Management System) стоит выделить более узкоспециализированные системы LMS(Learning Management System) и LCMS (Learning Content Management System).