III ЛИТОВСКИЙ МАТЕМАТИЧЕСКИЙ СБОРНИК 1
LIETUVOS MATEMATIKOS RINKINYS

1963

ПЕРЕХОДНЫЕ ЯВЛЕНИЯ В ПРОЦЕССАХ МАССОВОГО ОБСЛУЖИВАНИЯ I

Ю. В. ПРОХОРОВ

Рассматриваются системы массового обслуживания при условии, что нагрузка близка к критическому значению (т. е. к значению, отделяющему случай стационарной очереди от случая неограниченно растущей очереди). По-видимому, ряд утверждений, полученных автором в специальном случае (например, оценка времени вхождения системы в стационарный режим, показательность предельного распределения) имеет место в значительно более общей обстановке.

Характер обсуждаемых вопросов проще всего объяснить на следующем примере. В моменты $t_1=\xi_1$, $t_2=\xi_1+\xi_2$, ..., где ξ_1, ξ_2, ... – независимые неотрицательные одинаково распределенные величины, на обслуживающий прибор поступают вызовы. В случае, когда прибор занят, вызовы становятся в очередь и обслуживаются в порядке очереди. Время обслуживания j-го вызова равно η_j. Последовательность η_1, η_2, ... состоит из независимых одинаково распределенных случайных величин и как целое не зависит от $\{\xi_j\}$.

Пусть $\mathbf{M}\xi_j=\alpha$, $\mathbf{M}\eta_j=\beta$. Не ограничивая общности, примем $\beta=1$. Известно, (см., напр. [1]), что при $\alpha>1$ и $n\to\infty$ распределение времени w_n ожидания n-ым вызовом начала обслуживания стремится к стационарному. При α, приближающемся к единице, это стремление становится все более медленным. Цель статьи состоит в изучении явлений, возникающих при двойном предельном переходе $\alpha\downarrow 1$ (или $\delta=1-\alpha\downarrow 0$) и $n\to\infty$. Например, один из результатов $\big($формула (8)$\big)$ в применении к случаю, когда ξ_j и η_j распределены показательно со средними $\alpha=1/\lambda=1/(1-\delta)$ и $\beta=1/\mu=1$, принимает форму:

при $n\delta^2\to\tau>0$

$P\left\{w_n\leqslant\frac{v}{\delta}\right\}\sim P\left\{\xi(t)\leqslant\frac{x}{\sqrt{2}}+\frac{t}{\sqrt{2}}\text{ при всех }0\leqslant t\leqslant\tau\right\}$, где $\xi(t)$ – стандартный винеровский процесс (с $\mathbf{M}\Delta\xi=0$ и $\mathbf{D}\Delta\xi=\Delta t$). В частности, при $n\delta^2\to\infty$ имеем

$$P\left\{w_n\leqslant\frac{v}{\delta}\right\}\sim 1-e^{-v}.$$

Аналитический метод для случая $n\delta^2\to\infty$ разрабатывался Э. Г. Самандаровым [2]. Поведение стационарного распределения $F(x)=\lim_{n\to\infty}P\{w_n\leqslant x\}$ при $\delta\to 0$ исследовалось Кингманом [3] и И. Коваленко. В частности, этими авторами были даны достаточные условия того, чтобы $F(x)$ было асимптотически показательно: при $\delta\to 0$

$$F\left(\frac{v}{\delta}\right)\to 1-e^{-v}.$$

Во второй половине статьи изучается поведение при $\delta\to 0$ случайного процесса $\nu_\delta(t)$, где $\nu_\delta(t)$ — число абонентов, занятых в системе в момент t. Грубо говоря, результат исследования таков. При малом δ и $t\delta^2$ значительно меньшем 1, очередь ведет себя так же, как в случае $\delta=0$. При $t\delta^2$ значительно большем 1 очередь входит в стационарный режим. При $t\delta^2$ по порядку сравнимым с 1 наблюдается некоторый промежуточный режим (наглядные иллюстрации поведения $\mathbf{M}\nu_0(t)$ приведены в работе [4]).

§ 1. Легко видеть, что величины w_n связаны следующими соотношениями:

$$w_1=0, \quad w_{n+1}=(w_n+u_n)^+, \quad n\geqslant 1, \tag{1}$$

где $u_n=\eta_n-\xi_{n+1}$ и $a^+=a$ при $a\geqslant 0$ и $=0$ при $a<0$. Как показано в работе [1] из (1) вытекает равенство: при $x>0$

$$F_{n+1}(x)=P\{w_{n+1}\leqslant x\}=P\left\{\max_{1\leqslant k\leqslant n}\sum_{j=1}^{k}u_j\leqslant x\right\}. \tag{2}$$

Мы произведем в последней формуле предельный переход от распределения максимума сумм независимых случайных величин к распределению соответствующего функционала от винеровского процесса. Этим путем мы и получим желательную асимптотическую формулу для F_n. Указанный предельный переход связан с „принципом инвариантности“ Эрдеша — Каца [5] и Донскера [6] и может быть обоснован ссылкой на результаты работы [7] (глава 3). Оценку остаточного члена в приводимых формулах можно получить, например, или из [7] (глава 4) или из [8].

§ 2. Мы приступим теперь к доказательству основных формул. Допустим, что каждому δ, $0<\delta<1$ соответствует своя схема с $\xi_j=\xi_j^{(\delta)}$, $\eta_j=\eta_j^{(\delta)}$, $u_j=u_j^{(\delta)}=\xi_j^{(\delta)}-\eta_{j+1}^{(\delta)}$ *), причем при $\delta\to 0$

$$\mathbf{M}u_j\sim-\delta, \tag{3}$$

$$\sigma_{u_j}^2\sim\sigma^2, \quad 0<\sigma^2<\infty, \tag{4}$$

и при $A\to\infty$ равномерно относительно δ

$$\int_{|y|\geqslant A}y^2\,dF_{uj}(y)\to 0. \tag{5}$$

Как легко показать, для (5) достаточно, чтобы при $A\to\infty$ равномерно относительно δ

$$\int_{|x|>A}x^2\,dF_\xi\to 0 \quad \text{и} \quad \int_{|y|>A}y^2\,dF_\eta\to 0. \tag{6}$$

А. Пусть при $\delta\to 0$

$$n_\delta\cdot\delta^2\to\tau, \quad 0<\tau<\infty. \tag{7}$$

Из (2) при фиксированном v, $0<v<\infty$ получаем

$$F_{n+1}(v/\delta)=P\left\{\frac{\sum_{j=1}^{k}(u_j-\mathbf{M}u)}{\sqrt{n}\,\sigma_u}\leqslant\frac{v-\frac{k}{n_\delta}\cdot n_\delta\,\mathbf{M}u\cdot\delta}{\sqrt{n_\delta}\cdot\delta\cdot\sigma_u}\quad\text{при всех}\quad 0\leqslant\frac{k}{n_\delta}\leqslant 1\right\}.$$

*) Индекс δ наверху мы будем иногда сохранять, а иногда опускать, в зависимости от удобства записи.

В силу (3) и (4) при $\delta\to 0$

$$-n_\delta \mathbf{M}u\cdot\delta\to\tau,\quad \sqrt{n_\delta}\cdot\delta\cdot\sigma_u\to\sigma\sqrt{\tau}.$$

По теореме 3.1 работы [7] мы получаем

$$F_{n_\delta}\left(\frac{v}{\delta}\right)\to P\left\{\xi(t)\leqslant\frac{v+t\tau}{\sigma\sqrt{\tau}}\quad\text{при всех}\quad 0\leqslant t\leqslant 1\right\}=$$

$$=P\left\{\xi(t)\leqslant\frac{v+t}{\sigma}\quad\text{при всех}\quad 0\leqslant t\leqslant\tau\right\}=p(v,\ \sigma,\ \tau). \tag{8}$$

Б. Пусть при $\delta\to 0$ $n_\delta\cdot\delta^2\to\infty$. Фиксируем $\varepsilon>0$ и выберем τ_ε так, чтобы

$$p(v,\ \sigma,\ \tau_\varepsilon)\leqslant p(v,\ \sigma,\ \infty)+\frac{\varepsilon}{2}. \tag{9}$$

Возьмем, далее, δ_0 таким, что при $\delta\leqslant\delta_0 n_\delta\geqslant n'_\delta=[\tau_\varepsilon/\delta^2]$ и δ_1 таким, чтобы при $\delta\leqslant\delta_1 F_{n'_\delta}(v\delta^{-1})$ отличалось от левой части (9) не более, чем на $\varepsilon/2$. Но, как видно из (2)

$$F_{n_\delta}(v\delta^{-1})\leqslant F_{n'_\delta}(v\delta^{-1}). \tag{10}$$

Из (8), (9) и (10) заключаем, что при $0<\delta<\min(\delta_0,\ \delta_1)$ $F_{n_\delta}(v\delta^{-1})\leqslant p(v,\ \sigma,\ \infty)+\varepsilon$, т. е.

$$\varlimsup_{\delta\to 0} F_{n_\delta}(v\delta^{-})\leqslant p(v,\ \sigma,\ \infty). \tag{11}$$

Докажем теперь, что для $\varliminf\limits_{\delta\to 0}$ имеет место обратное неравенство. Для этого нам потребуется.

Лемма. При достаточно малых δ

$$P\left\{\sup_{k\geqslant k_0}\sum_{j=1}^{k}u_j>0\right\}\leqslant\frac{32\,\sigma_u^2}{k_0\cdot\delta^2}. \tag{12}$$

Действительно, событие, вероятность которого оценивается в лемме, может быть записано в форме $\{s_k>-k\mathbf{M}u$ хотя бы при одном $k\geqslant k_0\}$, где $s_k=\sum\limits_{j=1}^{k}(u_j-\mathbf{M}u_j)$. При $\delta\leqslant\delta_2$ имеем $-\mathbf{M}u\geqslant\frac{\delta}{2}$ и потому вероятность предыдущего события не превосходит вероятности события $\left\{s_k>k\frac{\delta}{2}\right.$ хотя бы при одном $\left.k\geqslant 2^l\right\}$, где $2^l\leqslant k_0<2^{l+1}$. Применяя известное неравенство Колмогорова, легко устанавливаем, что последняя вероятность не превосходит

$$\sum_{j=l}^{\infty}P\left(\max_{2^j\leqslant k<2^{j+1}}s_k>\frac{\delta}{2}\cdot 2^j\right)\leqslant\sum_{j=l}^{\infty}\frac{2^{j+1}\sigma_u^2}{\delta^2\cdot 2^{2j}}\leqslant\frac{32\,\sigma_u^2}{k_0\cdot\delta^2}$$

ч. тр. д.

Фиксируем теперь какое-либо число v_1, $0<v_1<\infty$. Тогда при любом n и при $v_1\leqslant k_0\delta^2<v_1+\delta^2$, $\delta\leqslant\delta_3$ имеем

$$F_n(v\delta^{-1})\geqslant F(v\delta^{-1})=P\left\{\sup_{1\leqslant k<\infty}\sum_{j=1}^{k}u_j\leqslant v\delta^{-1}\right\}\geqslant$$

$$\geqslant P\left\{\sup_{1\leqslant k\leqslant k_0-1}\sum_{j=1}^{k}u_j\leqslant v\delta^{-1}\right\}-32\,\sigma_n^2 v_1^{-1}=F_{k_0}(v\delta^{-1})-32\,\sigma_u^2 v_1^{-1}\geqslant$$

$$\geqslant p(v,\ \sigma,\ v_1)-\varepsilon-64\,\sigma^2 v_1^{-1}.$$

Из выписанных неравенств следует, что

$$\varliminf_{\delta\to 0} F_{n_\delta}(v\delta^{-1}) \geqslant p(v, \sigma, \infty). \tag{13}$$

Из (11) и (13) вытекает

$$\lim_{\delta\to\infty} F_{n_\delta}(v\delta^{-1}) = p(v, \sigma, \infty) = 1 - e^{-\frac{2v}{\sigma^2}} \tag{14}$$

(по поводу вычисления значения $p(v, \sigma, \infty)$ см., напр. [9]).

Замечание. Из предыдущих рассуждений видно также, что предельное распределение $F(x) = \lim\limits_{n\to\infty} F_n(x)$ при $\delta\to 0$ асимптотически показательно.

В. Пусть теперь при $\delta\to 0$ $n_\delta\to\infty$, $n_\delta\cdot\delta^2\to 0$. Рассуждениями, сходными с предшествующими, нетрудно получить формулу: при $v>0$

$$F_{n_\delta}(v\sigma\sqrt{n_\delta}) \to P\{\max_{0\leqslant t\leqslant 1} \xi(t) \leqslant v\} = 2\Phi(v) - 1.$$

где $\Phi(v)$ — см. нормальная функция распределения.

§ 3. Заключительные замечания. А. Как можно подсчитать переходом к преобразованиям Лапласа, плотность распределения $p(v, \sigma, \tau)$, т. е., $\frac{\partial p}{\partial v}$ определяется выражением

$$\frac{1}{\delta}\sqrt{\frac{2}{\pi\tau}}\exp\left\{-\frac{(v+\tau)^2}{\sigma^2\tau}\right\} + \frac{2}{\sigma^2\sqrt{\pi}}e^{-\frac{2v}{\sigma^2}}\int\limits_{\frac{v-\tau}{\sigma\sqrt{2\tau}}}^{\infty}\exp(-x^2)\,dx.$$

Последнее выражение совпадает с мгновенной плотностью $w(v, \tau, 0)$ в момент τ для процесса $\eta(t)$ — гауссовского процесса с независимыми приращениями с $\mathbf{M}\Delta\eta = -\Delta t$, $\mathbf{D}\Delta\eta = \sigma^2\Delta t$, начинающегося из точки $+0$ при наличии в нуле отражающего экрана (см. [11], гл. III, § 5). К этой формуле легко прийти из эвристических соображений, если заметить, что процесс $\eta_\delta(t)$, определяемый равенством

$$\eta_\delta(n\delta^2) = \delta w_n$$

„очень похож" на процесс $\eta(t)$. Однако формальное обоснование проще всего достигается через формулу (2). Заметим, что из приведенной формулы можно оценить, с какой скоростью распределение $\eta(t)$ стремится при $t\to\infty$ к предельному показательному распределению. Б. Усиливая требование (5) (соотв. (6)), мы можем получить теоремы о сходимости моментов. В качестве иллюстрации можно рассмотреть поведение математических ожиданий. Допустим, что (5) заменено требованием: существует такая константа K, что при всех достаточно малых δ

$$\rho_{3,u} = \sigma_u^{-3}\mathbf{M}|u_j - \mathbf{M}u_j|^3 \leqslant K \tag{5'}$$

(для (5′) достаточно, например, чтобы при тех же δ и некотором K_1

$$\mathbf{M}|\xi_j|^3 + \mathbf{M}|\eta_j|^3 \leqslant K_1).$$

Учитывая (5′), мы усилим оценку (12). Для этого мы используем принадлежащее Б. А. Рогозину [10] утверждение: пусть $\zeta_1, \zeta_2, \ldots, \zeta_n$ — независимые одинаково распределенные случайные величины с

$$\mathbf{M}\zeta_k = 0, \quad \mathbf{D}\zeta_k = \sigma_\zeta^2 \quad \text{и} \quad \rho_3 = \sigma_\zeta^{-3}\mathbf{M}|\zeta_k|^3 < \infty,$$

тогда

$$\left|P\{\zeta_1+\zeta_2+\ldots+\zeta_n>x\sigma_\zeta\sqrt{n}\}-\left(1-\Phi(x)\right)\right|\leqslant\frac{C\rho_3}{\sqrt{n}\,(1+x^3)}$$

(буква C здесь и в дальнейшем обозначает абсолютные постоянные). Из последнего неравенства при $x\geqslant 1$ получаем

$$P\{\zeta_1+\zeta_2+\ldots+\zeta_n>x\sigma_\zeta\sqrt{n}\}\leqslant\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}+\frac{C\rho_3}{\sqrt{n}\,(1+x^3)}\,.$$

Применим это соотношение к суммам s_n (см. лемму выше). При этом вместо упоминавшегося там неравенства Колмогорова мы используем другое, а именно

$$P\{\max_{1\leqslant k\leqslant nk}\zeta_k\geqslant y\}\leqslant 2P\{\zeta_n\geqslant y-\sqrt{2\mathbf{D}\zeta_n}\}\,.$$

Тогда мы получим

$$P\left\{\max_{2^j\leqslant k\leqslant 2^{j+1}}s_k>\frac{\delta}{2}\cdot 2^j\right\}\leqslant 2P\left\{s_{2^{j+1}}>\frac{\delta}{2}\cdot 2^j-\sqrt{2\cdot 2^{j+1}\cdot\sigma_u}\right\}. \qquad (15)$$

В дальнейшем мы рассмотрим только те значения δ, для которых выполняется (5′) и для которых $\sigma_u^2\leqslant 2\sigma^2$. Допустим, что $k_0\delta^2\geqslant 2^8\cdot\sigma^2$. Тогда правая часть неравенства (15) не превосходит

$$2P\left\{s_{2^{j+1}}>\frac{\delta}{8}\cdot 2^{j+1}\right\}=2P\left\{s_{2^{j+1}}>\frac{\delta\sqrt{2^{j+1}}}{8\sigma_n}\cdot\sigma_n\sqrt{2^{j+1}}\right\}\leqslant$$

$$\leqslant\frac{2}{\sqrt{2\pi}}\exp\left\{-\frac{\delta^2\cdot 2^{j+1}}{2\cdot 64\cdot\sigma_u^2}\right\}+2C\frac{K\cdot 64\cdot\sigma_u^2}{\sqrt{2^{j+1}}\cdot\delta^2\cdot 2^{j+1}}\,.$$

Отсюда, так же, как в доказательстве леммы, получаем

$$\left(\Theta=2^{\frac{3}{2}},\quad x_j=\frac{\delta^2\cdot 2^j}{2\cdot 64\cdot\sigma_u^2}\right)$$

$$P\left\{\sup_{k\geqslant k_0}\sum_{j=1}^{k}u_j>0\right\}\leqslant\sum_{j=l}^{\infty}P\left\{\max_{2^j\leqslant k\leqslant 2^{j+1}}s_k>\frac{\delta}{2}\cdot 2^j\right\}\leqslant$$

$$\leqslant\frac{2}{\sqrt{2\pi}}\cdot 2\sum_{j\geqslant l}\exp\{-x_{j+1}\}(x_{j+1}-x_j)+\frac{256\cdot C\cdot K\cdot\sigma^2}{\delta}\sum_{j\geqslant l}\frac{1}{\Theta^{j+1}}\leqslant$$

$$\leqslant\frac{4}{\sqrt{2\pi}}\exp\left\{-\frac{k_0\delta^2}{2^9\cdot\sigma^2}\right\}+K_1\cdot\frac{1}{\delta^2}\cdot\frac{1}{k_0^{\frac{3}{2}}}\,.$$

Следовательно, при $v>0$

$$1-F^{(\delta)}(v\delta^{-1})=P\left\{\text{хотя при одном } k\ s_k>\frac{v}{\delta}-\mathbf{M}u\right\}\leqslant$$

$$\leqslant P\left\{\max_{1\leqslant k\leqslant k_0}s_k>\frac{v}{\delta}\right\}+P\left\{\sup_{k\geqslant k_0}\sum_{j=1}^{k}u_j>0\right\}\leqslant$$

$$\leqslant\frac{k_0\cdot\sigma_u^2\cdot\delta^2}{v^2}+\frac{4}{\sqrt{2\pi}}\exp\left\{-\frac{k_0\delta^2}{2^9\cdot\sigma^2}\right\}+K_1\cdot\frac{\delta}{k_0\delta^2\sqrt{k_0\cdot\delta^2}}\,.$$

Допустим теперь, что $v\geqslant 1$ и что δ достаточно мало. Выберем $k_0\geqslant 2$ так, что $(k_0-1)\delta^2<v^{\frac{4}{5}}\leqslant k_0\delta^2$. Тогда автоматически $k_0\delta^2\leqslant 2v^{\frac{4}{5}}$ и

$$1-F^{(\delta)}(v\delta^{-1})\leqslant 4\sigma^2v^{-\frac{6}{5}}+\frac{4}{\sqrt{2\pi}}\exp\left\{-2^{-9}\sigma^{-2}v^{\frac{4}{5}}\right\}+K_1\cdot v^{-\frac{6}{5}}\leqslant K_2\cdot v^{-\frac{6}{5}}.$$

Отсюда выводим, что при всех $v \geqslant 0$

$$1-F^{(\delta)}(v\delta^{-1}) \leqslant K_3(1+v)^{-\frac{6}{5}}.$$

Так как при любых v, δ и n

$$1-F_n^{(\delta)}(v\delta^{-1}) \leqslant 1-F^{(\delta)}(v\delta^{-1}) \leqslant K_3(1+v)^{-\frac{6}{5}},$$

то из формулы

$$\mathbf{M}(\delta w_n) = \int\limits_0^\infty \left(1-F_n^{(\delta)}(v\delta^{-1})\right) dv$$

и теоремы о мажорированной сходимости получаем: при $n_\delta \cdot \delta^2 \to \tau$, $0<\tau\leqslant\infty$

$$\mathbf{M}(\delta w_{n_\delta}) \to \mathbf{M}\eta(\tau).$$

Равенство

$$\mathbf{M}\eta(\tau) = \frac{\sigma^2}{2}\left[1+\frac{\sqrt{\Theta}}{\sqrt{\pi}}e^{-\frac{\Theta}{4}} - (2+\Theta)(1-\Phi)\left(\sqrt{\frac{\Theta}{2}}\right)\right] =$$

$$= \frac{\sigma^2}{2}\left[1-\frac{8}{\Theta\sqrt{\Theta}\sqrt{\pi}}e^{-\frac{\Theta}{4}}+\dots\right],$$

где $\Theta=\frac{2\tau}{\sigma^2}$, дает представление о том, каково время вхождения системы в стационарный режим (если его оценивать по поведению среднего времени ожидания). При $\Theta=2$ (относительное) отклонение $\mathbf{M}\eta(\tau)$ от стационарного значения $\frac{\sigma^2}{2}$ составляет около 15%, при $\Theta=8$ — около 3%. Очевидно, что при достаточно малых δ аналогичное положение имеет место и для $\mathbf{M}(\delta w_{n_\delta})$. В. Путем некоторых вычислений из предыдущего можно вывести, что при

$$\delta \to 0, \quad n_\delta \to \infty, \quad n_\delta \cdot \delta^2 \to \tau, \quad 0 \leqslant \tau \leqslant \infty,$$

случайная величина $w_{n_\delta}/\mathbf{M}w_{n_\delta}$ имеет невырожденное предельное распределение. В случае $\tau=0$ оно равно $2\Phi\left(\sqrt{\frac{2}{\pi}}\,x\right)-1$, а в случае $\tau=\infty$ — равно $1-e^{-x}$.

Математический институт им. В. А. Стеклова АН СССР

Поступила в редакцию 28. VI. 1962

ЛИТЕРАТУРА

1. D. V. Lindley, Proc. Cambr. Phil. Soc. 48: 1 (1952), 277—289.
2. Э. Г. Самандаров, 2-ой Коллоквиум по предельным теоремам, Фергана, 1962, 15—16.
3. J. F. C. Kingman, Proc. Cambr. Phil. Soc. 57: 4 (1961), 902—904.
4. N. T. J. Bayley, J. Roy. Stat. Soc. 19: 2, Ser. B (1957), 326—333.
5. P. Erdös, M. Kac, Bull. A. M. S. 53 (1947), 1011.
6. M. D. Donsker, Mem. A. M. S., 6 (1951), 1—12.
7. Ю. В. Прохоров, Теория вероят. и ее примен. 1:2 (1956), 177—238.
8. А. В. Скороход, ДАН СССР, 133:1 (1960), 34—35.
9. D. A. Darling, A. J. F. Siegert, Ann. Math. Stat. 24: 2 (1953), 624—639.
10. Б. А. Рогозин, Диссертация, Москва, МГУ, 1962.
11. Чандрасекар, Стохастические проблемы в физике и астрономии.

PEREINAMIEJI REIŠKINIAI MASINIO APTARNAVIMO PROCESUOSE I

J. V. PROCHOROVAS

(Reziumė)

Straipsnyje nagrinėjamos masinio aptarnavimo sistemos, esant sąlygai, kad apkrovimas yra artimas kritiškai reikšmei, t. y. reikšmei skiriančiai, stacionarinės eilės atvejį nuo neaprėžtai didėjančios eilės.

THE TRANSITIONAL PHENOMENA IN THE QUEUING PROCESSES I

YU. V. PROHOROV

(Summary)

The systems of queues under condition that the traffic is close to critical value, i.e., to value which separates the case of stationary queue from the case of unfinitely increasing queue, are considered.