УДК 519.872

И.Н. КУЗНЕЦОВ

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗЫ ОБ АСИМПТОТИЧЕСКОЙ НОРМАЛЬНОСТИ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ ЧИСЛА ТРЕБОВАНИЙ В СИСТЕМЕ $GI/G/\infty$ В УСЛОВИЯХ БОЛЬШОЙ ЗАГРУЗКИ

Аннотация. Рассмотрена система обслуживания $GI/G/\infty$ в условиях большой загрузки. Статистические данные сгенерированы методом Монте-Карло. Исследована возможность применения некоторых статистических критериев для проверки гипотезы об асимптотически нормальном распределении количества требований в системе в стационарном режиме. Приведены численные примеры.

Ключевые слова: система обслуживания, большая загрузка, стационарное распределение, статистическая гипотеза, асимптотически нормальное распределение, метод Монте-Карло.

ВВЕДЕНИЕ

Одной из важнейших составляющих современной информационной структуры общества являются телекоммуникационные сети. Их развитию, исследованию и внедрению уделяется повышенное внимание. Наиболее адекватные модели телекоммуникационных сетей — системы массового обслуживания, функционирующие в условиях большой загрузки, когда количество поступающих заявок исчисляется сотнями. Во многих работах входящие потоки предполагаются пуассоновскими, что существенно упрощает исследование и в ряде случаев позволяет получить явные аналитические формулы. В то же время далеко не всегда статистические данные подтверждают пуассоновское свойство входящих потоков. В этом случае используют приближенные методы, в частности, асимптотические, позволяющие при значениях некоторых параметров, близких к критическим, строить весьма точные оценки искомых характеристик системы. Существенный вклад в развитие асимптотических методов анализа систем обслуживания в условиях большой загрузки внесла и украинская школа теории вероятности [1–8].

В настоящей статье исследуется система обслуживания $GI/G/\infty$, состоящая из бесконечного числа обслуживающих каналов. Обслуживание начинается в момент поступления требования, время обслуживания имеет функцию распределения $B(x)$ и конечную среднюю длительность τ. В систему поступает рекуррентный поток требований, задаваемый функцией распределения $A(x)$; θ — конечная средняя длительность интервала между последовательно поступающими требованиями.

Предполагается, что функция $A(x)$ имеет абсолютно непрерывную компоненту, что гарантирует существование стационарного режима работы системы. Пусть $B(x)=B^{(0)}(x/\rho)$, причем $\rho\to\infty$. Без ограничения общности будем считать $\theta=1$ и $\tau=\rho$. В этом случае величина ρ определяет загрузку системы.

Если обозначить ν число требований в системе в стационарном режиме, то есть основания полагать [9–11], что случайная величина $(\nu-\mathbf{M}\nu)/\sqrt{\mathbf{D}\nu}$ имеет асимптотически нормальное распределение при $\rho \to \infty$. В настоящей статье исследована возможность применения χ^2-критерия, а также критерия, основанного на равномерной метрике (теорема Гливенко), для проверки гипотезы асимптотической нормальности, когда статистические данные генерируются методом Монте-Карло. Описаны сложные проблемы, возникающие при использовании указанных критериев.

АЛГОРИТМ ПРОВЕРКИ ГИПОТЕЗЫ АСИМПТОТИЧЕСКОЙ НОРМАЛЬНОСТИ χ^2-КРИТЕРИЕМ

Сбор статистических данных методом Монте-Карло организуем так же, как описано в [12]. Предположим, что система обслуживания $GI/G/\infty$ в момент $t=0$ находится в стационарном режиме (иначе говоря, этот момент является бесконечно удаленным от момента начала функционирования системы). Число ν требований, обслуживаемых в стационарном режиме, определяется моментами их поступления до этого момента и длительностями обслуживания. Начиная с момента $t=0$, направим временну́ю ось не в «будущее», а в «прошлое». Моменты поступления требований — это стационарный рекуррентный поток, определяемый последовательностью независимых случайных величин $\{\xi_1^{(0)},\ \xi_i,\ i \ge 2\}$, где все $\{\xi_i,\ i \ge 2\}$ имеют одно и то же распределение $A(x)$, а $\mathbf{P}\{\xi_1^{(0)} < x\} = \int_0^x [1 - A(u)]\, du$, $x > 0$, поскольку предполагается $\theta = 1$. Если обозначить $\{\eta_i,\ i = 1, 2, \ldots\}$ последовательность независимых одинаково распределенных случайных величин с функцией распределения $B(x)$, то количество требований ν в стационарный момент времени определяется равенством

$$\nu = \sum_{n=1}^{\infty} I(\eta_n > \xi_1^{(0)} + \xi_2 + \ldots + \xi_n)\,, \tag{1}$$

где $I(\cdot)$ — индикатор соответствующего события. Цель исследования — проверить, не противоречат ли численные данные гипотезе: при $\rho \to \infty$ нормированная случайная величина $(\nu - \mathbf{M}\nu)/\sqrt{\mathbf{D}\nu}$ имеет асимптотически стандартное нормальное распределение.

Очевидно, что с увеличением n вклад соответствующих членов ряда (1) стремится к нулю. Поэтому без сколь-нибудь заметной потери точности ряд (1) можно аппроксимировать суммой

$$\nu(\varepsilon) = \sum_{n=1}^{m(\varepsilon)} I(\eta_n > \xi_1^{(0)} + \xi_2 + \ldots + \xi_n)\,, \tag{2}$$

где $m(\varepsilon) = \max\{j\colon\ \xi_1^{(0)} + \xi_2 + \ldots + \xi_j \le W(\varepsilon)\}$, $W(\varepsilon) = \inf\{u : 1 - G(u) < \varepsilon\}$, а ε — некоторый пороговый уровень, определяющий значения вероятностей событий, которыми будем пренебрегать (например, $\varepsilon = 10^{-7}$). Иначе говоря, каждое событие $\{\eta_n > \xi_1^{(0)} + \xi_2 + \ldots + \xi_n\}$, $n \ge m(\varepsilon)$, имеет вероятность меньшую, чем ε.

Если методом Монте-Карло моделировать независимые последовательности случайных величин $\{\xi_1^{(0)},\ \xi_i,\ 2 \le i \le m(\varepsilon)\}$ и $\{\eta_i\ ,\ 2 \le i \le m(\varepsilon)\}$, то по формуле (2) можно строить независимые реализации $\{\nu^{(1)}, \ldots, \nu^{(N)}\}$ случайной величины $\nu(\varepsilon)$, а также выборочные значения математического ожидания и дисперсии

$$\bar{\nu}(N)=\frac{1}{N}\sum_{i=1}^{N}\nu^{(i)},\quad \bar{\sigma}^2(N)=\frac{1}{N-1}\left\{\sum_{i=1}^{N}[\nu^{(i)}]^2-N[\bar{\nu}(N)]^2\right\}. \tag{3}$$

Построим новую последовательность

$$x^{(i)}=\frac{\nu^{(i)}-\bar{\nu}(N)}{\bar{\sigma}(N)},\ i=1,2,\ldots,N. \tag{4}$$

Введем гипотезу H_0: наблюдения $\{x^{(i)},\ i=1,\ldots,N\}$ — это реализации нормально распределенной случайной величины с математическим ожиданием 0 и дисперсией 1. В настоящей статье предлагается подход, основанный на χ^2-критерии, а также критерии равномерной метрики, позволяющий оценить изменение степени доверия к данной гипотезе при увеличении ρ и N.

Для проверки гипотезы H_0 воспользуемся χ^2-критерием Пирсона. Это наиболее универсальный критерий, широко применяемый на практике для проверки разнообразных гипотез. Он основан на дискретизации распределения наблюдаемой случайной величины. В то же время этот критерий предоставляет исследователю значительную свободу действий, в частности, при выборе областей, на которые разбивается область наблюдения $R=(-\infty,+\infty)$.

Пусть значения ρ и N фиксированы. Для формирования областей проведем предварительные N наблюдений $\{\nu^{(1)},\ \ldots\ ,\ \nu^{(N)}\}$ и по формуле (3) вычислим $\bar{\nu}(N)$, $\bar{\sigma}^2(N)$. Пусть n_{gr} — некоторое натуральное число. Область значений случайной величины $\nu(\varepsilon)$ разобьем на области таким образом, чтобы в каждую область попадало не менее n_{gr} наблюдений. Предположим, что таких областей оказалось r:

$$C_1=\{0,1,\ldots,l_1\},\ C_i=\{l_{i-1}+1,\ldots,l_i\},\ i=2,3,\ldots,r-1,$$

$$C_r=\{l_{r-1}+1,l_{r-1}+2,\ldots\},$$

где $0\le l_1<l_2<\ldots<l_{r-1}<\infty$. Тогда $R=\bigcup_{k=1}^{r}U_k$, где $U_k=(u_{k-1},u_k)$, $k=1,\ldots,r$, $u_0=-\infty$, $u_k=(l_k+0{,}5-\bar{\nu}(N))/\bar{\sigma}(N)$, $k=1,\ldots,r-1$, $u_r=+\infty$. Вычислим теоретические вероятности попадания случайной величины $\nu(\varepsilon)$ в области $\{C_k\}$ при условии, что гипотеза H_0 справедлива:

$$p_1=\Phi(u_1),\ p_k=\Phi(u_k)-\Phi(u_{k-1}),\ k=2,\ldots,r-1,\ p_r=1-\Phi(u_{r-1}),$$

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-v^2/2}\,dv.$$

Пусть разбиение R на области фиксированное. Построим новые реализации $\{\nu^{(1)},\ldots,\nu^{(N)}\}$ случайной величины $\nu(\varepsilon)$. Предположим, что в область C_k попало μ_k наблюдений, $\mu_1+\ldots+\mu_r=N$. Тогда составляем сумму Пирсона

$$\Delta_N^{(r)}=N\sum_{k=1}^{r}\frac{(\mu_k/N-p_k)^2}{p_k}. \tag{5}$$

Согласно теореме Пирсона асимптотическое распределение (при $N\to\infty$) случайной величины $\Delta_N^{(r)}$ является χ^2-распределением с $r-3$ степенями свободы. Это распределение обозначим $W_{r-3}(z)$. Пусть γ — уровень значимости критерия. Тогда по таблицам для χ^2-распределения находим пороговый уровень

$z_{r-3}(\gamma)$, являющийся решением уравнения $W_{r-3}(z)=1-\gamma$. Если вычисления согласно (5) показали, что

$$\Delta_N^{(r)} \le z_{r-3}(\gamma), \qquad (6)$$

то полученные статистические данные не противоречат гипотезе H_0. Если неравенство (6) не выполняется, то следует принять альтернативную гипотезу H_1.

Если значения ρ и N фиксированы, то по одной серии испытаний весьма сложно судить о том, насколько асимптотическое распределение близко к нормальному. При одних статистических данных, полученных методом Монте-Карло, неравенство (6) имеет место, а при других — нет. Поэтому более объективную информацию может дать последовательность из K независимых серий испытаний. Выполнение неравенства (6) назовем «успехом». Количество успехов обозначим $\sigma(\rho, N, K)$. Если гипотеза H_0 справедлива, то с возрастанием ρ следует ожидать неубывание и $\sigma(\rho, N, K)$ при фиксированных N и K. Данную особенность наглядно демонстрирует приведенный далее численный пример.

Пример 1. Рассмотрим систему обслуживания $GI/G/\infty$, определяемую следующими распределениями:

$$A(x)=1-\exp\{-(\rho\,\alpha\,x)^2\},\ B(x)=1-\exp\{-(\alpha\,x)^2\},\ x\ge 0,$$

где $\alpha=\sqrt{\pi}/2$. В этом случае $\theta=1$ и $\tau=1$. Выберем $\varepsilon=10^{-7}$. Это означает, что в сумме (2) оставлены лишь события, вероятности которых не менее ε. Исследуем поведение количества успехов $\sigma(\rho, N, K)$ в зависимости от ρ и N. Выберем $K=10$. При группировании статистических данных для формирования областей в качестве n_{gr} выберем $n_{gr}=N/50$. Иначе говоря, величина каждой группы пропорциональна общему количеству наблюдений, причем количество степеней свободы в χ^2-распределении не превысит 50. Результаты проведенного численного эксперимента представлены в табл. 1 (уровень значимости $\gamma=0.05$).

При одном и том же значении ρ с изменением N может изменяться количество областей, что отражено в последних двух колонках табл. 1. При фиксированном объеме статистических данных N количество успехов монотонно неубывает, что вполне соответствует выдвинутой гипотезе: с увеличением загрузки стационарное распределение числа требований в системе стремится к нормальному распределению. В то же время на первый взгляд может показаться странным, что с увеличением объема выборки (при фиксированном ρ) количество успехов уменьшается. Чтобы объяснить этот феномен отметим, что рассматривается асимптотика. Иначе говоря, при любом фиксированном ρ асимптотическая оценка p_k вероятности попадания во множество C_k является смещенной.

Таблица 1. Значения $\sigma(\rho, N, K)$ как функции от ρ и N

ρ	N					r	$z_{r-3}(\gamma)$
	10 000	20 000	40 000	80 000	160 000		
100	8	3	0	0	0	26	35.17
200	8	6	0	1	0	32; 34	42.56; 44.99
400	9	8	7	4	1	38; 39	49.80; 51,00
800	9	9	8	6	1	35; 37	46.19; 48.60
1 600	9	10	9	8	2	40; 41	52.19; 53.38

Рассмотрим отдельное слагаемое в формуле (5). Если значение p_k лежит внутри доверительного интервала с центром μ_k / N, то квадрат разности $(\mu_k / N - p_k)^2$ пропорционален $1/N$. Учитывая множитель N перед суммой, видим, что данное слагаемое вносит лишь незначительный вклад в сумму, слабо зависящий от значения N. При увеличении N сужается доверительный интервал и поэтому существенно возрастает вероятность выхода p_k за пределы доверительного интервала (ввиду смещенности асимптотической оценки). Учет множителя N перед суммой приводит к значительным «выбросам» отдельных слагаемых, и в результате сумма в правой части (5) может существенно превышать предельно допустимое значение $z_{r-3}(\gamma)$. Так, например, при $N = 160000$ значение $\Delta_N^{(r)}$ изменяется от 179.2 до 297.2 (при $\rho = 100$) и от 95.2 до 145.8 (при $\rho = 200$). При дальнейшем увеличении ρ значение $\Delta_N^{(r)}$ «в среднем» продолжает убывать, что вполне соответствует гипотезе H_0.

АЛГОРИТМ ПРОВЕРКИ ГИПОТЕЗЫ АСИМПТОТИЧЕСКОЙ НОРМАЛЬНОСТИ КРИТЕРИЕМ РАВНОМЕРНОЙ МЕТРИКИ

В случае если выборка $\{x^{(i)},\ i = 1, \ldots, N\}$ представляет собой независимые реализации некоторой случайной величины и существует обоснованное предположение, что ее распределение является нормальным, то для проверки данной гипотезы вполне подходит критерий Колмогорова, основанный на асимптотическом (при $N \to \infty$) распределении равномерной метрики между эмпирическим и предполагаемым распределениями. Однако $\{x^{(i)},\ i = 1, \ldots, N\}$ в определенном смысле зависимые случайные величины (см. (4)), которые могут принимать одинаковые значения. Кроме того, при фиксированном ρ эмпирическое распределение не является несмещенным. Тем не менее критерий равномерной метрики вполне можно использовать для статистической проверки гипотезы о сходимости к нормальному распределению. Суть данного подхода состоит в следующем.

Пусть $X = \{x^{(i)},\ i = 1, \ldots, N\}$ — выборка, полученная согласно (4). Построим эмпирическую функцию распределения. Упорядочив значения $\{x^{(i)}\}$ в порядке возрастания, получим вариационный ряд: $y_1 < y_2 < \ldots < y_r$, причем значение y_i в выборке X встречается k_i раз, $k_1 + \ldots + k_r = N$. Эмпирическая функция распределения строится по формуле

$$\hat{F}_N(t; X) = \begin{cases} 0, & \text{если } t \le y_1, \\ \dfrac{1}{N} \sum\limits_{i=1}^{j} k_i, & \text{если } y_j < t \le y_{j+1},\ j = 1, \ldots, r-1, \\ 1, & \text{если } t > y_r. \end{cases}$$

Разность между эмпирической функцией распределения и теоретической (нормальное распределение) в равномерной метрике имеет вид

$$D_N(X) =$$

$$= \sup_{-\infty < t < +\infty} |\hat{F}_N(t; X) - \Phi(t)| = \max_{1 \le j \le r} \max \left\{ \left| \frac{1}{N} \sum_{i=1}^{j-1} k_i - \Phi(y_j) \right|, \left| \frac{1}{N} \sum_{i=1}^{j} k_i - \Phi(y_j) \right| \right\}.$$

Таблица 2. Значения $\varphi(\rho, N, K)$ как функции от ρ и N

ρ	N			
	2 500	10 000	40 000	160 000
100	0.041	0.037	0.035	0.035
200	0.032	0.026	0.026	0.024
400	0.028	0.022	0.018	0.018
800	0.025	0.016	0.015	0.013
1 600	0.021	0.014	0.011	0.010

Если имеет место сходимость к нормальному распределению, то естественно ожидать (теорема Гливенко), что

$$\lim_{\rho\to\infty} \lim_{N\to\infty} \mathbf{M}D_N(X) = 0. \tag{7}$$

Следующий пример показывает, насколько реальные статистические данные согласуются с соотношением (7).

Пример 2. Рассмотрим систему обслуживания с теми же характеристиками, что и в примере 1. Предположим, что X_i, $i=1, \ldots, K$, — независимые реализации случайного вектора X при фиксированных ρ и N. В качестве оценки для $\mathbf{M}D_N(X)$ выберем $\varphi(\rho, N, K) = \frac{1}{K}\sum_{i=1}^{K} D_N(X_i)$. Тенденция к изменению $\varphi(\rho, N, K)$ при изменении ρ и N наблюдается даже для небольших значений K. Выберем $K=10$. Результаты проведенного численного эксперимента представлены в табл. 2.

Приведенные в табл. 2 численные данные полностью соответствуют соотношению (7): с возрастанием объема N выборки уменьшается средняя погрешность аппроксимации в равномерной метрике (но не стремится к нулю); дальнейшее убывание средней погрешности наблюдается при увеличении загрузки системы. Таким образом, можно сделать вывод: статистические данные не противоречат гипотезе об асимптотически нормальном распределении стационарного распределения числа требований при $\rho \to \infty$.

ЗАКЛЮЧЕНИЕ

В данной статье для системы обслуживания $GI/G/\infty$ в условиях большой загрузки исследовалась возможность применения χ^2-критерия, а также критерия, основанного на равномерной метрике, для проверки статистической гипотезы об асимптотически нормальном распределении стационарного распределения числа требований. Показано, что классические варианты этих критериев не дают ответа на поставленный вопрос. Численный пример подтвердил возможность практического применения предложенной методики для проверки указанной гипотезы для систем обслуживания с конкретными характеристиками.

СПИСОК ЛИТЕРАТУРЫ

1. Королюк В.С., Турбин А.Ф. Математические основы фазового укрупнения сложных систем. Киев: Наук. думка, 1978. 217 с.
2. Kovalenko I.N. Approximation of queues via small-parameter method. *Advances in Queueing.* Boca Raton: CRC Press, 1995. P. 481–506.
3. Kovalenko I.N. Ergodic and light-traffic properties of a complex repairable system. *Math. Meth. Operat. Res.* 1997. Vol. 45, N 2. P. 387–409.

4. Korolyuk V.S., Korolyuk V.V. Stochastic models of systems. Boston: Kluwer, 1999. 185 p.
5. Kovalenko I.N., Atkinson J.B., Mikhalevich K.V. Three cases of light-traffic insensitivity of the loss probability in a $GI / G / m / 0$ loss system to the shape of the service time distribution. *Queueing Systems.* 2003. Vol. 45, N 3. P. 245–271.
6. Меликов А.З., Пономаренко Л.А., Паладюк В.В. Телетрафик: модели, методы, оптимизация. Киев: Политехника, 2007. 256 с.
7. Меликов А.З., Пономаренко Л.А., Фаттахова М.И. Управление мультисервисными сетями связи с буферными накопителями. Киев: НАУ-друк, 2008. 156 с.
8. Ponomarenko L., Kim C.S., Melikov A. Performance analysis and optimization of multi-traffic on communication networks. New York: Springer, 2010. 208 p.
9. Боровков А. А. Эргодичность и устойчивость случайных процессов. Москва: Эдиториал УРСС, 1999. 450 с.
10. Ливинская А.В., Лебедев Е.А. Предельная теорема для перегруженных многоканальных сетей. *Кибернетика и системный анализ.* 2012. № 6. С. 106–113.
11. Lebedev E., Livinska G. Gaussian approximation of multi-channel networks in heavy traffic. *Communications in Computer and Information Science.* 2013. N. 356. P. 122–130.
12. Кузнєцов І.М. Використання методу Монте-Карло для статистичної перевірки асимптотичної нормальності стаціонарного розподілу кількості вимог у системі $GI / G / \infty$ у випадку великого завантаження. *Доповіді НАН України*. 2016. № 5. С. 30–35.

Надійшла до редакції 20.12.2016

І.М. Кузнєцов

СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗИ АСИМПТОТИЧНОЇ НОРМАЛЬНОСТІ СТАЦІОНАРНОГО РОЗПОДІЛУ КІЛЬКОСТІ ВИМОГ У СИСТЕМІ $GI / G / \infty$ В УМОВАХ ВЕЛИКОГО ЗАВАНТАЖЕННЯ

Анотація. Розглянуто систему обслуговування $GI / G / \infty$ в умовах великого завантаження. Статистичні дані згенеровано методом Монте-Карло. Досліджено можливість застосування деяких статистичних критеріїв для перевірки гіпотези асимптотичної нормальності розподілу кількості вимог у системі в стаціонарному режимі. Розглянуто числові приклади.

Ключові слова: система обслуговування, велике завантаження, стаціонарний розподіл, статистична гіпотеза, асимптотично нормальний розподіл, метод Монте-Карло.

I.N. Kuznetsov

STATISTICAL TESTING OF THE HYPOTHESIS THAT THE NUMBER OF CUSTOMERS IN QUEUEING SYSTEM $GI / G / \infty$ HAS ASYMPTOTICALLY NORMAL DISTRIBUTION IN HEAVY TRAFFIC

Abstract. The queueing system $GI / G / \infty$ in heavy traffic is considered. Statistical data are generated due to Monte Carlo simulation. The author considers the possibility of applying some statistical criteria to test the hypothesis that the number of customers has asymptotic normal distribution in steady state. Numerical examples are considered.

Keywords: queueing system, heavy traffic, steady-state distribution, statistical hypothesis, asymptotically normal distribution, Monte Carlo method.

Кузнецов Игорь Николаевич,
кандидат физ.-мат. наук, старший преподаватель Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского», e-mail: sea_hawk@icloud.com.