УДК 519.872

**И.Н. КУЗНЕЦОВ, А.А. ШУМСКАЯ**

# ОЦЕНКА СТАЦИОНАРНЫХ ВЕРОЯТНОСТЕЙ СОСТОЯНИЙ СИСТЕМЫ ОБСЛУЖИВАНИЯ $\bullet / G / \infty$ ПРИ РАЗЛИЧНЫХ ВИДАХ ВХОДЯЩЕГО ПОТОКА ТРЕБОВАНИЙ

**Аннотация.** Рассмотрены пять моделей входящего потока существенно более сложной структуры, чем пуассоновский, когда стационарные вероятности состояний системы $\bullet / G / \infty$ находятся в явном виде (распределение Пуассона). Для представленных моделей сочетание распределения Пуассона (аналитическая часть) со статистическим моделированием (статистическая часть) позволяет находить стационарные вероятности состояний ускоренным моделированием. Точность полученных оценок проиллюстрирована численными примерами.

**Ключевые слова:** система обслуживания, стационарные вероятности состояний, нестационарный пуассоновский процесс, регенерирующий процесс, полумарковский процесс, процесс скоплений, несмещенная оценка, относительная погрешность.

## ВВЕДЕНИЕ

Нахождению стационарных вероятностей состояний (а также условий их существования) различных систем обслуживания, начиная от «классических» и кончая весьма экзотическими, посвящено множество публикаций. Некоторое представление о многообразии исследований дают работы [1–6]. Столь же разнообразными являются и методы исследований, среди которых отметим следующие:

- аналитические, основанные на нахождении условий инвариантности стационарных распределений состояний относительно вида распределения длительности обслуживания [5, 7–9];
- метод малого параметра [10–12];
- метод асимптотической инвариантности [13–15];
- метод фазового укрупнения состояний [16, 17] и связанный с ним метод декомпозиции [18–20];
- асимптотический метод в условиях большой загрузки [21, 22];
- некоторые методы ускоренного моделирования [23–27].

Здесь не упоминается бесчисленное множество вариаций указанных общих подходов.

## ПОСТАНОВКА ЗАДАЧИ

В настоящей статье в качестве модели выбрана система массового обслуживания $\bullet / G / \infty$, состоящая из бесконечного числа обслуживающих каналов. Обслуживание начинается в момент поступления требования, время обслуживания имеет функцию распределения $G(x)$ и среднюю длительность $\tau$. Если входящий поток требований является пуассоновским с интенсивностью $\lambda$, то результат относится к «классическим», т.е. число $\nu$ занятых каналов в стационарном режиме имеет распределение Пуассона с параметром $\lambda\tau$:

$$P_k = \mathbf{P}\{\nu = k\} = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}, \ k \geq 0. \qquad (1)$$

Данная формула характерна тем, что стационарные вероятности состояний не зависят от функционального вида $G(x)$ при фиксированном среднем $\tau$. Возникает естественный вопрос, при каких иных потоках требований стационарные вероятности состояний вычисляются с использованием распределения Пуассона: либо непосредственно по явной аналитической формуле, либо с дополнительным этапом моделирования параметра распределения Пуассона

$$P_k = \mathbf{P}\{\nu = k\} = \mathbf{M}\left\{\frac{\theta^k}{k!}\, e^{-\theta}\right\}, \ k \ge 0\,, \tag{2}$$

где $\theta$ — случайный параметр, подлежащий моделированию. Поиск ответа на данный вопрос и является целью исследований настоящей статьи.

Рассматриваются пять моделей входящих потоков. Однако лишь для первой, наиболее простой и, по-видимому, хорошо изученной модели, стационарные вероятности находятся по явной аналитической формуле. В остальных случаях используется формула (2): аналитическому вычислению предшествует статистическое моделирование параметра $\theta$. Численные примеры показывают, что использование явных аналитических формул для вычисления малых вероятностей в случае большой загрузки системы позволяет существенно понизить дисперсию оценок при статистическом моделировании системы. Предложенный поход можно использовать и для других моделей входящих потоков, не рассмотренных в настоящей статье.

### ВСПОМОГАТЕЛЬНАЯ ФОРМУЛА

Предположим, что система с бесконечным числом обслуживающих каналов в момент $t^* = 0$ находится в стационарном режиме (иначе говоря, момент $t^*$ является бесконечно удаленным от момента начала функционирования системы). Число $\nu$ обслуживаемых в момент $t^*$ требований определяется моментами их поступления до момента $t^*$ и длительностями обслуживания. Начиная с момента $t^* = 0$, направляем временну́ю ось не в «будущее», а в «прошлое».

Пусть в систему, находящуюся в стационарном режиме, поступает нестационарный пуассоновский поток, т.е. нестационарный ординарный поток без последействия с интенсивностью $\mu(u)$, $u \ge 0$. Поскольку временна́я ось направлена в «прошлое», $\mu(u)$ является интенсивностью поступления требования в момент $t^* - u$. Для данной весьма простой модели число требований в системе в момент $t^*$ имеет распределение Пуассона.

**Лемма 1.** Если $\theta = \int\limits_0^\infty \mu(u)\,[1 - G(u)]\,du < \infty$, то $P_k = \mathbf{P}\{\nu = k\} = \dfrac{\theta^k}{k!}\, e^{-\theta}$, $k \ge 0$.

**Доказательство.** Обозначим $Q_k(u)$ вероятность того, что в интервале $(t^* - u, t)$ поступит $k$ требований, обслуживание которых не окончится к моменту $t^*$. Тогда

$$P_k = \lim_{u \to \infty} Q_k(u)\,. \tag{3}$$

Функции $\{Q_k(u),\ k \ge 0\}$ удовлетворяют следующим соотношениям:

$$Q_0(u + h) = Q_0(u)\,\{1 - \mu(u)\,[1 - G(u)]\,h + o(h)\},$$

$$Q_k(u + h) = Q_k(u)\,\{1 - \mu(u)\,[1 - G(u)]h + o(h)\} + Q_{k-1}(u)\,\mu(u)\,[1 - G(u)]h + o(h),\ k \ge 1\,.$$

Таким образом, получаем систему дифференциальных уравнений

$$Q_0'(u) = -\mu(u)\,[1 - G(u)]\,Q_0(u)\,,$$

$$Q_k'(u) = -\mu(u)\,[1 - G(u)]\,Q_k(u) + Q_{k-1}(u)\,\mu(u)\,[1 - G(u)],\ k \ge 1\,,$$

с начальными условиями $Q_0(0) = 1$, $Q_k(0) = 0$, $k \ge 1$.

Сделаем замену $Q_k(u)=S_k(u)\exp\left\{-\int\limits_0^u \mu(v)[1-G(v)]\,dv\right\},\ k\ge 0$. В результате элементарных преобразований для $\{S_k(u)\}$ имеем систему уравнений

$$S_0'(u)=0,$$

$$S_k'(u)=S_{k-1}(u)\,\mu(u)\,[1-G(u)],\ k\ge 1,$$

с начальными условиями $S_0(0)=1,\ S_k(0)=0,\ k\ge 1$. Используя метод математической индукции, имеем

$$S_k(u)=\frac{1}{k!}\left\{\int\limits_0^u \mu(v)[1-G(v)]\,dv\right\}^k,\quad k\ge 0.$$

Таким образом,

$$Q_k(u)=\frac{1}{k!}\left\{\int\limits_0^u \mu(v)[1-G(v)]\,dv\right\}^k \exp\left\{-\int\limits_0^u \mu(v)[1-G(v)]\,dv\right\},\ k\ge 0. \qquad (4)$$

Утверждение леммы 1 следует из соотношений (3), (4).

Если $\mu(u)\equiv\lambda$, то получим формулу (1). Рассмотрим несколько моделей входящих потоков требований, когда стационарное распределение числа требований в системе находится с помощью соотношения (2). Лишь для первой модели параметр $\theta$ является детерминированным, для всех остальных — случайным, подлежащим моделированию.

#### ПЕРИОДИЧЕСКАЯ ИНТЕНСИВНОСТЬ С ПОСТОЯННЫМ ПЕРИОДОМ

Пусть в систему обслуживания поступает нестационарный пуассоновский поток с интенсивностью $\lambda(t),\ t\ge 0$, которая предполагается конечной и периодической с постоянным периодом $T$, т.е. $\lambda(t+T)=\lambda(t),\ t\ge 0$. В качестве момента $t^*=0$ выберем бесконечно удаленный $(n\to\infty)$ момент $nT+\tau$. Иначе говоря, процесс поступления требований к моменту $t^*$ уже длится бесконечно долго, причем очередной период начался $\tau$ единиц времени тому назад. Пусть $\nu(\tau)$ — число требований, обслуживаемых в системе в момент $t^*$. Обозначим $P_k(\tau)=\mathbf{P}\{\nu(\tau)=k\},\ k\ge 0$ (функции $\{P_k(\tau)\}$ являются периодическими с периодом $T$), а также

$$\mu(u;\tau)=\lambda\left(T+\tau-u+T\left[\frac{u-\tau}{T}\right]\right),\quad u\ge 0,\ 0\le\tau\le T,$$

где $\left[\frac{u-\tau}{T}\right]$ — целая часть соответствующего числа. При фиксированном $\tau\ge 0$ функция $\mu(u;\tau)$ задает интенсивность поступления требования в момент $t^*-u$. Пусть

$$\theta(\tau)=\int\limits_0^\infty \mu(u;\tau)[1-G(u)]\,du,\ 0\le\tau\le T.$$

Из сформулированной ранее леммы 1 вытекает следующее утверждение.

**Теорема 1.** Если $\sup\limits_{\tau\in[0,T]}\theta(\tau)<\infty$, то

$$P_k(\tau)=\frac{[\theta(\tau)]^k}{k!}e^{-\theta(\tau)},\ k\ge 0.$$

Данный случай отличается от предыдущего тем, что период $T$ является случайным, т.е. задана пара $\{T, \lambda(t), 0 \le t < T\}$, где $T$ — случайная величина с функцией распределения $H(x)$ и средним значением $\rho > 0$. Предполагается, что функция $H(x)$ имеет абсолютно непрерывную компоненту, а $\lambda(t)$, $t \ge 0$, — некоторая детерминированная функция (интенсивность поступления требования). Иначе говоря, входящий поток является нестационарным пуассоновским, интенсивность которого определяется простым процессом восстановления. Предположим, что в рассматриваемый момент $t^* = 0$ процесс восстановления стационарный. Как и ранее, временную ось направим в «прошлое». Именно интенсивность поступления требований до момента $t^*$ влияет на количество $\nu$ требований в момент $t^*$. Обозначим $P_k = \mathbf{P}\{\nu = k\}$, $k \ge 0$. Пусть $\{z_i,\ i \ge 1\}$ — фиксированные моменты восстановления, отсчитываемые от момента $t^* = 0$ в «прошлое»: $t_1 = z_1$, $t_k = z_k - z_{k-1}$, $k \ge 2$. Обозначим

$$\theta_n(t_1, \ldots, t_n) = \int_0^{z_n} \mu_n(u; t_1, \ldots, t_n)[1 - G(u)]\, du, \tag{5}$$

где

$$\mu_n(u; t_1, \ldots, t_n) = \begin{cases} \lambda(z_1 - u), & \text{если} \quad 0 \le u \le z_1, \\ \lambda(z_k - u), & \text{если} \quad z_{k-1} < u \le z_k, \quad k = 2, \ldots, n, \end{cases} \tag{6}$$

является интенсивностью поступления требования в момент $u$ при известных длительностях $t_1, \ldots, t_n$ последних $n$ периодов восстановления. Справедливо следующее утверждение.

**Теорема 2.** Предположим, что выполнено условие

$$\sup_{n \ge 1} \sup_{t_1 \ge 0, \ldots, t_n \ge 0} \theta_n(t_1, \ldots, t_n) < \infty. \tag{7}$$

Тогда вероятности $\{P_k,\ k \ge 0\}$ вычисляются по формуле

$$P_k = \lim_{n \to \infty} \underbrace{\int_0^\infty \ldots \int_0^\infty}_{n-1} dH(t_2) \ldots dH(t_n) \int_0^\infty \frac{[\theta(t_1, \ldots, t_n)]^k}{k!} e^{-\theta(t_1, \ldots, t_n)} \frac{1 - H(t_1)}{\rho} dt_1, \quad k \ge 0. \tag{8}$$

**Замечание 1.** Конечность среднего времени обслуживания и равномерная ограниченность интенсивности $\lambda(t)$, $t \ge 0$, являются достаточным условием выполнения соотношения (7). В то же время выполнение данного соотношения возможно и при значительно более слабых условиях (при этом случай монотонно возрастающей интенсивности $\lambda(t)$, $t \ge 0$, не исключен).

С формальной точки зрения формула (8) является явной аналитической, позволяющей вычислять вероятности $\{P_k\}$. Однако в случае большой загрузки системы обслуживания для достижения приемлемой точности оценки необходимо рассматривать значения $n > 100$. Столь высокая кратность интегрирования не позволяет использовать каких-либо численных методов (даже на самых современных вычислительных комплексах). В этом случае наиболее эффективным и простым для использования является статистическое моделирование.

Пусть $\varepsilon$ — некоторое малое число (граница отсечения), $W(\varepsilon) = \inf\{u : 1 - G(u) < \varepsilon\}$. Иначе говоря, если требование поступит до момента $t^* - W(\varepsilon)$, то вероятность его наличия в системе в момент $t^*$ не превышает $\varepsilon$. При моделировании ограничимся лишь рассмотрением требований, поступивших в интервале $(t^* - W(\varepsilon), t^*)$. Алгоритм построения асимптотически несмещенной (при $\varepsilon \to 0$) оценки $\hat{P}^{(1)}(A; \varepsilon)$ в одной реализации для $P(A) = \sum_{k \in A} P_k$, $A \subset \{0, 1, 2, \ldots\}$, основан на формуле (8) и формулируется следующим образом.

1. Согласно распределению с плотностью $\frac{1-H(u)}{\rho}$, $u \geq 0$, моделируем момент $t_1$ начала периода, длящегося в момент $t^*$.

2. Согласно распределению $H(u)$ моделируем длительности $t_2, t_3, \ldots, t_n$ периодов, где $n = \max\{m: t_1 + \ldots + t_m \leq W(\varepsilon)\}$.

3. По формулам (5), (6) вычисляем $\theta(t_1, \ldots, t_n)$.

4. В качестве оценки $\hat{P}^{(1)}(A;\varepsilon)$ выбираем

$$\hat{P}^{(1)}(A;\varepsilon) = \sum_{k \in A} \frac{[\theta(t_1, \ldots, t_n)]^k}{k!} e^{-\theta(t_1, \ldots, t_n)}. \quad (9)$$

Число реализаций алгоритма, необходимых для построения оценки с заданными достоверностью и относительной погрешностью, определяется по известным формулам [28]. Выбором $\varepsilon$ можно добиться практически несмещенной оценки.

Аналитическая часть (9) алгоритма позволяет существенно уменьшить дисперсию оценки. Особенно это заметно, когда вероятность попадания во множество состояний $A$ мала. Рассмотрим численный пример.

**Пример 1.** Исследуем среднее количество $s$ требований в системе в стационарном режиме и вероятность нахождения в системе не менее $r$ требований, т.е. $A = \{r, r+1, \ldots\}$. Сравним результаты, полученные двумя методами: непосредственным моделированием системы (метод Монте-Карло) и ускоренным моделированием, основанным на формулах (5), (6) и (9). Проиллюстрируем преимущество ускоренного моделирования при возрастании значений $r$. В качестве параметров модели выберем следующие: $T$ — случайная величина, равномерно распределенная в промежутке $[0, 0.5]$, $\rho = 0.25$; $\lambda(t) = 100\,t$, $t \geq 0$; $G(x) = 1 - e^{-\pi x^2/4}$, $x \geq 0$, $\tau = 1$; $r = 20 + 5\,i$, $i = 0, 1, \ldots, 6$.

Выберем $\varepsilon = 10^{-7}$. В табл. 1 использованы следующие обозначения (нижние индексы MC и fast означают, что соответствующие оценки получены методом Монте-Карло и ускоренным моделированием):

- $\hat{S}_{\mathrm{MC}}$ и $\hat{S}_{\mathrm{fast}}$ — оценки среднего числа требований в системе;
- $\hat{R}(\hat{S}_{\mathrm{MC}})$ и $\hat{R}(\hat{S}_{\mathrm{fast}})$ — оценки относительной погрешности для $\hat{S}_{\mathrm{MC}}$ и $\hat{S}_{\mathrm{fast}}$;
- $\hat{P}_{\mathrm{MC}}(r;\varepsilon)$ и $\hat{P}_{\mathrm{fast}}(r;\varepsilon)$ — оценки стационарных вероятностей пребывания в системе не менее $r$ требований;
- $\hat{R}(\hat{P}_{\mathrm{MC}}(r;\varepsilon))$ и $\hat{R}(\hat{P}_{\mathrm{fast}}(r;\varepsilon))$ — оценки относительной погрешности для $\hat{P}_{\mathrm{MC}}(r;\varepsilon)$ и $\hat{P}_{\mathrm{fast}}(r;\varepsilon)$.

Все приведенные в табл. 1 оценки построены по 100 000 реализациям алгоритма. При этом $\hat{S}_{\mathrm{MC}} = 16.07$, $\hat{R}(\hat{S}_{\mathrm{MC}}) = 0.24\%$, $\hat{S}_{\mathrm{fast}} = 16.67$, $\hat{R}(\hat{S}_{\mathrm{fast}}) = 0.13\%$.

**Таблица 1**

| Требования, $r$ | Оценки стационарных вероятностей и относительных погрешностей | | | |
|---|---|---|---|---|
| | $\hat{P}_{\mathrm{MC}}(r;\varepsilon)$ | $\hat{R}(\hat{P}_{\mathrm{MC}}(r;\varepsilon))$, % | $\hat{P}_{\mathrm{fast}}(r;\varepsilon)$ | $\hat{R}(\hat{P}_{\mathrm{fast}}(r;\varepsilon))$, % |
| 20 | 0.27 | 1.34 | 0.27 | 0.55 |
| 25 | $6.15 \cdot 10^{-2}$ | 3.18 | $6.10 \cdot 10^{-2}$ | 0.97 |
| 30 | $7.89 \cdot 10^{-3}$ | 9.13 | $7.79 \cdot 10^{-3}$ | 1.54 |
| 35 | $8.40 \cdot 10^{-4}$ | 28.08 | $6.17 \cdot 10^{-4}$ | 2.38 |
| 40 | $3.00 \cdot 10^{-5}$ | 148.67 | $2.97 \cdot 10^{-5}$ | 3.58 |
| 45 | $1.00 \cdot 10^{-5}$ | 257.50 | $9.21 \cdot 10^{-7}$ | 5.29 |
| 50 | 0 | – | $1.93 \cdot 10^{-8}$ | 7.33 |

Если оцениваемая вероятность не слишком мала, то оба метода дают близкие результаты. С увеличением $r$ точность оценок методом Монте-Карло заметно снижается и уже при $r \geq 40$ для достижения приемлемой относительной погрешности следует значительно увеличить количество реализаций. Эффект применения ускоренного моделирования очевиден: использование распределения Пуассона позволяет обеспечивать устойчивость моделирования в широком диапазоне изменения оцениваемой вероятности. Хотя с возрастанием $r$ и наблюдается рост относительной погрешности, однако он незначителен и слабо влияет на увеличение вычислительных затрат.

## ИНТЕНСИВНОСТЬ ВХОДЯЩЕГО ПОТОКА — РЕГЕНЕРИРУЮЩИЙ СЛУЧАЙНЫЙ ПРОЦЕСС

Предположим, что в систему обслуживания поступает дважды стохастический пуассоновский поток требований, интенсивность которого — случайный процесс, в данном случае регенерирующий. Иначе говоря, на вероятностном пространстве $(\Omega, \Im, \mathbf{P})$ определена пара $\{T(\omega),\ \lambda(t,\omega), 0 \le t < T(\omega)\}$, где $T(\omega)$ — неотрицательная случайная величина с функцией распределения $H(x)$ и конечным средним значением $\rho$, $H(+0)=0$, а $\lambda(t)=\lambda(t,\omega)$, $0 \le t < T(\omega)$, — обрывающийся случайный процесс, принимающий значения в измеримом пространстве $(\mathbb{R}^+, \mathfrak{R}^+)$, здесь $\mathbb{R}^+ = [0;\infty)$, а $\mathfrak{R}^+$ — $\sigma$-алгебра борелевых множеств из $\mathbb{R}^+$. Предполагается, что функция $H(x)$ имеет абсолютно непрерывную компоненту.

Пусть длительность периода регенерации $T(\omega)=\tau$ фиксирована, а $X_\tau$ — пространство функций, определенных на $[0,\tau)$ и принимающих значения в $\mathbb{R}^+$. Обозначим $U_\tau$ минимальную $\sigma$-алгебру подмножеств из $X_\tau$, содержащую все цилиндрические множества вида $\{\gamma(\cdot)\colon \gamma(t_1)\in A_1, \ \dots, \gamma(t_k)\in A_k\}$, $0 \le t_1 < \dots < t_k < \tau$, $A_1, \ \dots, A_k \in \mathbb{R}^+$, $k \ge 1$.

Вероятностная мера $\mathbf{P}\{\cdot\}$ на $\Im$ порождает соответствующее распределение $\mathbf{Q}_\tau\{\cdot\}$ на $U_\tau$.

Обозначим $\tau_i$, $i \ge 1$, последовательность независимых случайных величин с функцией распределения $H(x)$, а $\gamma_i(u)$, $0 \le u < \tau_i$, — соответствующий обрывающийся процесс, построенный согласно распределению $\mathbf{Q}_{\tau_i}\{\cdot\}$. При фиксированной последовательности $\{\tau_i\}$ интенсивность поступления требований в систему в момент $t \in (\tau_1 + \dots + \tau_{k-1}, \tau_1 + \dots + \tau_k]$ $(k=1,2,\dots)$ равна $\gamma_k(t-(\tau_1+\dots+\tau_{k-1}))$.

Рассмотрим теперь стационарный режим работы системы обслуживания. Обозначим $\tau_1^*$ промежуток времени от последнего момента регенерации перед $t^*=0$ до $t^*$ («недоскок»), а $\tau_0^*$ — время до следующего после $t^*=0$ момента регенерации («перескок»). Из теории восстановления известно, что $\tau_1^*$ имеет плотность распределения $\dfrac{1-H(x)}{\rho}$, $x \ge 0$. При фиксированном значении $\tau_1^* = t_1$ случайная величина $\tau_0^*$ имеет распределение $\dfrac{H(x+t_1)-H(t_1)}{1-H(t_1)}$, $x \ge 0$. Предположим, что построены реализации случайных величин $\tau_0^* = t_0$, $\tau_1^* = t_1$, $\tau_i = t_i$, $i \ge 2$, а также траектории процессов $\{\gamma_1(u),\ 0 \le u < t_0 + t_1\}$, $\{\gamma_i(u),\ 0 \le u < t_i\}$, $i \ge 2$. Тогда интенсивность поступления требования в момент $u \in (t_1 + \dots + t_{k-1}, t_1 + \dots + t_k]$ (фактически в момент $t^*-u$) будет

$$\mu(u;\{t_i\},\{\gamma_i(\cdot)\}) = \gamma_k(t_1 + \dots + t_k - u). \tag{10}$$

Обозначим

$$\theta_n(\{t_i\},\{\gamma_i(\cdot)\}) = \int_0^{t_1+\dots+t_n} \mu(u;\{t_i\},\{\gamma_i(\cdot)\})\,[1-G(u)]\,du. \tag{11}$$

Пусть $\nu$ — число требований в системе обслуживания в момент $t^*$, $P_k = \mathbf{P}\{\nu = k\}$, $k \ge 0$. Справедливо следующее утверждение.

**Теорема 3.** Предположим, что выполнены условия, гарантирующие существование приведенных интегралов. Тогда стационарные вероятности $\{P_k,\ k \geq 0\}$ состояний системы вычисляются по формуле

$$P_k = \lim_{n\to\infty} \underbrace{\int_0^\infty \ldots \int_0^\infty}_{n-1} dH(t_2)\ldots dH(t_n) \int_0^\infty \frac{1-H(t_1)}{\rho}\, dt_1 \int_0^\infty \frac{dH(t_0+t_1)}{1-H(t_1)} \times$$

$$\times \underbrace{\int_{X_{t_0+t_1}} \int_{X_{t_2}} \ldots \int_{X_{t_n}}}_{n} \frac{[\theta_n(\{t_i\},\{\gamma_i(\cdot)\})]^k}{k!} e^{-\theta_n(\{t_i\},\{\gamma_i(\cdot)\})} \times$$

$$\times \mathbf{Q}_{t_0+t_1}\{d\gamma_1(\cdot)\}\mathbf{Q}_{t_2}\{d\gamma_2(\cdot)\} \ldots \mathbf{Q}_{t_n}\{d\gamma_n(\cdot)\},\ k \geq 0.$$

**Замечание 2.** Наличие случайной величины $t_0$ в правой части последней формулы обусловлено тем, что мера $\mathbf{Q}_\tau\{\cdot\}$ определена на пространстве траекторий процесса на протяжении периода регенерации при фиксированной его длительности $\tau$. Поскольку $t_1$ — время, прошедшее с момента начала периода регенерации, длящегося в момент $t^*$, а $t_0$ — время от момента $t^*$ до его окончания, то общая длительность периода регенерации будет $t_1 + t_0$. Именно при фиксированной общей длительности и задается мера $\mathbf{Q}_{t_0+t_1}\{d\gamma_1(\cdot)\}$.

**Замечание 3.** Существование стационарных вероятностей $\{P_k,\ k \geq 0\}$ вытекает из следующих условий:

- среднее время обслуживания конечно;
- существует постоянная $h < \infty$ такая, что выполняется соотношение $\mathbf{P}\left\{\sup\limits_{0\leq t<\tau(\omega)} \lambda(t,\omega) \leq h\right\} = 1.$

Алгоритм ускоренного моделирования стационарных вероятностей аналогичен приведенному ранее.

1. Моделируем длительности $\{t_i\}$ периодов регенерации (отдельный алгоритм используется для моделирования длительности периода регенерации, оканчивающегося после момента $t^*$ и состоящего из двух частей).
2. Моделируем траектории функций $\{\gamma_k(\cdot)\}$, задающих интенсивности поступления требований.
3. По формулам (10), (11) вычисляем $\theta_n(\{t_i\},\{\gamma_i(\cdot)\})$ (количество $n$ периодов регенерации выбираем из того же условия, что и ранее).
4. Аналитически согласно распределению Пуассона с параметром $\theta_n(\{t_i\},\{\gamma_i(\cdot)\})$ вычисляем требуемые вероятности.

## ПОТОК ПОСТУПАЮЩИХ ТРЕБОВАНИЙ — ПРОЦЕСС СКОПЛЕНИЙ СПЕЦИАЛЬНОГО ВИДА

В качестве поступающего потока требований рассмотрим процесс скоплений специального вида [29].

Пусть $(\Omega, \Im, \mathbf{P})$ — некоторое вероятностное пространство и каждому $\omega \in \Omega$ соответствует счетная последовательность действительных чисел $\{s_i\}$ без предельных точек, называемая точечным процессом. Обозначим $N(A)$ число точек последовательности $\{s_i\}$, попавших в борелево множество $A \in \Re$, где $\Re$ — $\sigma$-алгебра борелевых множеств из $\mathbb{R} = (-\infty, \infty)$. Неотрицательная целочисленная счетно-аддитивная функция $N(\cdot)$ борелевых множеств называется считающей мерой. Ввиду того, что в последовательности $\{s_i\}$ не имеется предельных точек, эта мера с вероятностью 1 является конечной на любых ограниченных борелевых множествах из $\Re$. Обозначим $M$ множество таких мер, а $\aleph$ — соответствующую минимальную $\sigma$-алгебру подмножеств, порожденную цилиндрическими множествами вида $C = \{N(\cdot): N(A_1) = j_1, \dots, N(A_k) = j_k\}$, $k \ge 1$, $j_1 \ge 0, \dots, j_k \ge 0$, $A_1, \dots, A_k \in \Re$, $A_i \cap A_r = \varnothing$, $i \ne r$.

Поскольку между $\Omega$ и $M$ существует взаимно-однозначное соответствие, распределение $\mathbf{P}\{\cdot\}$ на $\Im$ порождает соответствующее распределение $\mathbf{Q}\{\cdot\}$ на $\aleph$. Предположим, что рассматриваемый точечный процесс является стационарным и «развернутым» в «прошлое», т.е. $\ldots < -s_2 < -s_1 < t^* = 0 < s_0 < s_{-1} < s_{-2} < \ldots$ Моменты $\{-s_i,\ i \ge 1,\ s_i,\ i \le 0\}$ — стационарный поток центров скоплений. Каждый центр скопления порождает свой нестационарный пуассоновский поток требований (члены скопления), задаваемый интенсивностью $\alpha(u)$, $u \ge 0$. При этом предположим $\overline{\alpha} = \int\limits_0^\infty \alpha(u)\, du < \infty$, что гарантирует конечность числа членов в одном скоплении с вероятностью 1 (распределение числа членов в скоплении является пуассоновским с параметром $\overline{\alpha}$). Суперпозиция всех потоков членов скоплений образует входящий поток требований. Если $\{s_i\}$ — фиксированная последовательность центров скоплений, то в систему обслуживания поступает нестационарный пуассоновский поток требований с интенсивностью в момент $u$ (фактически в момент $t^* - u$)

$$\mu(u; \{s_i\}) = \sum_{i \ge 1:\, s_i > u} \alpha(s_i - u). \tag{12}$$

Обозначим

$$\theta(\{s_i\}) = \int\limits_0^\infty \mu(u; \{s_i\})[1 - G(u)]\, du. \tag{13}$$

Пусть $\nu$ — число требований в системе обслуживания в момент $t^*$, $P_k = \mathbf{P}\{\nu = k\}$, $k \ge 0$. Справедливо следующее утверждение.

**Теорема 4.** Предположим, что выполнены условия, гарантирующие существование приведенного интеграла. Тогда стационарные вероятности $\{P_k,\ k \ge 0\}$ состояний системы вычисляются по формуле

$$P_k = \int\limits_M \frac{[\theta(\{s_i\})]^k}{k!} e^{-\theta(\{s_i\})} \mathbf{Q}\{dN(\cdot)\}, \quad k \ge 0. \tag{14}$$

Алгоритм ускоренного моделирования стационарных вероятностей $\{P_k,\ k \ge 0\}$ основан непосредственно на формулах (12)–(14).

### НЕСТАЦИОНАРНЫЙ ПУАССОНОВСКИЙ ПОТОК ТРЕБОВАНИЙ, УПРАВЛЯЕМЫЙ ПОЛУМАРКОВСКИМ ПРОЦЕССОМ

Пусть $\eta(t)$, $t \ge 0$, — непрерывный справа полумарковский процесс с конечным множеством состояний $E$. Переходные вероятности вложенной цепи Маркова обозначим $\{p_{ij}\}_{i,j \in E}$. Начальное состояние имеет распределение $\{p_i^{(0)}\}_{i \in E}$. Предполагается, что цепь Маркова является однородной, неприводимой и непериодической, что гарантирует существование ее эргодического распределения $\{\pi_i\}_{i \in E}$. Кроме того, заданы условные функции распределения $F_{ij}(x)$, $i, j \in E$, времени пребывания полумарковского процесса в состоянии $i$ при условии, что следующим состоянием будет $j$. Предполагается, что функции распределения $\{F_{ij}(x)\}$ нерешетчатые, причем существует такое $\delta > 0$, что $\sum\limits_{j \in E} F_{ij}(\delta) < 1$ для всех $i \in E$ (при этом условии невозможен случай, когда в конечном промежутке с положительной вероятностью происходит бесконечное число переходов полумарковского процесса). Предполагается также конечность средних времен $\{a_{ij}\}$ между соответствующими переходами. Введенные условия гарантируют эргодичность полумарковского процесса.

Для каждого $i \in E$ задана функция $\lambda_i(x),\ x \ge 0$, имеющая следующий смысл: если процесс $\eta(t)$ к моменту $t$ находится в состоянии $i$ в течение времени $x$, то вероятность поступления требования в интервале $(t, t+\Delta t)$ равна $\lambda_i(x)\Delta t + o(\Delta t)$. Вероятность поступления более одного требования в данном интервале равна $o(\Delta t)$.

Траектория полумарковского процесса — это двухмерная последовательность $\{(\nu_k, \tau_k),\ k \ge 0\}$, задаваемая моментами скачков $\{\tau_k\}$ и состояниями $\{\nu_k\}$ в эти моменты:

$$\tau_0 = 0, \mathbf{P}\{\nu_0 = i\} = p_i^{(0)},\ i \in E,\ \ \mathbf{P}\{\nu_{k+1} = j, \tau_{k+1} - \tau_k < x \mid \nu_k = i\} = p_{ij} F_{ij}(x), i, j \in E,\ x > 0.$$

Рассмотрим теперь стационарный момент $t^* = 0$. Приведенные условия гарантируют существование эргодического распределения полумарковского процесса. Если моделировать траекторию полумарковского процесса от момента $t^*$ «в прошлое» [30], то получим траекторию $S = \{(\nu_k, \tau_k),\ k \ge 0\}$, причем

$$\tau_0 = 0, \quad \mathbf{P}\{\nu_0 = i\} = \pi_i a_i / a,\ \ i \in E, \quad a = \sum_{j \in E} \pi_j a_j, \quad a_j = \sum_{l \in E} p_{jl} a_{jl}.$$

Обозначим $M$ множество всех таких траекторий, $\aleph$ — соответствующую $\sigma$-алгебру подмножеств $M$, $\mathbf{Q}\{\cdot\}$ — вероятностную меру на $\aleph$ (данная мера однозначно определяется переходными вероятностями $\{p_{ij}\}$ и распределениями $\{F_{ij}(x)\}$). При фиксированной последовательности $S$ интенсивность поступления требования в систему в момент $u$ определяется по формуле: если $\tau_{m-1} < u \le \tau_m$ $(m \ge 1)$, то

$$\mu(u; S) = \lambda_{\nu_m}(\tau_m - u). \tag{15}$$

Обозначим

$$\theta(S) = \int_0^\infty \mu(u; S)\,[1 - G(u)]\, du. \tag{16}$$

Пусть $\nu$ — число требований в системе обслуживания в момент $t^*$, $P_k = \mathbf{P}\{\nu = k\},\ k \ge 0$. Справедливо следующее утверждение.

**Теорема 5.** Предположим, что выполнены условия, гарантирующие существование приведенного интеграла. Тогда стационарные вероятности $\{P_k,\ k \ge 0\}$ состояний системы вычисляются по формуле

$$P_k = \int_M \frac{[\theta(S)]^k}{k!} e^{-\theta(S)}\, \mathbf{Q}\{dS\},\ \ k \ge 0. \tag{17}$$

Пусть $\varepsilon$ — некоторое малое число (граница отсечения), $W(\varepsilon) = \inf\{u : 1 - G(u) < \varepsilon\}$. Как и ранее, будем рассматривать лишь те члены последовательности $S = \{(\nu_k, \tau_k),\ k \ge 0\}$, для которых $\tau_k \le W(\varepsilon)$. Обозначим $S^{(n)} = \{(\nu_k, \tau_k),\ 0 \le k \le n\}$, где $n = \max\{k : \tau_k \le W(\varepsilon)\}$. Алгоритм построения асимптотически несмещенной (при $\varepsilon \to 0$) оценки $\hat{P}^{(1)}(A; \varepsilon)$ в одной реализации для $P(A) = \sum_{k \in A} P_k,\ A \subset \{0, 1, 2, \ldots\}$, основан на формулах (15)–(17) и формулируется следующим образом.

1. Моделируем последовательность $S^{(n)} = \{(\nu_k, \tau_k),\ 0 \le k \le n\}$ (начиная с момента $t^*$, моделируем моменты перехода и состояния полумарковского процесса «в прошлое»). Полагаем $\tau_0 = 0$ и моделируем состояние полумарковского процесса в момент $t^*$: $\nu_1 = j \in E$ с вероятностью $\pi_j a_j / a$. Пусть $\nu_1 = j$.

2. Моделируем время $\gamma_1$, прошедшее с момента предыдущего изменения состояния процесса $\mathbf{P}\{\gamma_1 < x\} = \dfrac{1}{a_j} \displaystyle\int_0^x [1 - F_j(v)]\, dv, x \ge 0$, где $F_j(v) = \sum_{l \in E} F_{jl}(v) p_{jl}$. Полагаем $\tau_1 = \tau_0 + \gamma_1$.

3. Предположим, что для некоторого $m \geq 0$ построены момент перехода $\tau_m$ и состояние $\nu_m$ процесса в момент $\tau_m - 0$. Состояние $\nu_{m+1}$ моделируем согласно распределению: $\nu_{m+1} = i \in E$ с вероятностью $p_{i\nu_m} \Big/ \sum_{k \in E} p_{k\nu_m}$.

4. Моделируем время $\gamma_{m+1}$ между переходами, $\mathbf{P}\{\gamma_{m+1} < x\} = F_{\nu_{m+1}\nu_m}(x)$, $x \geq 0$. Полагаем $\tau_{m+1} = \tau_m + \gamma_{m+1}$. Если $\tau_{m+1} > W(\varepsilon)$, то полагаем $n = m$, алгоритм завершен и траектория $S^{(n)}$ полумарковского процесса построена. В противном случае увеличиваем $m$ на единицу и переходим к п. 3 алгоритма.

5. Вычисляем (см. (15), (16))

$$\theta(S^{(n)}) = \sum_{m=1}^{n} \int_{\tau_{m-1}}^{\tau_m} \lambda_{\nu_m}(\tau_m - u)\,[1 - G(u)]\,du\,. \tag{18}$$

6. В качестве оценки $\hat{P}^{(1)}(A;\varepsilon)$ выбираем

$$\hat{P}^{(1)}(A;\varepsilon) = \sum_{k \in A} \frac{[\theta(S^{(n)})]^k}{k!}\, e^{-\theta(S^{(n)})}. \tag{19}$$

Как и для всех рассмотренных моделей входящих потоков требований, аналитическая часть метода, основанная на формулах (18), (19), позволяет существенно сократить дисперсию оценки, что демонстрирует приведенный далее пример.

**Пример 2.** В качестве параметров модели выберем $E = \{1, 2, 3\}$; $F_{ij}(x) = 1 - \exp\{-(\alpha_{ij} x)^2\}$, $x > 0$, $a_{ij} = \dfrac{\sqrt{\pi}}{2\alpha_{ij}}$; матрицу $P = (p_{ij})$ переходных вероятностей и матрицу $B = (a_{ij})$ средних времен пребывания в состояниях определим так:

$$P = \begin{pmatrix} 0 & 0.7 & 0.3 \\ 0.8 & 0 & 0.2 \\ 0.9 & 0.1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 & 4 \\ 3 & 0 & 6 \\ 2 & 4 & 0 \end{pmatrix};$$

интенсивности поступления требований зададим формулой $\lambda_i(x) = 50\, i\, x$, $i \in E$; время обслуживания требований имеет распределение Вейбулла $G(x) = 1 - e^{-x^2}$, $x \geq 0$, $\tau = \sqrt{\pi}/2$.

Исследуются среднее количество $s$ требований в системе в стационарном режиме и вероятность нахождения в системе не менее $r$ требований, т.е. $A = \{r, r+1, \ldots\}$.

Решив систему линейных уравнений, получим стационарные вероятности состояний вложенной цепи Маркова: $\pi_1 = 0.45581$, $\pi_2 \approx 0.33954$, $\pi_3 \approx 0.20465$. Выберем $\varepsilon = 10^{-7}$. В табл. 2 использованы те же обозначения, что и в табл. 1. Все приведенные оценки построены по 100 000 реализациям алгоритма. При этом $\hat{S}_{\text{fast}} = 74.43$, $\hat{R}(\hat{S}_{\text{fast}}) = 0.80\%$.

Как и следовало ожидать, для оценки малых вероятностей наиболее эффективно ускоренное моделирование. Так, при переходе от $r = 300$ к $r = 350$ вероятность уменьшилась на пять порядков, в то время как относительная погрешность возросла всего в полтора раза.

**Таблица 2**

| Требования, $r$ | Оценки стационарной вероятности и относительной погрешности | |
|---|---|---|
| | $\hat{P}_{\text{fast}}(r;\varepsilon)$ | $\hat{R}(\hat{P}_{\text{fast}}(r;\varepsilon))$, % |
| 150 | 0.19 | 1.59 |
| 200 | 0.10 | 2.21 |
| 250 | $2.69 \cdot 10^{-2}$ | 3.76 |
| 300 | $2.83 \cdot 10^{-4}$ | 6.42 |
| 350 | $5.39 \cdot 10^{-9}$ | 9.10 |

## ЗАКЛЮЧЕНИЕ

В работе предложен аналитико-статистический метод оценки стационарных вероятностей состояний немарковских систем массового обслуживания при различных моделях входящих потоков требований. Вначале методом Монте-Карло моделируется некоторый параметр системы (статистическая часть), а затем аналитически с помощью распределения Пуассона вычисляется стационарная вероятность нахождения системы в определенном множестве состояний. Данный метод наиболее эффективен при оценке малых вероятностей, связанных с наступлением редких событий. Предложенный поход можно использовать и для ряда других моделей входящих потоков, не рассмотренных в статье.

СПИСОК ЛИТЕРАТУРЫ


1. Климов Г.П. Стохастические системы обслуживания. Москва: Наука, 1966. 244 с.
2. Takács L. On Erlang's formula. *Ann. Math. Stat.* 1969. Vol. 40, N 1. P. 71–78.
3. König D., Stoyan D. Methoden der Bedienungstheorie. Berlin.: Akademie-Verlag, 1976. 238 S.
4. Саати Т. Элементы теории массового обслуживания и ее приложения. Москва: Либроком, 2010. 520 с.
5. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. Москва: ЛКИ, 2012. 400 с.
6. Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности. Москва: Либроком, 2013. 584 с.
7. Севастьянов Б.А. Предельная теорема для марковских процессов и ее приложение к телефонным системам с отказами. *Теория вероятностей и ее применения.* 1957. Т. 2, вып. 1. С. 106–116.
8. König D., Mattes K., Nawrotzki K. Verallgemeinerungen der Erlangschen und Engsetschen Formeln (eine Methode in der Bedienungstheorie). Berlin: Akademie-Verlag, 1967.
9. Франкен П., Кениг Д., Арндт У., Шмидт Ф. Очереди и точечные процессы. Киев: Наук. думка, 1984. 283 с.
10. Kovalenko I.N. Rare events in queueing systems — a survey. *Queueing Systems*. 1994. Vol. 16, N 1. P. 1–49.
11. Blaszczyszyn B., Rolski T., Schmidt V. Light-traffic approximations in queues and related stochastic models. Advances in Queueing. Boca Raton: CRC Press, 1995. P. 379–406.
12. Kovalenko I.N. Approximation of queues via small-parameter method. Advances in Queueing. Boca Raton: CRC Press, 1995. P. 481–506.
13. 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.
14. Atkinson J.B., Kovalenko I.N. On the practical insensitivity of the availability of some redundant repairable systems to the lifetime distribution in light traffic. *Proc. Int. Conf. on Probability Analysis of Rare Events.* Riga: Aviation University, 1999. P. 83–91.
15. 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.
16. Королюк В.С., Турбин А.Ф. Математические основы фазового укрупнения сложных систем. Киев: Наук. думка, 1978. 217 с.
17. Korolyuk V.S., Korolyuk V.V. Stochastic models of systems. Boston: Kluwer, 1999. 185 p.
18. Меликов А.З., Пономаренко Л.А., Паладюк В.В. Телетрафик: модели, методы, оптимизация. Киев: Политехника, 2007. 256 с.
19. Меликов А.З., Пономаренко Л.А., Фаттахова М.И. Управление мультисервисными сетями связи с буферными накопителями. Киев: НАУ-друк, 2008. 156 с.
20. Ponomarenko L., Kim C.S., Melikov A. Performance analysis and optimization of multi-traffic on communication networks. New York: Springer, 2010. 208 p.
21. Ливинская А.В., Лебедев Е.А. Предельная теорема для перегруженных многоканальных сетей. *Кибернетика и системный анализ.* 2012. № 6. С. 106–113.
22. 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.

23. Kovalenko I.N., Kuznetsov N.Yu., Pegg Ph. Mathematical theory of reliability of time dependent systems with practical applications. Chichester: Wiley, 1997. 303 p.

24. Ermakov S.M., Melas V.B. Design and analysis of simulation experiments. Dordrecht: Kluwer Academic Publ., 1995. 197 p.

25. Lassila P.E., Virtamo J.T. Efficient importance sampling for Monte Carlo simulation of loss systems. Proc. of the ITC-16, Teletraffic Engineering in a Competitive World. Edinburgh: Elsevier, 1999. P. 787–796.

26. Glasserman P., Heidelberger Ph., Shahabuddin P., Zajic T. Multilevel splitting for estimating rare event probabilities. *Operat. Res.* 1999. Vol. 47, N 4. P. 585–600.

27. Шумская А.А. Оценка стационарной вероятности потери в системе массового обслуживания с рекуррентными потоками требований. *Кибернетика и системный анализ*. 2004. № 2. С. 133–145.

28. Ермаков С. М. Метод Монте-Карло и смежные вопросы. Москва: Наука, 1975. 472 с.

29. Коваленко И.Н., Кузнецов Н.Ю., Шуренков В.М. Случайные процессы: справочник. Киев: Наук. думка, 1983. 366 с.

30. Шумська А.А. Прискорене моделювання стаціонарного розподілу кількості вимог в системі. *SMBAP*$|G|\infty$. *Системні дослідження та інформаційні технології.* 2004. № 3. С. 121–129.

*Надійшла до редакції 29.06.2016*

**І.М. Кузнєцов, А.А. Шумська**

**ОЦІНКА СТАЦІОНАРНИХ ЙМОВІРНОСТЕЙ СТАНІВ СИСТЕМИ ОБСЛУГОВУВАННЯ $\bullet/G/\infty$ ПРИ РІЗНИХ ВИДАХ ВХІДНОГО ПОТОКУ ВИМОГ**

**Анотація.** Розглянуто п'ять моделей вхідного потоку істотно більш складної структури, ніж пуасонівський, коли стаціонарні ймовірності станів системи $\bullet/G/\infty$ знаходяться у явному вигляді (розподіл Пуасона). Для наведених моделей поєднання розподілу Пуасона (аналітична частина) із статистичним моделюванням (статистична чистина) дозволяє знаходити стаціонарні ймовірності станів прискореним моделюванням. Точність оцінок проілюстровано числовими прикладами.

**Ключові слова:** система обслуговування, стаціонарні ймовірності станів, нестаціонарний пуасонівський процес, регенерувальний процес, напівмарковський процес, процес скупчень, незміщена оцінка, відносна похибка.

**I.N. Kuznetsov, A.A. Shumskaya**

**THE EVALUATION OF STEADY-STATE PROBABILITIES OF QUEUEING SYSTEM $\bullet/G/\infty$ FOR DIFFERENT INPUT FLOW MODELS**

**Abstract.** We consider five input flow models of more complicated structure than the Poisson one, where steady-state probabilities of the queueing system $\bullet/G/\infty$ can be found explicitly (the Poisson distribution). For these models, the combination of Poisson distribution (analytical part) with statistical simulation (statistical part) allows us to evaluate the steady-state probabilities with the fast simulation method. The accuracy of the estimates is illustrated by numerical examples.

**Keywords:** queueing system, steady-state probabilities, nonstationary Poisson flow, regenerative process, semi-Markov process, cluster process, unbiased estimate, relative error.

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

**Шумская Алла Антоновна**,
кандидат физ.-мат. наук, доцент Физико-технического института Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского»,
e-mail: shumska-aa@ukr.net.