ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
VIII (1965)

J. ŁUKASZEWICZ (Wrocław)

TEORIA KOLEJEK CZYLI OBSŁUGI MASOWEJ

1. Czym jest teoria kolejek? Teoria kolejek jest niezwykle szybko rozwijającą się(1) gałęzią badań operacyjnych. Z drugiej strony, ze względu na stosowany w tej teorii aparat matematyczny, teoria kolejek może być uważana za rozdział teorii procesów stochastycznych.

Początków teorii kolejek należy szukać w pracach duńskiego teletechnika A. K. Erlanga ([7]), który zajmował się zagadnieniami stochastycznymi występującymi w pracy central telefonicznych. Wkrótce jednak nowa teoria znalazła wiele innych zastosowań (w dziedzinie komunikacji i transportu, przemysłu, handlu, urządzeń socjalnych itp.) i wzbudziła żywe zainteresowanie wielu matematyków, statystyków, techników i ekonomistów.

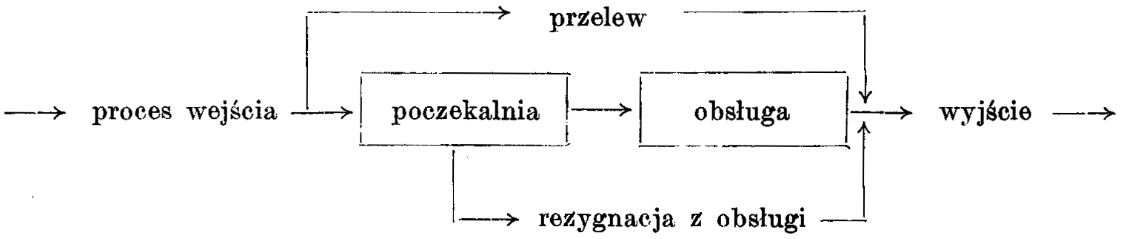
W Polsce teoria kolejek jest stosunkowo mało znana. Autor wykładał ją w roku 1959 na Uniwersytecie Wrocławskim i w roku 1963 na XVII Kursie Zastosowań Matematyki, organizowanym przez Instytut Matematyczny PAN w Warszawie. W roku 1963 wygłosił również cykl wykładów z teorii obsługi masowej B. W. Gniedenko, profesor Uniwersytetu Moskiewskiego, zaproszony do Polski przez Instytut Matematyczny PAN. Od roku 1963 na Uniwersytecie Wrocławskim prowadzone jest seminarium z teorii kolejek i zagadnień pokrewnych. Prace I. Kopocińskiej ([16]) i J. Rzytki ([21]) zawierają wyniki osiągnięte na tym seminarium.

W dostępnej w języku polskim literaturze naukowej omówienie pewnych zagadnień z teorii kolejek można znaleźć w książkach W. Fellera ([8]) i W. Sadowskiego ([23]). Zastosowania Matematyki drukowały już w tomie 7 dwie prace z tej dziedziny ([15] i [17]). Ogłaszając w niniejszym zeszycie kilka dalszych prac z teorii kolejek pragniemy je poprzedzić przeglądową pracą omawiającą podstawowe pojęcia teorii i stosowane w niej metody matematyczne. Praca ta powinna ułatwić polskiemu czytelnikowi czytanie dalszych prac i przyczynić się do spo-

(1) Bibliografia teorii kolejek, opublikowana przez pannę A. Doig ([6]) w roku 1957, wymienia 638 publikacji z tej dziedziny. Obecnie z tej dziedziny publikowanych jest ponad sto prac rocznie.

pularyzowania teorii, która może być z pożytkiem stosowana w wielu sytuacjach praktycznych. Czytelnikom, którzy chcieliby głębiej poznać tę teorię, polecamy obcojęzyczne opracowania tematu w książkach i pracach przeglądowych wyliczonych na końcu niniejszej pracy.

Większość urządzeń obsługi masowej, do których stosuje się teorię kolejek, ma następujący schemat



Rys. 1

Kolejne jednostki przybywają do urządzenia obsługującego w pewnych momentach czasu. Momenty te są zazwyczaj przypadkowo rozmieszczone w czasie i dokładna specyfikacja modelu wymaga probabilistycznego opisu procesu stochastycznego zgłoszeń, czyli tak zwanego *procesu wejścia.*

Gdy przybywająca jednostka zastaje obsługę zajętą, w poczekalni tworzy się kolejka jednostek czekających. Dla sprecyzowania modelu potrzebne jest ustalenie *regulaminu kolejki,* określającego zachowywanie się jednostek w poczekalni oraz kolejność ich dopuszczenia do obsługi. Regulamin kolejki może w szczególności dotyczyć ograniczenia ilości miejsc w poczekalni (modele ze skończoną długością kolejki). Jednostki przybywające w momencie, gdy wszystkie miejsca w poczekalni są zajęte muszą zrezygnować z obsługi tworząc tak zwany *przelew.* Przelew jest możliwy także w przypadku, kiedy niektóre jednostki rezygnują z obsługi, gdy w momencie ich przybycia są jeszcze wolne miejsca w kolejce, ale według indywidualnej oceny przybywającej jednostki kolejka jest zbyt długa by warto było czekać na obsługę (filtracja procesu wejścia) lub gdy czekając przez pewien czas nie zostaną jeszcze obsłużone (zniecierpliwienie czekających). Najczęściej rozpatrywane są modele, w których przestrzegana jest kolejność obsługi zgodna z kolejnością zgłoszenia. Jeżeli obsługa ma kilka równoległych kanałów, to w poczekalni mogą się tworzyć oddzielne kolejki przed wejściem do każdego kanału (model kas biletowych na dworcu kolejowym) lub też jedna kolejka, z której jednostki wchodzą do wszystkich aktualnie dostępnych kanałów (model kolejki w zakładzie fryzjerskim). Rozpatruje się niekiedy systemy bez kolejki w poczekalni. W systemach takich wybiera się losowo jednostkę do obsługi spośród wszystkich jednostek czekających w poczekalni. Inne modele zakładają podział jednostek na klasy o różnych stopniach uprzywilejowania. W takich systemach do obsługi wybiera się każdorazowo najdłużej czekającą jednostkę z klasy o najwyższym stopniu uprzywilejowania. Ilość różnych klas uprzywilejowania może być skończona (np. pociągi towarowe, osobowe, pośpieszne i ekspresowe) lub nieskończona (np. wtedy, gdy z maszyn czekających na naprawę wybiera się w pierwszej kolejności tę, której przewidywany czas naprawy jest najkrótszy).

Obsługa może się składać z jednego lub więcej kanałów, w każdym z nich może być niezależnie obsługiwana jedna jednostka. Niekiedy w kanałach obsługi wyróżnia się kolejne stacje, na których wykonywane są poszczególne etapy obsługi. Obsługiwana jednostka przechodzi kolejno przez wszystkie stacje kanału, przy czym można rozpatrywać modele takie, że obsługiwana jednostka blokuje cały kanał i takie, w których możliwa jest jednoczesna obsługa na wszystkich stacjach kanału. W ostatnim przypadku mogą występować dodatkowe poczekalnie między poszczególnymi stacjami w kanale obsługi, które zapobiegają blokadzie stacji kończącej obsługę jednostki w chwili, gdy następna stacja jest jeszcze zajęta obsługiwaniem poprzedniej jednostki. Zakładając możliwość rozgałęzień i wewnętrznych połączeń między kanałami dochodzimy do modeli kolejkowych w sieciach. Do modeli tych możemy z drugiej strony dojść od zagadnień transportowych.

Czas potrzebny do obsługi jednostki w danym kanale, lub na danej stacji, jest zazwyczaj zmienną losową i dla sprecyzowania modelu należy podać łączne rozkłady wszystkich tych zmiennych losowych. W przypadku zagadnienia z priorytetem jednostki różnych klas uprzywilejowania mogą mieć różne rozkłady czasów obsługi. Zazwyczaj zakłada się, że czasy obsługi jednostek w różnych kanałach (lub na różnych stacjach) są między sobą niezależne, a także niezależne od procesu wejścia i sytuacji w poczekalni. Takie założenie znacznie ułatwia konstrukcję modelu matematycznego, ale w wielu sytuacjach praktycznych rzeczywistość wygląda inaczej.

Po dokładnym sprecyzowaniu modelu urządzenia obsługi masowej można szukać odpowiedzi na różne pytania o doniosłym znaczeniu praktycznym. Jedno z takich pytań może dotyczyć statystycznej stabilności systemu. System nazywamy *stabilnym*, jeżeli niezależnie od stanu wyjściowego z biegiem czasu ustala się pewien graniczny rozkład prawdopodobieństw możliwych stanów systemu. Przykładem niestabilnych systemów mogą być takie systemy, w których z dodatnim prawdopodobieństwem długość kolejki rośnie do nieskończoności wraz z upływem czasu, albo takie, w których występuje periodyczność zmian rozkładu prawdopodobieństwa stanów. Dla procesów stabilnych w większości przypadków praktycznych zadowalamy się znalezieniem granicznych prawdopodobieństw stanów. W niektórych sytuacjach, nie tylko w przypadku systemów niestabilnych, mogą nas interesować prawdopodobieństwa poszczególnych stanów jako funkcje czasu.

Można pytać o straty czasu jednostek czekających na obsługę, a w przypadku systemów z przelewem o frekwencję nieobsłużonych jednostek. Czas czekania jednostki jest zmienną losową i może nas interesować zarówno wartość oczekiwana czasu czekania jak też pełny rozkład tej zmiennej losowej. Inne pytania mogą dotyczyć obciążeń kanałów obsługi, rozkładu czasów nieprzerwanej pracy, wykorzystania poczekalni międzystacyjnych itp.

Odpowiedzi na te pytania (i wiele innych) dostarczają wskaźników efektywności systemu obsługi, pozwalają racjonalnie planować nowe urządzenia i ulepszać już istniejące. Przy uwzględnieniu kosztów inwestycji i kosztów operacyjnych można formułować i rozwiązywać różne zagadnienia optymalizacji ekonomicznej systemów.

2. Proces Poissona. Niech $X(t)$, dla $t \geqslant 0$, będzie ilością jednostek, które przybywają do urządzenia obsługi masowej w przedziale czasowym $\langle 0, t\rangle$. W teoretycznym modelu urządzenia $X(t)$ jest procesem stochastycznym. Każda realizacja tego procesu jest niemalejącą funkcją schodkową o wartościach całkowitych nieujemnych. Dla dowolnie ustalonego momentu $t' \geqslant 0$, $X(t')$ jest zmienną losową o wartościach całkowitych nieujemnych. Dla każdego t' i t'' $(t'' > t' \geqslant 0)$ różnica $X(t'')-X(t')$ jest ilością jednostek przybywających do systemu w przedziale czasowym $(t', t''\rangle$.

W różnych modelach teorii kolejek zakłada się często, że proces $X(t)$ jest *procesem Poissona*. Można go zdefiniować jako proces $X(t)$ spełniający następujące trzy warunki:

(a) Dla każdych $t \geqslant 0$ i $\tau > 0$ przyrost $X(t+\tau)-X(t)$ ma rozkład prawdopodobieństwa zależny tylko od długości τ przedziału czasowego $(t, t+\tau)$, a nie od położenia tego przedziału na osi czasowej. Można więc wprowadzić oznaczenie

$$P\{X(t+\tau)-X(t) = k\} = v_k(\tau), \qquad k = 0, 1, \ldots$$

(b) Dla dowolnej liczby rozłącznych przedziałów czasowych $(a_1, b_1\rangle$, $(a_2, b_2\rangle, \ldots, (a_n, b_n\rangle$; gdzie $0 \leqslant a_1 < b_1 \leqslant a_2 < b_2 \leqslant \ldots \leqslant a_n < b_n$, przyrosty procesu w tych przedziałach

$$X(b_1)-X(a_1), \quad X(b_2)-X(a_2), \quad \ldots, \quad X(b_n)-X(a_n)$$

są niezależnymi (*en bloc*) zmiennymi losowymi.

(c) Prawdopodobieństwo tego, że w małym przedziale czasowym o długości τ przyrost procesu $X(t)$ przekroczy jedność, dąży do zera szybciej niż długość przedziału

$$1 - v_0(\tau) - v_1(\tau) = o(\tau). \tag{1}$$

Z tych trzech warunków wynika (dowód można znaleźć np. w książce Fellera [8] lub Chinczyna [3]), że zmienna losowa $X(t+\tau)-X(t)$ ma rozkład Poissona

$$v_k(\tau) = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}, \qquad k = 0, 1, \ldots, \tag{2}$$

gdzie parametr $\lambda > 0$ jest oczekiwaną ilością jednostek przybywających do systemu w jednostce czasu.

Ze wzoru (2) wynika także, że

$$\lim_{\tau\to 0}\frac{1-v_0(\tau)}{\tau} = \lambda$$

albo inaczej

$$1-v_0(\tau) = \lambda\tau+o(\tau).$$

Ostatnia równość znaczy, że dla krótkich przedziałów czasu prawdopodobieństwo tego, że przybędzie co najmniej jedna jednostka jest, z dokładnością do małych niższego rzędu, proporcjonalne do długości przedziału. Parametr λ, zwany także *intensywnością procesu wejścia*, jest współczynnikiem tej proporcjonalności.

Niech $t_1, t_2, \ldots$ oznaczają momenty przybycia kolejnych jednostek do systemu. Momenty te są zmiennymi losowymi podobnie jak wielkości

$$z_1 = t_1, \quad z_2 = t_2-t_1, \quad z_3 = t_3-t_2, \quad \ldots \tag{3}$$

będące odstępami czasowymi między kolejno przybywającymi jednostkami. Łatwo jest sprawdzić, że proces wejścia $X(t)$ jest procesem Poissona wtedy i tylko wtedy, gdy spełnione są następujące dwa warunki:

(α) Odstępy czasowe z_i $(i = 1, 2, \ldots)$ są niezależnymi zmiennymi losowymi.

(β) Wszystkie odstępy z_i mają identyczny rozkład wykładniczy

$$F_i(x) = P\{z_i < x\} = \begin{cases} 0 & \text{dla} \quad x \leqslant 0; \\ 1-e^{-\lambda x} & \text{dla} \quad x > 0. \end{cases} \tag{4}$$

Ze wzoru (4) łatwo jest obliczyć wartość oczekiwaną odstępu czasowego między przybyciem dwu sąsiednich jednostek w procesie Poissona

$$Ez = \int_0^\infty x\,dF(x) = \frac{1}{\lambda}. \tag{5}$$

Wynik ten można było przewidzieć pamiętając o tym, że λ jest oczekiwaną ilością jednostek przybywających w jednostce czasu.

Wariancja odstępów między sąsiednimi zgłoszeniami jest równa

$$\sigma_z^2 = E(z-Ez)^2 = \int_0^\infty \left(x-\frac{1}{\lambda}\right)^2 dF(x) = \frac{1}{\lambda^2}. \tag{6}$$

Rozkład wykładniczy (4) ma charakterystyczną własność: dla dowolnych $x \geqslant 0$ i $a \geqslant 0$ zachodzi równość

$$P\{z > x+a \mid z > a\} = P\{z > x\}. \tag{7}$$

Równość ta ma w przypadku procesu wejścia następującą interpretację. Jeżeli obserwujemy system począwszy od ustalonego momentu t', to okres czekania na przybycie pierwszej jednostki jest niezależny od tego, jak dawno przed momentem t' przybyła do systemu poprzednia jednostka i czas czekania na pierwszą jednostkę ma taki sam rozkład jak cały odstęp między kolejnymi jednostkami([2]).

3. Uogólnienia procesu Poissona. W wielu przypadkach urządzeń obsługi masowej założenie, że proces wejścia jest procesem Poissona jaskrawo odbiega od rzeczywistości i trzeba szukać innych modeli lepiej przybliżających rzeczywistość. W tym celu mogą być wyzyskane pewne uogólnienia procesu Poissona, z których najczęściej stosowane omówimy tu w zarysie. Uogólnienia te będziemy konstruowali odrzucając lub osłabiając warunki (a), (b) i (c) składające się na definicję procesu Poissona.

Rozpatrzmy najpierw przypadek procesu spełniającego warunki (b) i (c). Po odrzuceniu warunku (a) prawdopodobieństwa przybycia k jednostek w przedziale czasowym $(t, t+\tau\rangle$ zależą zarówno od długości przedziału τ jak i od jego położenia t

$$P\{X(t+\tau)-X(t) = k\} = v_k(\tau, t),$$

gdzie $t \geqslant 0$, $\tau > 0$, $k = 0, 1, 2, \ldots$

W związku z nieistnieniem funkcyj $v_k(\tau)$ w warunku (c) musimy teraz zakładać, że dla każdego $t \geqslant 0$ i $\tau \to 0$ mamy

$$1-v_0(\tau, t)-v_1(\tau, t) = o(\tau).$$

Załóżmy dodatkowo, że dla każdego $t \geqslant 0$ istnieje granica

$$\lim_{\tau\to 0} \frac{1-v_0(\tau, t)}{\tau} = \lambda(t),$$

którą nazwiemy *chwilową intensywnością procesu w momencie t*.

([2]) Tak np. jeżeli na pewnej linii autobusy kursują przypadkowo, tak że momenty przyjazdu autobusów na ustalony przystanek tworzą proces Poissona, to czas czekania pasażera na autobus ma ten sam rozkład (i tę samą wartość oczekiwaną $1/\lambda$) co odstęp czasowy między kolejnymi autobusami.

Zdefiniowany w ten sposób proces znany jest pod nazwą *procesu Poissona o zmiennej intensywności* $\lambda(t)$. Ilość jednostek przybywających do systemu w dowolnym przedziale $(t, t+\tau\rangle$ ma rozkład Poissona

$$v_k(\tau, t) = e^{-\Lambda(\tau,t)} \frac{[\Lambda(\tau, t)]^k}{k!}, \tag{8}$$

gdzie $\Lambda(\tau, t) = \int\limits_t^{t+\tau} \lambda(u)du$.

Oczekiwana ilość jednostek przybywających w danym przedziale czasu jest więc całką z chwilowej intensywności procesu w tym przedziale.

Następne uogólnienie procesu Poissona otrzymujemy odrzucając warunek (c). Z warunków (a) i (b) wynika, że proces stochastyczny momentów, w których jednostki przybywają do systemu jest procesem Poissona a w każdym z tych momentów jedna lub więcej jednostek przybywa do systemu. Ilości jednostek przybywających do systemu w kolejnych momentach przybycia są niezależnymi (*en bloc*) zmiennymi losowymi mającymi ten sam rozkład $P\{n = k\} = a_k$ $\left(\sum\limits_{k=1}^{\infty} a_k = 1\right)$. Dla pełnego określenia procesu należy podać parametr λ (oczekiwaną ilość momentów przybycia w jednostce czasu) i ciąg prawdopodobieństw $\{a_k\}$. Oczekiwana liczba jednostek przybywających do systemu w jednostce czasu (jeśli istnieje(³)) jest równa sumie

$$\lambda \sum_{k=1}^{\infty} k a_k. \tag{9}$$

4. Proces Palma i stacjonarny proces punktowy. Zajmiemy się teraz klasą tak zwanych *procesów Palma*. Definicję procesu Palma uzyskamy zastępując warunek (b) w zespole trzech warunków określających proces Poissona przez słabszy(⁴) warunek (α) ze str. 17. Zamiast zakładać, że ilości przybywających jednostek w rozłącznych przedziałach czasowych są niezależne, przyjmiemy teraz, że odstępy czasowe (3) między przybyciem kolejnych jednostek są niezależne. Można wykazać (patrz [3]), że dla ustalonego procesu Palma wszystkie odstępy (3) mają ten sam rozkład

$$P\{z_i < x\} = F(x), \quad i = 2, 3, \ldots \tag{10}$$

z wyjątkiem pierwszego odstępu z_1, który ma rozkład

$$P\{z_1 < x\} = \lambda \int_0^x [1-F(u)]du,$$

(³) Suma (9) może być nieskończona. Takie przypadki nie są jednak ciekawe z punktu widzenia zastosowań teorii obsługi masowej.

(⁴) Warunek (α) jest słabszy od warunku (b) tylko w klasie procesów spełniających dodatkowo warunki (a) i (c).

gdzie współczynnik λ można wyznaczyć z warunku

$$P\{z_1 < \infty\} = \lambda \int_0^\infty [1-F(u)]du = 1.$$

W tym celu musi oczywiście istnieć skończona wartość całki niewłaściwej

$$\int_0^\infty [1-F(u)]du < \infty. \tag{11}$$

Ten ostatni warunek jest równoznaczny z postulatem, by istniała skończona wartość oczekiwana odstępu między przybyciem kolejnych jednostek

$$Ez = \int_0^\infty u\,dF(u) = \int_0^\infty [1-F(u)]du = \frac{1}{\lambda}. \tag{12}$$

Tak więc proces Palma jest jednoznacznie określony przez zadanie dystrybuanty $F(x)$ odstępów między przybyciem kolejnych jednostek, przy założeniu istnienia ograniczonej wartości oczekiwanej (12). Proces Poissona jest szczególnym przypadkiem procesu Palma o wykładniczym rozkładzie odstępów (4).

Specjalnie ciekawym przypadkiem procesu Palma, często stosowanym do opisu procesu wejścia, jest proces, dla którego $F(x) = F_{k,k\lambda}(x)$ jest dystrybuantą sumy k niezależnych zmiennych losowych o identycznych rozkładach wykładniczych z parametrem $k\lambda$. Jest wtedy

$$F_{k,k\lambda}(x) = \begin{cases} 0 & \text{dla} \quad x \leqslant 0; \\ \dfrac{(k\lambda)^k}{(k-1)!}\displaystyle\int_0^x e^{-k\lambda u}u^{k-1}du & \text{dla} \quad x > 0. \end{cases} \tag{13}$$

Rozkład (13) znany jest w statystyce jako *rozkład* χ^2 *Pearsona.* W teorii kolejek jest on zwykle nazywany *rozkładem Erlanga.* Jeżeli odstępy z mają rozkład (13), to ich wartość oczekiwana i wariancja są równe

$$Ez = \frac{1}{\lambda}, \quad \sigma_z^2 = \frac{1}{k\lambda^2}$$

i dla $k \to \infty$ mamy $\sigma_z^2 \to 0$.

Dystrybuanta (13) jest więc dla $k = 1$ dystrybuantą (4) rozkładu wykładniczego a przy $k \to \infty$ ma jako granicę funkcję schodkową

$$F_{\infty,\lambda}(x) = \begin{cases} 0 & \text{dla} \quad x \leqslant \dfrac{1}{\lambda}; \\ 1 & \text{dla} \quad x > \dfrac{1}{\lambda}, \end{cases} \tag{14}$$

będącą dystrybuantą stałej wartości $z = 1/\lambda$.

Zauważmy, że dowolna dystrybuanta $F(x)$ nieujemnej zmiennej losowej z o skończonej wartości oczekiwanej (12) może być jednostajnie aproksymowana przez liniową kombinację dystrybuant (13). Wynika to z faktu, że każdą dystrybuantę można jednostajnie aproksymować przez niemalejące funkcje schodkowe, które mogą być przedstawione jako liniowe kombinacje funkcyj (14).

Jeszcze ogólniejszym procesem niż proces Palma jest *stacjonarny proces punktowy*. Należałoby go poprawniej nazwać *procesem punktowym o stacjonarnych przyrostach*. Punktowość procesu polega na tym, że realizacjami są linie schodkowe, mające w dyskretnych momentach czasu dodatnie przyrosty całkowite a w każdym ograniczonym przedziale czasowym całkowity przyrost jest z prawdopodobieństwem jeden skończony. Stacjonarność przyrostów oznacza, że wszystkie probabilistyczne charakterystyki procesu, dotyczące przyrostów na dowolnym układzie odcinków czasowych, są niezmiennikami sztywnej translacji danego układu odcinków.

5. Symbolika Kendalla. W pracy ogłoszonej w roku 1953 D. G. Kendall ([13]) zaproponował przejrzystą symbolikę, która pozwala w skrócie zapisać niektóre założenia o modelu obsługi masowej. Jak widzieliśmy, pełna specyfikacja modelu wymaga określenia procesu wejścia, rozkładu czasów obsługi, ilości kanałów oraz regulaminu kolejki. Kendall proponuje, aby założenia precyzujące pierwsze trzy charakterystyki systemu zapisywać w postaci symbolu

$$X/Y/n,$$

gdzie: X charakteryzuje proces wejścia;
Y charakteryzuje rozkład czasów obsługi;
n jest liczbą równoległych kanałów obsługi.

Oto najczęściej używane symbole na miejscu X i Y:

D — deterministyczny proces wejścia (stałe odstępy między zgłoszeniami kolejnych jednostek) lub stały czas obsługi;

M — proces Poissona na wejściu lub wykładniczy rozkład czasów obsługi;

E_k — rozkład Erlanga czasów obsługi lub proces Palma na wejściu z rozkładem Erlanga odstępów między kolejnymi zgłoszeniami;

G — dowolny proces wejścia lub dowolny rozkład czasów obsługi;

GI — (symbol używany tylko na miejscu X) proces Palma o dowolnej dystrybuancie niezależnych odstępów.

Tak np. symbol $M/G/5$ oznacza pięciokanałowy model obsługi masowej z procesem Poissona na wejściu i dowolnym rozkładem czasów obsługi. Jeśli nie podano słownie dodatkowych założeń, dotyczących dyscypliny kolejki itp. należy się domyślać, że chodzi o najprostszy model z jedną kolejką, obsługą w kolejności wejścia do systemu, brak priorytetów i rezygnacji oraz nieograniczoną pojemność poczekalni.

Symbolika Kendalla bywa w razie potrzeby uzupełniana przez różnych autorów. Tak np. F. A. Haight ([11]) używa litery S na oznaczenie stacjonarnego procesu punktowego i rozważa w swej pracy modele typu $M/S/1$ i $S/M/1$.

6. Metody matematyczne stosowane w teorii kolejek. Przy rozwiązywaniu różnych zagadnień związanych z konkretnymi modelami urządzeń obsługi masowej stosowane są różnorodne metody matematyczne. Nie siląc się na pełną systematykę stosowanych metod postaramy się wskazać najczęściej spotykane podejścia.

(a) Dyskretne procesy Markowa. Stan systemu obsługi masowej w ustalonym momencie t najprościej jest scharakteryzować przez podanie ilości jednostek $N(t)$, które w danej chwili znajdują się w systemie (w obsłudze i w poczekalni łącznie). Gdy t jest zmienne i przebiega wartości z półprostej $t \geqslant 0$ ilość jednostek w systemie $N(t)$ jest procesem stochastycznym o skończonej lub przeliczalnej ilości stanów. W najprostszych systemach typu $M/M/n$ proces ten jest *procesem Markowa* ([5]). Dzięki tej własności można napisać układ liniowych równań różniczkowych rzędu pierwszego na prawdopodobieństwa poszczególnych stanów systemu, traktowane jako funkcje czasu. Układ ten można niekiedy sprowadzić do jednego równania różniczkowego cząstkowego na funkcję tworzącą procesu. Dla jednoznaczności rozwiązania potrzebne są pewne warunki brzegowe, najczęściej określane przez podanie rozkładu prawdopodobieństw stanów początkowych (dla $t = 0$).

W wielu przypadkach, gdy mamy do czynienia z modelem stabilnym, zadowalamy się wyznaczeniem granicznych prawdopodobieństw stanów. Przechodząc do granicy, przy $t \to \infty$, z układu równań różnicz-

---

([5]) Dyskretny proces $N(t)$ jest procesem Markowa, jeżeli dla dowolnego ciągu momentów

$$t_{-s} < \ldots < t_{-1} < t_0 < t_1 < \ldots < t_r$$

i dowolnego ciągu możliwych stanów procesu

$$k_{-s}, \ldots, k_{-1}, k_0, k_1, \ldots, k_r$$

zachodzi równość

$$P\{X(t_1) = k_1, \ldots, X(t_r) = k_r / X(t_0) = k_0, X(t_{-1}) = k_{-1}, \ldots, X(t_{-s}) = k_{-s}\} =$$
$$= P\{X(t_1) = k_1, \ldots, X(t_r) = k_r / X(t_0) = k_0\}.$$

Innymi słowy, jeśli znany jest stan systemu w chwili t_0, to dodatkowa informacja o stanach systemu w chwilach wcześniejszych nie ułatwia prognozy stanu systemu w dowolnych chwilach późniejszych.

kowych uzyskamy jednorodny układ algebraicznych równań liniowych. Dodatkowym warunkiem gwarantującym jednoznaczność rozwiązania jest sumowalność granicznych prawdopodobieństw do jedności.

W niektórych bardziej skomplikowanych modelach (np. przy istnieniu priorytetów albo specjalnych regulaminów kolejkowych) stan systemu jest wektorowym procesem Markowa (którego współrzędne oznaczają np. ilości jednostek w systemie o różnych klasach uprzywilejowania). Ponieważ jednak nadal ilość różnych stanów jest co najwyżej przeliczalna, więc ich prawdopodobieństwa (zależne od czasu lub graniczne) wyznacza się z analogicznych układów równań liniowych (różniczkowych lub algebraicznych) jak i w rozważanym przypadku jednowymiarowym.

(b) Ciągłe procesy Markowa. Niekiedy stan systemu wygodniej jest opisywać za pomocą ciągłego procesu stochastycznego (to jest takiego procesu, dla którego zbiór różnych możliwych stanów jest zbiorem o mocy continuum). Tak np. stan systemu można opisać podając w każdej chwili t potencjalny czas czekania $W(t)$, to jest czas, który upłynie do zakończenia obsługi wszystkich jednostek znajdujących się w danej chwili w systemie. Tyle czasu musiałaby czekać na rozpoczęcie obsługi jednostka, która weszłaby do systemu w momencie t. L. Takács ([25]) pokazał, że w systemach $M/G/1$ proces $W(t)$ jest procesem Markowa. Dzięki tej własności funkcja $F(t, x)$, oznaczająca dystrybuantę procesu $W(t)$ w chwili t, spełnia pewne równanie różniczkowo-całkowe, które można rozwiązać za pomocą transformaty Laplace'a-Stieltjesa.

W podobny sposób można również potraktować niektóre inne modele obsługi masowej. Nawet wtedy, gdy w efekcie zastosowania tej metody dochodzi się do równań trudnych lub niemożliwych do rozwiązania, można na ich podstawie uzyskać pewne informacje o stabilności rozwiązań, momentach itp.

(c) Zlepianie procesu Markowa. Niech $X(t)$ będzie procesem stochastycznym. Dowolna funkcja f określona na zbiorze możliwych stanów procesu $X(t)$ wyznacza nowy proces $Y(t) = f[X(t)]$. Jeśli takie odwzorowanie nie jest jedno-jednoznaczne, to pojedyncze stany procesu $Y(t)$ mogą odpowiadać wieloelementowym podzbiorom stanów procesu $X(t)$. W takim przypadku odwzorowanie f nazywamy *zlepianiem*, a proces $Y(t)$ — *procesem zlepionym*. Tak np. jeżeli $X(t)$ jest wektorowym procesem stochastycznym $X(t) = [X_1(t), X_2(t), \ldots, X_n(t)]$, to każda z jego współrzędnych jest procesem zlepionym. Podobnie procesem zlepionym jest suma

$$Y(t) = \sum_{i=1}^{n} X_i(t).$$

Zlepianie nie zachowuje na ogół markowskości procesu. Jeżeli $X(t)$ jest procesem Markowa, to $Y(t)$ na ogół nie ma już tej własności. Znajomość probabilistycznych charakterystyk procesu $X(t)$ umożliwia jednak pełny opis procesu $Y(t)$. Tę możliwość wyzyskuje się do badania szerokiej klasy systemów obsługi masowej. Zdarza się bowiem często, że stan systemu $Y(t)$ nie jest procesem Markowa, ale można skonstruować proces Markowa $X(t)$ taki, że $Y(t)$ jest procesem zlepionym w stosunku do $X(t)$. Tak np. w modelu $M/E_k/1$ proces $Y(t)$ oznaczający ilość jednostek w systemie w chwili t nie jest procesem Markowa. Jeśli jednak rozbudujemy model przyjmując fikcyjnie, że kanał obsługi ma k stacji o wykładniczym czasie obsługi na każdej z nich (zakładamy przy tym, że obsługa prowadzona na jednej stacji blokuje cały kanał), to wektorowy proces

$$X(t) = [X_0(t), X_1(t), \ldots, X_k(t)],$$

gdzie $X_0(t)$ oznacza ilość jednostek w poczekalni w chwili t;
$X_i(t)$ oznacza ilość jednostek na i-tej stacji([6]), $i = 1, 2, \ldots, k$,
jest procesem Markowa, a interesujący nas proces

$$Y(t) = \sum_{i=0}^{k} X_i(t)$$

jest procesem zlepionym.

W ten sam sposób, wprowadzając fikcyjne elementy, można skonstruować wektorowe procesy Markowa takie, że zlepione procesy $Y(t)$ opiszą stan systemu w modelach $E_k/M/1$ i $E_k/E_1/1$, a także w analogicznych modelach z większą ilością równoległych kanałów. W podobny sposób B. Kopociński ([17]), wprowadzając fikcyjną kwarantannę na wejściu do systemu, skonstruował zlepiony proces opisujący stan systemu z opóźnionym sprzężeniem zwrotnym regulującym intensywność procesu wejścia w zależności od ilości jednostek w systemie.

(d) Metody aproksymacji systemu. Odpowiednie rozbudowanie systemu przez wprowadzenie fikcyjnych elementów i wielowymiarowych procesów wektorowych pozwala sprowadzić do badania procesów Markowa każde zagadnienie dotyczące systemów, w których (przy dowolnej ilości kanałów) czasy obsługi i niezależne odstępy między kolejnymi zgłoszeniami mają rozkłady będące kombinacjami liniowymi rozkładów Erlanga (13). Ponieważ, jak wspomnieliśmy na str. 21, każdy rozkład można dowolnie dokładnie aproksymować przez kombinacje liniowe rozkładów Erlanga, więc stosując taką aproksymację i metodę zlepiania procesu Markowa można w przybliżony sposób rozwiązać każ-

([6]) $X_i(t)$ dla $i = 1, 2, \ldots, k$ przyjmuje tylko dwa stany 0 lub 1, przy czym suma $\sum_{i=1}^{k} X_i(t)$ także przyjmuje tylko dwie wartości 0 lub 1.

de zagadnienie typu $GI/G/n$. Niestety nie ma dotychczas żadnych opracowań teoretycznych dających analizę dokładności takiej aproksymacji zagadnienia.

(e) Włożone łańcuchy Markowa. Dla pewnych niemarkowskich procesów $X(t)$, opisujących stan systemu obsługi masowej dla ciągłego czasu t, udaje się wybrać taki dyskretny ciąg momentów $\{t_i\}$, że stany procesu w tych wybranych momentach tworzą ciąg zmiennych losowych $\{X(t_i)\}$ będący łańcuchem Markowa. Tak np. w modelu $M/G/1$ ilość jednostek w systemie $X(t_i)$ w momencie t_i wyjścia z obsługi i-tej jednostki jest takim *włożonym łańcuchem Markowa*. Proces $X(t)$, zależny od ciągłego parametru, jest w tym modelu procesem Markowa tylko wtedy, gdy rozkład czasów obsługi jest rozkładem wykładniczym (a więc gdy system $M/G/1$ redukuje się do systemu $M/M/1$). Inne przykłady włożonych łańcuchów Markowa można znaleźć np. w pracy I. Kopocińskiej [16].

Wyznaczenie probabilistycznych charakterystyk włożonego łańcucha Markowa $\{X(t_i)\}$ nie daje na ogół pełnej znajomości procesu $X(t)$. Wiele zagadnień dotyczących systemów obsługi masowej może być jednak rozwiązanych przy ograniczeniu się do badania tylko włożonego łańcucha Markowa. Rezygnacja z badania niemarkowskiego procesu $X(t)$ i ograniczenie się do badania włożonego łańcucha Markowa $\{X(t_i)\}$ jest w pewnym sensie postępowaniem odwrotnym do omówionej poprzednio metody traktującej stan systemu $X(t)$ jako zlepienie pewnego rozbudowanego procesu Markowa.

(f) Inne metody. W krótkim artykule przeglądowym autor nie pretenduje do omówienia wszystkich metod stosowanych w konkretnych przypadkach zagadnień kolejkowych. Tak np. F. A. Haight ([11]) wyzyskuje w swej pracy wyniki uzyskane metodą kombinatoryczną. J. Rzytka ([21]) próbuje stosować do modeli obsługi masowej teorię zdarzeń rekurencyjnych.

Można by jeszcze mnożyć przykłady innych ujęć i prób, jednakże, mimo całej różnorodności środków matematycznych i metod, zakres wszystkich dotychczas rozwiązanych zagadnień jest jeszcze daleko niewystarczający w różnorodnych sytuacjach praktycznych. Najczęściej powtarzające się założenia teoretyczne, przyjmowane zazwyczaj dla ułatwienia uzyskania analitycznych rozwiązań, są niejednokrotnie nie do przyjęcia w konfrontacji z rzeczywistością.

Uniwersalną metodą, coraz częściej stosowaną w praktyce, jest badanie pracy urządzeń obsługi masowej przez modelowanie tych systemów na elektronicznych maszynach cyfrowych. Metoda Monte Carlo([7]) może być w zasadzie zastosowana do każdego zagadnienia teorii kolejek,

([7]) Zasady stosowania metody Monte Carlo, także w odniesieniu do zagadnień obsługi masowej, znaleźć można np. w książce N. P. Buslenki i współautorów [2].

jeśli tylko pozwalają na to parametry techniczne dostępnej maszyny cyfrowej. Największe trudności przy stosowaniu tych metod polegają na poprawnej ocenie dokładności uzyskiwanych rozwiązań.

Prace cytowane

[1] V. E. Beneš, *General stochastic processes in the theory of queues*, Reading 1963.

[2] Н. П. Бусленко, Д. И. Голенко, И. М. Соболь, В. Г. Срагович, Й. А. Шрейдер, *Метод статистических испытаний (метод Монте-Карло)*, Москва 1962.

[3] А. Я. Хинчин, *Математические методы теории массового обслуживания*, Труды Математического Института им. В. А. Стеклова 49 (1955), wydanie drugie z komentarzami B. W. Gniedenki ogłoszone jest w zbiorze [4].

[4] — *Работы по математической теории массового обслуживания*, Москва 1963.

[5] D. R. Cox, W. L. Smith, *Queues*, London 1961.

[6] A. Doig, *A bibliography on the theory of queues*. Biometrica 44 (1957), str. 490-514.

[7] A. K. Erlang, *Collected works*, Copenhagen 1948.

[8] W. Feller, *Wstęp do rachunku prawdopodobieństwa i jego zastosowań*, Warszawa 1960.

[9] B. W. Gniedenko, *Über einige Aspekte der Entwicklung der Theorie der Warteschlangen*, Mathematik-Technik-Wirtschaft 7 (1960), str. 162-166.

[10] Б. В. Гнеденко, *Курс лекций по теории массового обслуживания* (powielony tekst wykładu), София 1964.

[11] F. A. Haight, *The discrete busy period distribution for various single server queues*, Zastosow. Mat. 8 (1965), str. 37-46.

[12] D. G. Kendall, *Some problems in the theory of queues*, Journal of the Royal Statistical Society, series B, 13 (1951), str. 151-185.

[13] — *Stochastic processes occuring in the theory of queues and their analysis by the method of the imbedded Markov chain*, Ann. Math. Stat. 24 (1953), str. 338-354.

[14] — *Some recent work and further problems in the theory of queues*, Теория вероятностей и ее применения 9 (1964), str. 3-15.

[15] I. Kopocińska, *O pewnym modelu z teorii kolejek z uwzględnieniem zniecierpliwienia klientów*, Zastosow. Mat. 7 (1963), str. 41-50.

[16] — *Imbedded Markov chains in two examples of queueing models*, Zastosow. Mat. 8 (1965), str. 29-36.

[17] B. Kopociński, *On a model of a queue with delayed feedback*, Zastosow. Mat. 7 (1963), str. 51-57.

[18] P. LeGall, *Les systèmes avec ou sans attente et les processus stochastiques*, Paris 1962.

[19] Ph. M. Morse, *Queues, inventories and maintenance*, New York 1958.

[20] J. Riordan, *Stochastic service systems*, New York 1962.

[21] J. Rzytka, *Recurrent events in queues with independent arrival intervals*, Zastosow. Mat. 8 (1965), str. 47-56.

[22] Th. L. Saaty, *Elements of queueing theory with applications*, New York 1962.

[23] W. Sadowski, *Teoria podejmowania decyzji*, wydanie 2, Warszawa 1963.

[24] R. Syski, *Introduction to congestion theory in telephone systems*, Edinburgh and London 1960.

[25] L. Takács, *Investigation of waiting time problems by reduction to Markov processes*, Acta math. Acad. Sci. Hungar. 6 (1955), str. 101-129.

[26] — *Introduction to the theory of queues*, New York 1962.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 3. 10. 1964

Й. ЛУКАШЕВИЧ (Вроцлав)

ТЕОРИЯ ОЧЕРЕДЕЙ ИЛИ МАССОВОГО ОБСЛУЖИВАНИЯ

РЕЗЮМЕ

В статье дается краткий обзор основных понятий, вопросов и применяемых методов решения задач теории массового обслуживания.

J. ŁUKASZEWICZ (Wrocław)

QUEUEING THEORY OR THE THEORY OF MASS SERVICE

SUMMARY

In the paper author gives a short review of basic concepts, problems and methods applied in the theory of queues.