Рівненський державний гуманітарний університет

Факультет математики та інформатики

Кафедра вищої математики

Кваліфікаційна робота

магістерського рівня

на тему:

Застосування теорії масового обслуговування при моделюванні ймовірнісних моделей

Виконала: студентка II курсу магістратури

Спеціальності 014 Середня освіта (Математика)

Петрачик Лілія Олександрівна

Керівник: д.т.н., проф. Бичков О. С.

Рецензент д.т.н. проф. Сафоник А. П.

Рівне-2022 року

ЗМІСТ

ВСТУП

Складний характер ринкової економіки та сучасний рівень вимог, що пред’являються, стимулюють використання більш серйозних методів аналізу її теоретичних та практичних проблем. В останні десятиліття значної ваги в економічних дослідженнях набули математичні методи. Математичне моделювання все більше і більше стає одним з основних і найбільш плідних методів вивчення економічних процесів та об'єктів. Математичний аналіз економічних задач органічно перетворюється на частину економіки. Позитивна оцінка цього підтверджується і тим, що починаючи з 1969 р. Нобелівські премії у сфері економіки присуджуються, зазвичай, за економіко-математичні дослідження.

Одним із важливих розділів економіко-математичного моделювання є теорія масового обслуговування, що представляє собою теоретичні основи ефективного конструювання та експлуатації систем масового обслуговування. Системи масового обслуговування (СМО) зустрічаються в багатьох галузях економіки (виробництво, техніка, військова сфера, побут та ін) і призначені для багаторазового використання при виконанні однотипних завдань.

У боротьбу за клієнта у сучасній економіці вкладаються величезні кошти. За оцінкою західних економістів, завоювання фірмою нового клієнта обходиться їй у 6 разів дорожче, ніж утримання існуючих покупців. А якщо клієнт пішов незадоволеним, то на його повернення доводиться витратити в 25 разів більше коштів. У багатьох випадках незадоволеність клієнта викликана невдалою організацією його обслуговування (занадто довге очікування в черзі, відмова в обслуговуванні тощо). Використання теорії масового обслуговування дозволяє фірмі уникнути подібних неприємностей [28].

Основоположником теорії масового обслуговування вважається датський вчений А. К. Ерланг. Будучи співробітником Копенгагенської телефонної компанії, він опублікував у 1909 році роботу «Теорія ймовірностей та телефонні переговори», в якій розв’язав ряд задач з теорії систем масового обслуговування із відмовами.

Значний внесок у створення та розробку загальної теорії масового обслуговування вніс видатний радянський математик Олександр Якович Хінчін, котрий запропонував сам термін теорія масового обслуговування. У зарубіжній літературі частіше використовується назва теорія черг.

Мета роботи: систематизувати відомості про застосування теорії масового обслуговування до моделювання ймовірнісних моделей.

Об'єкт дослідження: теорія масового обслуговування.

Предмет дослідження: системи масового обслуговування.

Завдання роботи:

1) зробити огляд основних понять систем масового обслуговування;

2) розглянути математичні моделі різних конкретних систем масового обслуговування виробничих процесів та розрахувати їх функціональні показники;

3) створити методичний посібник на основі дослідження.

Робота складається зі вступу, основної частини, що містить три розділи, висновків та списку використаних джерел.

РОЗДІЛ 1. ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ МАСОВОГО ОБСЛУГОВУВАННЯ

1.1. Поняття черги

Очікування того чи іншого виду обслуговування є частиною нашого повсякденного життя. Ми очікуємо, щоб пообідати у ресторані, стоїмо у черзі до кас у продуктових магазинах та вишиковуємось у чергу у поштових відділеннях. Однак феномен очікування характерний не тільки для людей: деталі, поставлені в чергу для обробки на верстаті; група пасажирських літаків, що чекають дозволу на посадку в аеропорту; автомобілі, рух яких призупинено сигналом світлофора на шляху їхнього прямування тощо. На жаль, феномен очікування не можна виключити без надмірних витрат. І лише на одне ми можемо сподіватися на можливість скорочення часу не бажаного очікування у черзі до деяких прийнятних меж [21].

Вивчення черг у системах масового обслуговування дозволяє визначити критерії функціонування обслуговуючої системи, серед яких найбільш значимі є середній час очікування в черзі та середня тривалість черги. Ця інформація використовується для вибору належного рівня обслуговування, що продемонстровано в першому прикладі 3-го розділу.

1.2. Основні компоненти моделей масового обслуговування

Основними елементами моделі масового обслуговування є клієнт (заявка або вимога на обслуговування або просто «об'єкт обслуговування») та сервіс (обслуговуючий пристрій, засоби обслуговування тощо). Клієнти надходять у систему обслуговування із джерела. Звернувшись до сервісу, вони можуть відразу потрапити на обслуговування або чекати в черзі, якщо сервіс зайнятий. Після завершення процедури обслуговування сервіс автоматично «вибирає» із черги (якщо вона є) одного з клієнтів для того, щоб приступити до його обслуговування. Якщо ж черга відсутня, то сервіс стає незайнятим до прибуття нового клієнта [1].

Надходження клієнтів до системи обслуговування характеризується інтервалом між їхніми послідовними надходженнями, а обслуговування – часом обслуговування клієнта. У загальному випадку ці параметри можуть бути випадковими, як, наприклад, у роботі поштового відділення, та детермінованими, як, наприклад, прибуття на співбесіду кандидатів на вакантну посаду [6].

В аналізі систем обслуговування певну роль відіграє довжина черги, яка може бути скінченною, як у буферній зоні між двома послідовними обслуговуючими пристроями, та нескінченною, як у обслуговуючих операторів поштових відділень.

Важливим фактором під час аналізу систем обслуговування є дисципліна черги (або принципи побудови черги), що визначає порядок, відповідно до якого вибираються клієнти з черги обслуговування. Найбільш поширений принцип побудови черги заснований на правилі «першим прийшов – першим обслуговуєшся» (це правило часто позначається абревіатурою FIFO – від англійського First-In-First-Out, тобто, «першим увійшов – першим вийшов»). Серед інших правил, що визначають принципи побудови черг, зазначимо правило «останнім прийшов – першим обслуговуєшся» (зазвичай позначається, як LIFO – від англійського Last-In-First-Out, тобто «останнім увійшов – першим вийшов») та дисципліну черги, що визначається випадковим правилом відбору клієнтів (іноді позначається як SIRO – від англійського Service-In-Random-Out, тобто «обслуговування за випадковим принципом»). Крім того, клієнти можуть вибирати з черги відповідно до заданого пріоритету. Наприклад, у виробничому цеху термінові роботи виконуються раніше, ніж звичайні [10].

При аналізі систем із чергами важливим фактором є також поведінка індивідуума, який потребує обслуговування. Такі індивідууми, що виступають у ролі клієнтів, за наявності паралельного обслуговування можуть перейти з однієї черги до іншої в надії скоротити тривалість свого вимушеного очікування. Вони можуть також відмовитися від очікування в черзі, оскільки люди зазвичай не терплять тривалої бездіяльності, або залишити чергу, простоявши в ній якийсь час і дійшовши висновку, що й так аж надто багато часу втрачено.

Структура обслуговуючої системи може включати один сервіс або кілька: до таких засобів обслуговування, що працюють паралельно (наприклад, робота кількох клерків поштового відділення). Крім того, сервіси можуть бути розташовані послідовно (наприклад, обслуговування є комплексом робіт, що виконуються послідовно на різних верстатах).

Джерело, що генерує «клієнтів», що підлягають обслуговуванню, може мати скінченну чи нескінченну потужність. Джерело скінченної потужності обмежує кількість клієнтів, що надходять на обслуговування (наприклад, в цеху, що має N верстатів, сумарна кількість потенційних заявок на їх ремонт не перевищує N). Навпаки, джерело нескінченної потужності завжди має клієнтів удосталь (наприклад, дзвінки, що надходять на телефонну станцію).

Можна побудувати безліч моделей систем масового обслуговування, варіюючи перелічені вище операційні характеристики систем. У цьому розділі розглядається низка таких моделей.

1.3. Експоненціальний розподіл в системах масового обслуговування

1.3.1. Властивість відсутності післядії

У більшості систем масового обслуговування надходження клієнтів відбувається випадковим чином. Це означає, що настання події (наприклад, надходження клієнта або завершення обслуговування) не залежить від часу, що минув з настання попередньої події.

Час між послідовними надходженнями клієнтів та час їх обслуговування, будучи випадковими, при моделюванні систем масового обслуговування кількісно описуються експоненціальним розподілом, густина ймовірності якого має вигляд

$$f(t) = \lambda e^{-\lambda t},\ t > 0,$$

де $M\{t\} = 1/\lambda$. Те, що експоненціальний розподіл є випадковим, ілюструється наступним прикладом. Якщо зараз 8:20 та якась подія відбулася о 8:02, то відповідно до експоненціального закону розподілу ймовірність того, що наступна аналогічна подія відбудеться о 8:29, є функцією лише інтервалу часу від 8:20 до 8:29 і не залежить від інтервалу часу, що минув з моменту настання останньої події (від 8:02 до 8:20). Дану властивість експоненціального розподілу зазвичай називають відсутністю післядії або відсутністю пам'яті [9].

Нехай за час t настання будь-якої події розподілено за експотенціальним законом із функцією густини $f(t)$. Якщо S – час, що минув з моменту появи попередньої події, то властивість відсутності післядії виражається співвідношенням

$$P\{t > T + S | t > S\} = P\{t > T\}$$

Для доведення цієї рівності помітимо, що

$$P\{t > Y\} = 1 - P\{t < Y\} = e^{-\lambda Y}.$$

Отже,

$$P\{t > T + S | t > S\} = \frac{P\{t > T + S, t > S\}}{P\{t > S\}} = \frac{P\{t > T + S\}}{P\{t > S\}} = \frac{e^{-\lambda(T+S)}}{e^{-\lambda S}} = e^{-\lambda T} = P\{t > T\}.$$

1.3.2. Визначення експоненціального розподілу

Щоб усвідомити властивості експоненціального розподілу, сформулюємо основні аксіоми, на яких базується цей розподіл.

Аксіома 1. Якщо $N(t)$ число подій, що відбулися протягом інтервалу часу $(0,t)$, то ймовірнісний процес, що описує $N(t)$, має незалежні стаціонарні прирости; ймовірність появи події в інтервалі $(T, T+S)$ залежить лише від його довжини S.

Аксіома 2. Імовірність того, що подія настане на досить малому часовому інтервалі $h > 0$, більша нуля, але менше 1.

Аксіома 3. На досить малому часовому інтервалі $h > 0$ може з'явитися небільше однієї події, тобто $P\{N(h) > 1\} = 0$.

Позначимо через $p_n(t)$ ймовірність появ n подій на часовому інтервалі довжиною t. Відповідно до аксіоми 1 ймовірність того, що на інтервалі довжиною $t + h$, де $h > 0$ і досить мале, не з'явиться жодної події, дорівнює

$$p_0(t+h) = p_0(t)p_0(h)$$

Спираючись на дві інші аксіоми, можна показати, що розв'язком наведеного вище рівняння є

$$p_0(t) = e^{-\lambda t},\ t > 0$$

де λ – додатна константа. Нехай $f(t)$ – густина ймовірності розподілу довжини часового інтервалу між послідовними появами випадкової події $(t > 0)$.

Очевидно, що

$$P\left\{\begin{matrix}\text{інтервал часу між моментами}\\ \text{появою двох послідовних}\\ \text{випадкових подій не менше } T\end{matrix}\right\} = P\left\{\begin{matrix}\text{за } T \text{ подія}\\ \text{не появиться}\end{matrix}\right\}.$$

У математичній формі це виглядає так:

$$\int_T^\infty f(t)dt = p_0(T),\ T > 0$$

Підставивши вираз для $p_0(T)$ та виконавши невеликі перетворення, отримуємо наступне.

$$\int_0^T f(t)dt = 1 - e^{-\lambda T},\ T > 0$$

продиференціювавши праву та ліву частину цього співвідношення по T, отримуємо вираз для густини експотенціального розподілу

$$f(t) = \lambda e^{-\lambda t},\ t > 0$$

Математичне сподівання експоненціального розподілу дорівнює $1/\lambda$ часових одиниць, де λ – інтенсивність (кількість подій за одиницю часу), з якої виникають події.

1.4. Характеристики якості обслуговування

1.4.1. Системи з втратами

Для будь-якої телекомунікаційної системи важливою є оцінка ступеня задоволення потреби в обслуговуванні, або якість обслуговування (QoS – Quality of Service). В теорії телетрафіка якість обслуговування потоку вимог характеризується можливістю негайного обслуговування вимоги або тривалістю очікування початку обслуговування. З математичної моделі систем розподілу інформації (СРІ) випливає, що ці можливості визначаються обраною дисципліною обслуговування вимог. Через те для кожної дисципліни обслуговування вимог властивий певний набір основних і допоміжних характеристик якості обслуговування [3].

З економічних причин СРІ можуть проектуватися з дисципліною обслуговування із втратами, за якої вимозі, що надходить у систему в момент відсутності вільних пристроїв обслуговування, відмовляється в обслуговуванні і вона відразу ж покидає її та втрачається. Основною кількісною оцінкою якості обслуговування в цьому випадку є імовірність втрати вимоги P_B (B – blocking, втрати). Ймовірність P_B на відрізку часу (t_1,t_2)визначається як відношення кількості втрачених за цей відрізок часу вимог $C_B(t_1,t_2)$до загальної кількості вимог, що надійшли за той самий час до системи $C(t_1,t_2)$:

$$P_{\hat{A}} = \frac{C_{\hat{A}}(t_1,t_2)}{C(t_1,t_2)}$$

Допоміжними характеристиками QoS є імовірність втрати навантаження (кількість вимог на умовну одиницю часу) та імовірність втрат за часом, які використовуються рідко. Ймовірність втрати навантаження визначається як відношення інтенсивності втраченого навантаження до вхідного, а ймовірність втрат за часом – це сумарна частка часу з проміжку часу (t_1,t_2), за якої були зайняті всі сервери системи [17].

Середня кількість вимог у системі N характеризує ступінь завантаженості системи і співпадає з середньою кількістю зайнятих серверів, що є інтенсивністю обслуговуваного навантаження Y. Всі інші вимоги, що надходять і не знайшли вільного сервера, втрачаються і до системи не потрапляють.

Середній час перебування вимоги в системі T співпадає з середнім часом обслуговування вимоги $\bar{x}$. Для систем із втратами інтенсивність обслуженого навантаження менша вхідного навантаження на величину ΛP_B, тобто

$$Y = \Lambda(1 - P_B). \tag{2}$$

Величина ΛP_B є інтенсивністю надлишкового навантаження і його потік за своєю структурою суттєво відрізняється від потоку навантаження, що надходить до системи, більш нерівномірним характером [31].

1.4.2. Системи з чергами

Для кількісної оцінки якості обслуговування систем з чергою розраховують такі основні характеристики:

- ймовірність очікування $P_{w>0}$ або середню частку затриманих вимог;
- середню довжину черги Q;
- середню тривалість очікування для затриманих вимог t_q;
- середню тривалість очікування для будь-якої вимоги W.

Ймовірність $P_{w>0}$ на відрізку часу (t_1, t_2) визначається як відношення кількості вимог, що потрапили за цей відрізок часу в чергу $C_Q(t_1, t_2)$ до загальної кількості вимог, що надійшли за той же час до системи $C(t_1, t_2)$:

$$P_{w>0} = \frac{C_Q(t_1, t_2)}{C(t_1, t_2)}.$$

Довжина черги є ключовим параметром якості обслуговування (та показником ефективності функціонування СРІ) і визначається кількістю вимог, що очікують обслуговування. Довжина черги залежить від того, коли і скільки вимог надійшло в систему, скільки часу витрачено на обслуговування вимог, що надійшли і т.д. Оскільки довжина черги є випадковою величиною, то як показник довжини черги використовується її математичне сподівання Q.

Середній час очікування в черзі t_q утворюється за рахунок затримки вимог у черзі. Він залежить від кількості вимог, що знаходяться в даний момент у черзі, часу закінчення обслуговування всіх попередніх вимог і т.д.

Середній час очікування в системі W являє собою середнє значення часу очікування, віднесене до всіх вимог – затриманих і незатриманих. Цей параметр вводиться через те, що не всі вимоги потрапляють до черги, а частина з них за наявності вільних серверів системи обслуговується негайно.

Допоміжними характеристиками QoS є середня кількість вимог у системі N та середній час перебування вимоги в системі T. Вони є допоміжними тому, що їх можна розрахувати із основних характеристик.

Середня кількість вимог у системі N визначає ступінь завантаженості системи і за необмеженої черги складається з середньої кількості вимог, що надходять до системи Λ, та тих, що очікують в черзі Q:

$$N = \Lambda + Q.$$

Середній час перебування вимоги в системі T – це час, проведений однією вимогою в системі й усереднений за всіма вимогами (затриманим і незатриманим). Він складається із середнього часу обслуговування $\bar{x}$ і середнього часу очікування вимог у системі W:

$$T = \bar{x} + W.$$

Для кожної моделі потоку всі характеристики якості обслуговування перебувають у певній функціональній залежності [19].

1.4.3. Комбіновані системи

Система з чергою, в якій введено обмеження на максимальну кількість вимог, що можуть перебувати в черзі, або на максимальний час очікування початку обслуговування є системою з комбінованою дисципліною обслуговування. За обмеженої кількості місць очікування (максимальна довжина черги) у випадку надходження вимоги в момент, коли всі сервери і місця очікування зайняті попередніми вимогами, дана вимога втрачається. За обмеженого часу очікування якщо вимога перебуває в черзі понад допустимий час, то їй відмовляється в обслуговуванні і вона теж втрачається. Тому окрім характеристик QoS чистої системи з чергами розраховуються й такі:

- ймовірність втрати вимоги P_B (через обмежену довжини чергу);
- ймовірність очікування понад припустимий час $P_{w>t}$.

Ймовірність $P_{w>t}$ залежить від дисципліни обслуговування черги і найбільш простим для розрахунку є випадок упорядкованої черги *FIFO*. Цю імовірність ще називають умовними втратами, оскільки вимога, що очікує понад припустимий час t може втратити свою актуальність для користувача [9].

1.4.4. Пріоритетні системи

Сучасні телекомунікаційні системи і мережі характеризуються наявністю пріоритетного обслуговування переданих і оброблюваних даних. Аналітичні методи дослідження пріоритетних дисциплін обслуговування вимог розроблені, в основному, для дисциплін з одним класом пріоритетів, а більшість результатів отримані при різних допущеннях і припущеннях, що обмежують їхнє застосування на практиці. Детальне дослідження пріоритетних систем обслуговування припускає застосування комбінованого підходу до моделювання, що дозволяє, як показує практика, отримати результати, що значно змінюють уявлення про вплив пріоритетів на якість функціонування пріоритетних систем обслуговування [30].

Комбіновані системи з обмеженнями на довжину черги та час очікування є найбільш поширеними в телекомунікаціях. В умовах різнорідного трафіка (мова, відео, дані) цю систему доповнюють механізмом пріоритетів, в якому всі вимоги поділяють на категорії і вимоги більш високої категорії при обслуговуванні мають певні переваги (пріоритети) перед вимогами більш низької категорії. Для кількісної оцінки якості обслуговування систем із пріоритетами розраховуються такі ж характеристики, як і для системи з чергами, але для кожного з введених пріоритетів окремо. Наприклад, середній час очікування в черзі вимоги k-го пріоритету, середня кількість вимог у системі k-го пріоритету тощо.

Прикладом системи з втратами може бути телефонна станція – абонент, що є ініціатором телефонного виклику, отримує відмову в обслуговуванні, якщо необхідний канал зв'язку вже зайнятий (технологія комутації каналів). Модель системи з чергою, комбінованої системи та з пріоритетами застосовувана для мереж, що базуються на технології комутації пакетів.

1.4.5. Пропускна здатність і продуктивність

У будь-якій з наведених СРІ якість обслуговування істотно впливає на такі характеристики системи, як пропускна здатність і продуктивність. При цьому критеріями якості обслуговування для систем із втратами є імовірність втрати вимоги, а для систем з чергами – ймовірність очікування. Чим більше припустима норма втрат, тим менша якість обслуговування [18].

Пропускна здатність – це максимальна інтенсивність навантаження, що може надходити до системи при забезпеченні заданої якості обслуговування.

За малої імовірності втрат інтенсивність обслуженого навантаження близька до інтенсивності вхідного навантаження (пропускною здатністю часто вважають інтенсивність обслуженого навантаження). Однак при великих втратах ця відмінність суттєва і тому, чим менша якість обслуговування, тим більшою буде пропускна здатність (і навпаки). Крім того, чим вище норма якості обслуговування, тим більше серверів необхідно для забезпечення певної пропускної здатності.

Пропускна здатність системи не дорівнює кількості серверів в ній, оскільки інтенсивність обслуженого навантаження – це середня кількість зайнятих серверів. Через випадковість потоків навантаження ця кількість не досягає кількості каналів в системі. Задача лише полягає в наближенні середньої кількості зайнятих серверів до кількості серверів системи.

Продуктивність – це гранична, статистично усереднена кількість вимог, які будуть обслужені системою за одиницю часу при заданій якості обслуговування. Ця характеристика використовується, як правило, для оцінки систем управління та керуючих пристроїв.

Пропускна здатність та продуктивність СРІ залежать не тільки від імовірності втрат, але й від структури системи (кількості серверів і схеми їх включення) та дисципліни обслуговування і закону розподілу тривалості обслуговування. Крім того на пропускну здатність та якість обслуговування суттєво впливає вид потоку вимог або його математична модель.

При вимірюванні пропускної здатності системи розподілу інформації важливим є поняття блокування вимоги, що позначає подію, яка складається у відсутності вільних і доступних шляхів з'єднання до необхідного напрямку в момент надходження вимоги (або, точніше, у момент спроби встановити з'єднання). У системі з втратами блоковані вимоги втрачаються і основним показником пропускної здатності є частка втрачених вимог.

Для системи з чергою розрізняють випадок необмеженої та обмеженої черги (кількість місць очікування). Якщо нагромаджувач черги має необмежену ємність, то основними характеристиками є середній час очікування та імовірність очікування (імовірність блокування). У випадку нагромаджувача з обмеженою ємністю додається ще одна характеристика – імовірність втрати вимоги, тоді імовірність блокування дорівнює сумі двох ймовірностей – очікування та втрати вимоги.

У системі з повторними спробами підраховується кількість повторних спроб на одну вимогу та інші характеристики [29].

РОЗДІЛ 2. МОДЕЛЮВАННЯ СИСТЕМ МАСОВОГО ОБСЛУГОВУВАННЯ

2.1. Модель народження та загибелі

У цьому розділі розглядаються дві моделі обслуговуючих систем: у першій представлені лише надходження клієнтів (модель чистого народження), другий – тільки вихід клієнтів із системи (модель чистої загибелі). Прикладом моделі чистого народження є процес оформлення свідоцтв про народження дітей. Як модель чистої загибелі може бути випадкове вилучення запасів, що зберігаються на складі [4].

Дані моделі будуються на основі експоненціального розподілу, який задає інтервал часу між народженнями чи загибеллю. Побічним продуктом цих побудов є демонстрація тісного зв'язку між експоненціальним розподілом та розподілом Пуассона у тому сенсі, що одне з них автоматично визначає інше.

Нехай $p_0(t)$ – ймовірність відсутності подій надходження клієнтів) за період часу t. За умови, що довжина інтервалу часу T між надходженнями клієнтів описується експоненціальним розподілом з інтенсивністю λ, будем мати

$$p_0(t) = P\{\text{інтервал часу}\, T \geq t\} = 1 - P\{\text{інтервал часу}\, T \leq t\} = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t}.$$

При достатньо малому інтервалі часу $h > 0$ маємо

$$p_0(h) = e^{-\lambda h} = 1 - \lambda h + \frac{(\lambda h)^2}{2!} - \ldots = 1 - \lambda h + O\left(h^2\right).$$

Експоненціальний розподіл базується на припущенні, що на досить малому часовому інтервалі $h > 0$ може наступити не більше однієї події (надходження клієнта). Отже, при $h \to 0$

$$p_1(h) = 1 - p_0(h) \approx \lambda h.$$

Цей результат показує, що можливість надходження клієнта протягом інтервалу h прямо пропорційна h з коефіцієнтом пропорційності, рівним інтенсивності надходжень λ.

Щоб отримати розподіл числа клієнтів, що надійшли протягом деякого інтервалу часу, позначимо через $p_n(t)$, ймовірність надходження n клієнтів протягом часу t. При досить малому $h > 0$ маємо наступне

$$p_{\text{n}}(t+h) \approx p_{\text{n}}(t)(1-\lambda h) + p_{n-1}(t)\lambda h,\ n > 0$$

$$p_0(t+h) \approx p_0(t)(1-\lambda h),\ n = 0$$

З першого рівняння випливає, що надходження n клієнтів протягом часу $t+h$ можливо у двох випадках: якщо є n надходжень протягом часу t і немає надходжень за час h, або існує $n-1$ надходжень за час t і одне час h. Будь-які інші комбінації неможливі внаслідок того, що протягом малого періоду h можлива поява тільки однієї події. Відповідно до умови незалежності подій до правої частини рівняння застосуємо закон множення ймовірностей. У другому рівнянні відсутнє надходжень клієнтів на інтервалі $t+h$ можливе лиш тоді, коли немає надходження клієнтів за час h [26].

Перегруповуючи члени і переходячи до границі при $h \to 0$, отримуємо наступне.

$$p'_{\text{n}}(t) = \lim_{h \to 0} \frac{p_{\text{n}}(t+h) - p_{\text{n}}(t)}{h} = -\lambda p_{\text{n}}(t) + \lambda p_{n-1}(t),\ n > 0$$

$$p'_0(t) = \lim_{h \to 0} \frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_{\text{n}}(t),\ n = 0$$

де $p'_{\text{n}}(t)$ – похідна по t функції $p_n(t)$.

Розв'язки наведених вище різницево-диференціальних рівнянь мають наступний вид:

$$p_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!},\ n = 0,1,2,...$$

В даному випадку ми отримали дискретну густину ймовірності розподілу Пуассона з математичним очікуванням $M\{n|t\} = \lambda t$ надходжень за час t. Дисперсія розподілу Пуассона також дорівнює λt.

Отриманий результат означає, що кожен раз, коли часові інтервали між моментами послідовних надходжень заявок розподілені за екпоненціальним законом з математичним сподіванням $\frac{1}{\lambda}$, кількість надходжень заявок в інтервалі, що дорівнює t одиниць часу, характеризується розподілом Пуассона з математичним сподіванням λt. Правильним є і обернене твердження.

Відповідність між експотенціальним розподілом (з інтенсивністю надходжень λ) та розподілом Пуассона показано в наступній таблиці 2.1.

Таблиця 2.1

	Експоненціальний розподіл	Розподіл Пуассона
Випадкова змінна	Час t між появами подій	Кількість n появ подій протягом заданого періоду часу T
Значення випадкової величини	$t \geq 0$	$n = 0,1,2,\ldots$
Функція густини ймовірності	$f(t) = \lambda e^{-\lambda t},\ t \geq 0$	$p_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!},\ n = 0,1,2,\ldots$
Середнє значення (математичне сподівання)	$\frac{1}{\lambda}$ часових одиниць	λT впродовж часу T
Функція розподілу	$P\{t \leq A\} = 1 - e^{-\lambda A}$	$p_{nsN}(T) = p_0(T) + p_1(T) + \ldots + p_N(T)$
Ймовірність, що не з’явиться ні однієї події протягом часу	$P\{t > A\} = e^{-\lambda A}$	$p_0(A) = e^{-\lambda A}$

2.2. Загальна модель системи масового обслуговування

У цьому підрозділі розглядаються загальні системи масового обслуговування, у яких є як вхідний потік клієнтів, так і вихідний потік обслужених клієнтів. Час між послідовними надходженнями клієнтів та час обслуговування є експоненціально розподіленими випадковими величинами. Ця модель є основою при розгляді спеціалізованих моделей Пуассона, які розглядаються в наступному підрозділі.

При розгляді загальних систем масового обслуговування передбачається, що система функціонує протягом досить великого інтервалу часу, але через який у її роботі настає стаціонарний режим. Цей режим функціонування обслуговуючої системи протиставляється перехідному (або режиму, що не встановився), який переважає в початковий період функціонування системи [5].

У розглянутій в цьому підрозділі загальній моделі системи масового обслуговування передбачається, і інтенсивність надходження клієнтів, і інтенсивність вихідного потоку залежать від стану системи, що означає їхню залежність від числа клієнтів у системі обслуговування. Наприклад, збирач плати за проїзд по автомагістралі в години інтенсивного руху намагаються прискорити збір мита. Або в майстерні з фіксованою кількістю верстатів інтенсивність їх поломки зменшується в міру зростання числа аварійних верстатів, бо лише працюючі верстати можуть виходити з ладу.

Введемо такі позначення.

n – кількість клієнтів у системі обслуговування (у черзі та на обслуговуванні); λ_n – інтенсивність надходження до системи клієнтів за умови, що у системі вже знаходиться п клієнтів;

μ_n – інтенсивність вихідного потоку обслужених клієнтів за умови, що в системі знаходиться n клієнтів;

p_n – ймовірність того, що в системі знаходиться n клієнтів.

У загальній моделі системи масового обслуговування встановлюється функціональна залежність ймовірностей p_n від λ_n і μ_n. Ці ймовірності використовуються потім при визначенні функціональних характеристик системи обслуговування, таких як середня довжина черги, середній час очікування та середній коефіцієнт використання послуг.

Ймовірності p_n визначаються з діаграми інтенсивностей переходів, представленої на рис. 2.2. Обслуговуюча система знаходиться в стані n, якщо в ній є n клієнтів. Як показано в розділі 1.3, ймовірність появи більше одного нового клієнта протягом малого проміжку часу h прямує до нуля при $h \to 0$. Це означає, що за $n > 0$ стан n може бути змінено у двох можливих напрямках: $n-1$

, коли з інтенсивністю μ_n, обслужений клієнт вибуває із системи, $n+1$, коли клієнти надходять з інтенсивністю λ_n. Стан 0 може змінитися лише станом 1, коли має місце надходження клієнта з інтенсивністю λ_0. Зауважимо, що μ_0 не визначено, оскільки клієнти не можуть вибувати з порожньої системи обслуговування.

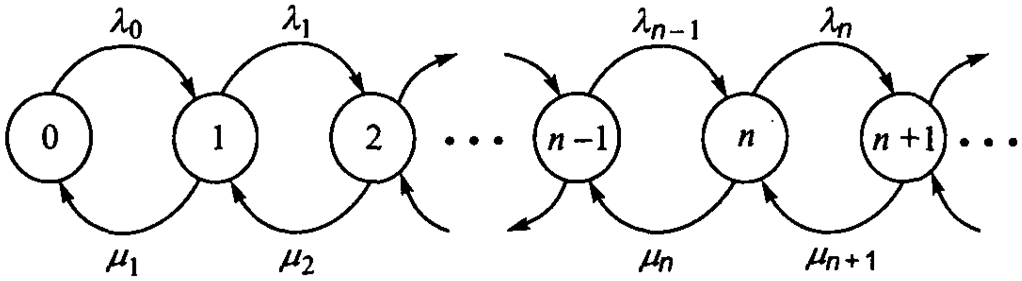

Рис. 2.2. Діаграми інтенсивностей переходів

При виконанні умов стаціонарності очікувані інтенсивності вхідного та вихідного потоків у стані $n(n>0)$ повинні бути рівні. Оскільки стан n може змінюватися лише станам $n-1$ і $n+1$, звідси слідує

$$\begin{pmatrix}\text{очікувана інтенсивність}\\ \text{вхідного потоку в стані } n\end{pmatrix} = \lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1}.$$

Аналогічно

$$\begin{pmatrix}\text{очікувана інтенсивність}\\ \text{вихідного потоку в стані } n\end{pmatrix} = (\lambda_n + \mu_n)p_n.$$

Прирівнюючи дві інтенсивності, отримуємо наступне рівняння балансу.

$$\lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1} = (\lambda_n + \mu_n)p_n,\ n = 1,2,...$$

Як видно із рис. 2.2, рівняння балансу, що відповідає $n=0$, має вигляд

$$\lambda_0 p_0 = \mu_1 p_1.$$

Рівняння балансу розв'язується рекурентно, послідовно виражаючи ймовірності p_1 через p_0 наступним чином: для $n=0$ маємо

$$p_1 = \left(\frac{\lambda_0}{\mu_1}\right)p_0.$$

Для $n=1$ отримуємо

$$\lambda_0 p_0 + \mu_2 p_2 = (\lambda_1 + \mu_1)p_1.$$

Підставляючи сюди $p_1 = \left(\frac{\lambda_0}{\mu_1}\right) p_0$ та спрощуючи отриманий вираз, маємо

$$p_2 = \left(\frac{\lambda_1 \lambda_0}{\mu_2 \mu_1}\right) p_0.$$

Методом індукції можна показати, що

$$p_n = \left(\frac{\lambda_{n-1} \lambda_{n-2} \ldots \lambda_0}{\mu_n \mu_{n-1} \ldots \mu_1}\right) p_0, \; n = 1,2,\ldots$$

Значення p_0 визначається з рівняння $\sum_{n=1}^{\infty} p_n = 1$

2.3. Спеціалізовані системи обслуговування

2.3.1. Спеціалізовані системи обслуговування з пуассонівським розподілом

На рис. 2.3 схематично представлена спеціалізована система обслуговування пуассонівського типу, в якій паралельно функціонують ідентичні сервіси (засоби обслуговування). Клієнт, що чекає, покидає чергу для обслуговування при першому вільному сервісі. Інтенсивність надходження клієнтів у систему дорівнює λ клієнтів за одиницю часу. Всі паралельні сервіси являються ідентичними; це означає, що інтенсивність обслуговування кожного сервісу дорівнює μ клієнтів за одиницю часу. Число клієнтів, що знаходяться в системі обслуговування, включає тих, хто вже обслуговується, та тих, хто перебуває в черзі.

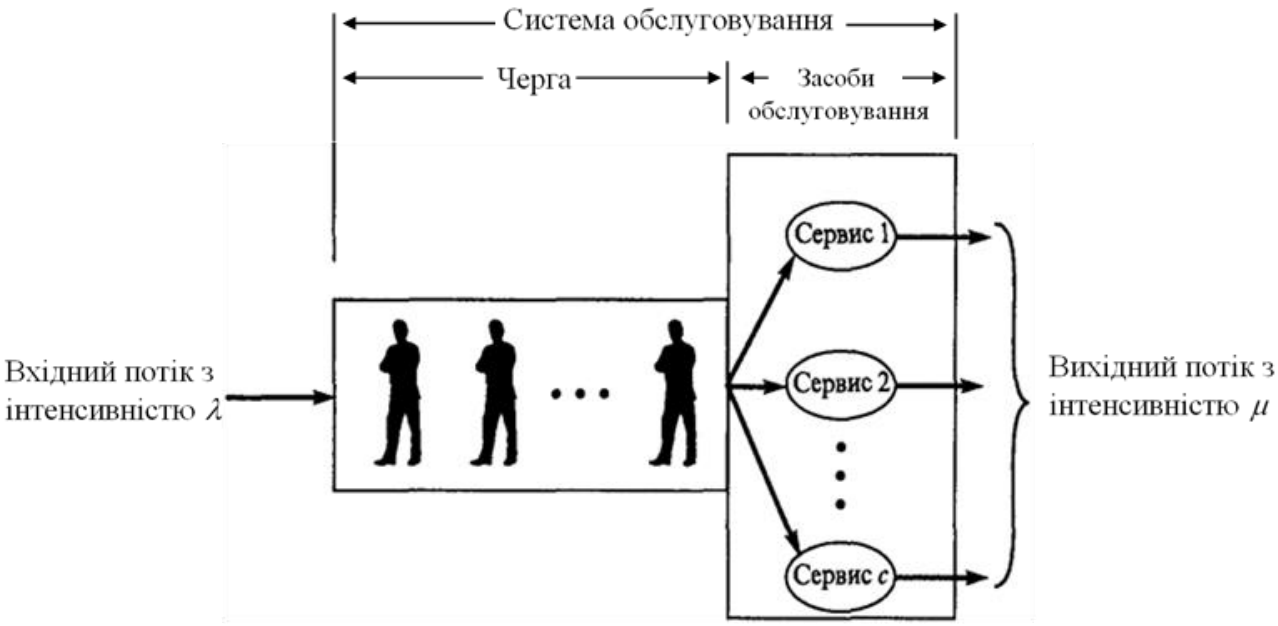

Рис. 2.3. Система обслуговування з кількома сервісами

Позначення, що найбільш підходять для характеристик системи обслуговування (рис. 2.3), мають таку структуру:

$$(a/b/c):(d/e/f),$$

де

a – тип розподілу моментів часу надходження клієнтів у систему;

b – тип розподілу часу між появою елементів вихідного потоку (часу обслуговування);

c – кількість паралельно працюючих сервісів $(=1,2,...,\infty)$;

d – дисципліна черги;

e – максимальна ємність (скінченна або нескінченна) системи (кількість клієнтів у черзі плюс кількість клієнтів, прийнятих на обслуговування);

f – ємність (скінченна або нескінченна) джерела, що генерує клієнтів.

Стандартними позначеннями для типів розподілів вхідного та вихідного потоків (символи a та b) є наступні.

M – марківський (або пуассонівський) розподіл моментів надходження клієнтів у систему або їх виходу з неї (або еквівалентний експоненціальний розподіл інтервалів часу між моментами послідовних надходжень чи тривалостей обслуговування клієнтів);

D – детермінований (фіксований) інтервал часу між моментами послідовних надходжень до системи клієнтів (або детермінована) (Фіксована) тривалість обслуговування клієнтів);

E_k – розподіл Ерланга, або гама-розподіл інтервалів часу (або, що те саме, розподіл суми незалежних випадкових величин, мають експотенціальний розподіл);

GI – довільний (загальний) тип розподілу моментів надходження клієнтів на обслуговування;

G – довільний (загальний) тип розподілу тривалості обслуговування клієнтів.

Для дисцинліни черги (символ d) використовуються такі позначення.

FCFS – першим прийшов - першим обслуговуєшся;

LCFS – останнім прийшов - першим обслуговуєшся;

SIRO – випадковий відбір клієнтів;

GD – довільний (загальний) тип дисципліни.

Для ілюстрації розглянемо структуру системи обслуговування, яка відповідає моделі $(M/D/10):(GD/N/\infty)$. Відповідно до прийнятих позначень тут йдеться про систему (і, відповідно, моделі) масового обслуговування з пуассонівським вхідним потоком (або експотенціальним розподілом) інтервалів часу між моментами послідовних надходжень клієнтів),
фіксованим часом обслуговування та десятьма паралельно функціонуючими сервісами. При цьому дисципліна черги не регламентована, і максимальна кількість клієнтів, що допускаються в систему, дорівнює N. Зрештою, джерело, «що породжує клієнтів», має необмежену ємність [8].

Як історичну довідку зауважимо, що перші три елементи $(a/b/c)$ розглянутого позначення були введені Кендаллом у 1953 році, і в літературі з теорії масового обслуговування вони фігурують як позначення Кендалла. Пізніше в 1966 Лі додав до них символи d та e. Хемді А Тахе в 1968 році був введений останній символ прийнятих позначень – f.

Перед детальним розглядом системи обслуговування пуасонівського типу покажемо, як за допомогою отриманих у підрозділі 2.2 ймовірностей p_n відповідних стаціонарному режиму, можна отримати функціональні характеристики системи.

2.3.2. Функціональні характеристики стаціонарних систем обслуговування

Основними функціональними характеристиками систем масового обслуговування є наступні:

L_s – середня кількість знаходяться в системі клієнтів,

L_q – середня кількість клієнтів в черзі,

W_s – середня тривалість перебування клієнта в системі,

W_q – середня тривалість перебування клієнта в черзі,

$\bar{c}$ – середня кількість зайнятих коштів обслуговування (сервісів).

Нагадаємо, що система включає як чергу, так і засоби обслуговування.

Покажемо, як перераховані функціональні характеристики виходять з ймовірностей p_n – ймовірностей того, що в системі знаходиться n клієнтів. Зокрема, маємо наступне.

$$L_s = \sum_{n=1}^{\infty} n p_n,$$

$$L_q = \sum_{n=c+1}^{\infty} (n-c) p_n.$$

Залежність між L_s і W_s (а також між L_q і W_q) відома в літературі з теорії масового обслуговування як формула Літтла, має вигляд

$$L_s = \lambda_{еф} W_s,$$

$$L_q = \lambda_{еф} W_q.$$

Ці співвідношення справедливі при досить загальних умовах. Параметр $\lambda_{еф}$ являє собою ефективну інтенсивність надходження клієнтів в систему обслуговування. Він дорівнює (вихідній) інтенсивності надходження клієнтів λ, коли всі клієнти мають можливість потрапити в систему обслуговування. Якщо ж деякі клієнти не мають такої можливості з тієї причини, що вона заповнена (наприклад, заповнена автостоянка), то $\lambda_{еф} < \lambda$. Пізніше ми покажемо, як обчислюється $\lambda_{еф}$.

Існує також пряма залежність між величинами W_s і W_q. За визначенням

$$\begin{pmatrix}\text{Середня ттривалість} \\ \text{перебування в системі}\end{pmatrix} = \begin{pmatrix}\text{Середній час} \\ \text{перебування в черзі}\end{pmatrix} + \begin{pmatrix}\text{Середній час} \\ \text{обслуговування}\end{pmatrix}.$$

$$W_s = W_q + \frac{1}{\mu}.$$

Звідси отримаємо формулу, що зв'язує L_s і L_q. Помноживши обидві частини останнього співвідношення на $\lambda_{еф}$ і використовуючи формулу Літтла, в результаті отримуємо

$$L_s = L_q + \frac{\lambda_{еф}}{\mu}.$$

За визначенням різниця між середнім числом клієнтів, що знаходяться в системі L_s, і середнім числом клієнтів в черзі L_q дорівнює середній кількості зайнятих вузлів обслуговування $\bar{c}$. Отже, маємо

$$\bar{c} = L_s - L_q = \frac{\lambda_{еф}}{\mu}.$$

Тому коефіцієнт використання вузлів обслуговування обчислюється як відношення $(c/\bar{c}) \times 100$.

2.3.3. Модель з одним сервісом

У цьому пункті представлені дві моделі обслуговуючої системи з одним засобом обслуговування (тобто $c = 1$). Передбачається, що клієнти надходять з постійною інтенсивністю λ. Інтенсивність обслуговування також є сталою і дорівнює μ клієнтів в одиницю часу. Перша модель не встановлює обмежень на місткість системи, в другій моделі передбачається, що вмістимість системи є обмеженою. У цих двох моделях джерело, що «породжує» клієнтів, має необмежену ємність.

Використовуючи позначення загальної моделі, маємо

$$\lambda_n = \lambda \text{ і } \mu_n = \mu \text{ для всіх } n = 0,1,2,\ldots \,.$$

Оскільки відсутні обмеження на ємність черги, і, отже, всі клієнти, які прибувають можуть потрапити в систему обслуговування, то $\lambda_{еф} = \lambda$, $\lambda_{втрат} = 0$.

Позначимо $\rho = \dfrac{\lambda}{\mu}$ трафік інтенсивності. Тоді вираз для ймовірності p_n в загальній моделі приймає наступний вигляд:

$$p_n = \rho^n p_0, \; n = 0,1,2,\ldots \,.$$

Для визначення величини p_0 використовується тотожність

$$p_0\left(1 + \rho + \rho^2 + \ldots\right) = 1.$$

Припускаємо, що $\rho < 1$, Тоді геометричний ряд має скінченну суму $\dfrac{1}{(1-\rho)}$, тому $p_0 = 1 - \rho$.

Отже, загальна формула для p_n має вигляд

$$p_n = (1-\rho)\rho^n, \; n = 0,1,2,\ldots \; (\rho < 1). \tag{2.1}$$

Ці значення ймовірностей p_n (включаючи ймовірність p_0) відповідають геометричному розподілу.

При виведенні формули для p_n передбачалося, що $\rho < 1$. Це означає, що для досягнення системою стаціонарного режиму функціонування необхідно, щоб інтенсивність надходження клієнтів λ була строго меншою за інтенсивність обслуговування μ. Якщо $\lambda \geq \mu$ геометричний ряд є розбіжним, і, отже, ймовірність p_n стаціонарного стану немає. В цьому випадку система обслуговування буде функціонувати у нестаціонарному режимі, коли довжина черги згодом необмежено зростає.

Середнє число клієнтів, що знаходяться в системі клієнтів L_s, як функціональна характеристика системи обслуговування обчислюється за такою формулою:

$$L_s = \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = (1-\rho)\rho \frac{d}{d\rho}\sum_{n=0}^{\infty}\rho^n = (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) = \frac{\rho}{1-\rho}.$$

Так як в даній моделі $\lambda_{еф} = \lambda$, То інші функціональні характеристики системи обслуговування, обчислюються з використанням співвідношень з підрозділі 2.3.2, що приводить до наступних результатів:

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)},$$

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)},$$

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho},$$

$$\bar{c} = L_s - L_q = \rho.$$

Виведення формули (2.1) для ймовірностей p_n загальної моделі системи обслуговування, абсолютно не залежить від дисципліни черги. Це означає, що математичні сподівання всіх функціональних параметрів W_s, W_q, L_s і L_q справедливі для системи з будь-якою дисципліною черги.

На відміну від середнього часу очікування у системі обслуговування, густина ймовірності його розподілу залежить від дисципліни черги. Проілюструємо це твердження шляхом побудови густини ймовірності часу очікування для моделі $(M/M/1)$ з дисципліною черги $FCFS$ (першим прийшов – першим обслуговуєшся).

Позначимо через τ кількість часу, який щойно прибулий клієнт проведе в системі від моменту прибуття до завершення обслуговування. Виходячи з дисципліни черги, першим прийшов – першим обслуговуєшся, якщо в системі вже знаходиться n клієнтів, які надійшли в систему перед тільки що прибулим, то

$$\tau = t_1' + t_2 + ... + t_{n+1},$$

де t_1' – час, необхідний для закінчення обслуговування клієнта, який уже знаходиться у засобі обслуговування системи, $t_2, t_3, ..., t_n$ – інтервали часу, які потрібні будуть для обслуговування $n-1$ клієнтів, які знаходяться в черзі.

Величина t_{n+1} являє собою час обслуговування клієнта, який щойно надійшов. Позначимо через $w(\tau(\|n+1))$ умовну густини ймовірності τ, де умовою служить наявність в обслуговуваній системі n клієнтів на момент прибуття нового. Оскільки час обслуговування в системі розподілено за експоненціальним законом, який має властивість відсутності післядії, величина t_1' також розподілена за експоненціальним законом. Отже, τ є сумою $n+1$ незалежних випадкових величин, кожна з яких підпорядковується тому самому експоненціальному розподілу. Як відомо з теорії ймовірностей, у цьому випадку функція $w(\tau(\|n+1))$ буде густиною ймовірності гамма-розподілу з параметрами μ і $n+1$. Звідси випливає, що

$$w(\tau)=\sum_{n=0}^{\infty} w(\tau(\|n+1))p_n+\sum_{n=0}^{\infty}\frac{\mu(\mu t)^n e^{-\mu t}}{n!}(1-\rho)\rho^n=$$

$$=(1-\rho)\mu e^{-\mu\tau}\sum_{n=0}^{\infty}\frac{(\lambda\tau)^n}{n!}=\mu(1-\rho)e^{-\mu(1-\rho)\tau},\quad \tau>0.$$

Таким чином, показано, що випадкова величина τ має експотенціальний розподіл з математичним сподіванням

$$W_s=\frac{1}{\mu}(1-\rho). \tag{2.2}$$

Модель $(M/M/1):(GD/N/\infty)$ відрізняється розглянутої вище тільки тим, що система вміщує не більше N клієнтів (максимальна довжина черги дорівнює $N-1$). Прикладами системи обслуговування такого типу служать виробничі ситуації, коли верстат може мати обмежену зону складування заготовок, а також ресторани швидкого харчування з одним пунктом обслуговування клієнтів на автомобілях і т.д.

Ситуація в даній моделі така, що, як тільки число клієнтів в системі досягає N, то жоден з додаткових клієнтів на обслуговування не приймається. З цієї умови випливає, що

$$\lambda_n=\begin{cases}\lambda, & n=0,1,\ldots,N-1,\\ 0, & n=N,N+1,\ldots,\end{cases}$$

$$\mu_n=\mu,\quad n=0,1,\ldots.$$

Використовуючи позначення $\rho = \dfrac{\lambda}{\mu}$, маємо

$$p_n = \begin{cases} \rho^n p_0, & n \le N, \\ 0, & n > N. \end{cases}$$

Значення ймовірності p_0 визначається з рівняння $\sum_{n=0}^{\infty} p_n = 1$, яке приймає наступний вигляд

$$p_0\left(1 + \rho + \rho^2 + \ldots + \rho^N\right) = 1. \tag{2.3}$$

З рівності (2.3) отримуємо

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}}, & \rho \ne 1, \\ \dfrac{1}{N+1}, & \rho = 1. \end{cases}$$

Отже,

$$p_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{N+1}}, & \rho \ne 1, \\ \dfrac{1}{N+1}, & \rho = 1. \end{cases} \quad n = 0, 1, \ldots, N.$$

Зауважимо, що в цій моделі значення параметра $\rho = \dfrac{\lambda}{\mu}$ не обов'язково має бути менше одиниці, так як надходження клієнтів в систему контролюються максимальним обсягом до системи N. Це означає, що в даному випадку в якості інтенсивності надходження клієнтів швидше виступає $\lambda_{еф}$, ніж λ. Так як клієнти будуть втрачені в тому випадку, якщо в системі знаходиться N клієнтів, тоді, як показано на рисунку 2.4,

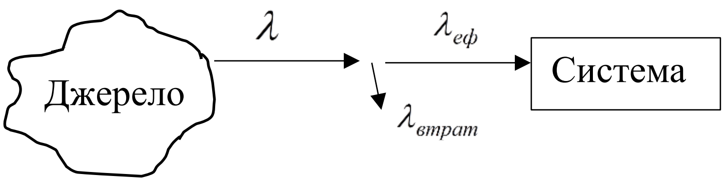

Рис. 2.4. Клієнти із джерела надходять з інтенсивністю λ

$$\lambda_{втрат} = \lambda p_n,$$

$$\lambda_{еф} = \lambda - \lambda_{втрат} = \lambda(1 - p_N).$$

Варто чекати, що $\lambda_{еф}$ буде менше μ.

Середнє число клієнтів в системі обчислюється за формулою

$$L_s = \sum_{\text{n}=0}^{\infty} np_n = \frac{1-\rho}{1-\rho^{\text{N}+1}} \sum_{s\ \text{n}=0}^{\infty} np_n = \frac{(1-\rho)\rho}{1-\rho^{\text{N}+1}} \frac{d}{d\rho}\left(\frac{1-\rho^{\text{N}+1}}{1-\rho}\right) =$$

$$= \frac{\rho\left\{1-(N+1)\rho^{\text{N}} + N\rho^{\text{N}+1}\right\}}{(1-\rho)\left(1-\rho^{\text{N}+1}\right)}, \quad \rho \neq 1.$$

при $\rho = 1$ $L_s = \dfrac{N}{2}$ використовуючи значення L_s і $\lambda_{еф}$, Можна також отримати вирази для W_s , W_q і L_q, як це зроблено в підрозділі 2.3.2.

2.3.4. Моделі з паралельними сервісами

У підрозділі розглядаються три моделі систем масового обслуговування з кількома паралельно працюючими засобами обслуговування (сервісами). Перші дві моделі є узагальнення моделей, розглянутих у пункті 2.3.3, для ситуації кількох паралельно працюючих сервісів. У третій моделі розглядається нескінченна кількість паралельно працюючих сервісів.

Модель $(M/M/c):(GD/\infty/\infty)$ передбачає роботу з паралельних засобів обслуговування. Інтенсивність вхідного потоку клієнтів дорівнює λ, а інтенсивність обслуговування клієнтів – μ для кожного сервісу. Оскільки відсутні обмеження кількість клієнтів у системі, то $\lambda_{еф} = \lambda$.

Результатом використання з паралельних сервісів є пропорційне збільшення інтенсивності обслуговування клієнтів системою до $n\mu$, якщо $n \le c$, і до $c\mu$, якщо $n > c$. Отже, у термінах загальної моделі системи обслуговування λ_n і μ_n визначаються наступним чином.

$$\lambda_n = \lambda, \ n \ge 0,$$

$$\mu_n = \begin{cases} n\mu, \ n \le c, \\ c\mu, \ n > c. \end{cases}$$

Отже,

$$p_n = \begin{cases} \dfrac{\lambda^n}{\mu(2\mu)(3\mu)\ldots(n\mu)} p_0 = \dfrac{\lambda^n}{n!\,\mu^n} p_0, \ n \le c, \\ \dfrac{\lambda^n}{\mu(2\mu)\ldots(c-1)\mu(c\mu)^{n-c+1}} p_0 = \dfrac{\lambda^n}{c!\,c^{n-c}\mu^n} p_0, \ n > c. \end{cases}$$

Значення ймовірності p_0 визначається рівності $\sum\limits_{n=0}^{\infty} p_n = 1$. Якщо $\rho = \lambda/\mu$.

І припустивши, що $\rho/c < 1$, приходимо до наступної формули для p_0:

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\rho}{c} \right)^{n-c} \right\}^{-1} = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \left(\frac{1}{1 - \frac{\rho}{c}} \right) \right\}^{-1}, \ \frac{\rho}{c} < 1.$$

Вираз для λ можна знайти таким чином.

$$L_q = \sum_{n=c}^{\infty} (n-c) p_n = \sum_{k=0}^{\infty} k p_{k+c} = \sum_{k=0}^{\infty} k \frac{\rho^{k+c}}{c^k c!} p_0 = \frac{\rho^c \rho}{c!\,c} \sum_{k=0}^{\infty} k \left(\frac{\rho}{c} \right)^{k-1} p_0 = \frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2} p_0.$$

Оскільки $\lambda_{еф} = \lambda$, то $L_s = L_q + \rho$ значення для W_s та W_q можна знайти, розділивши на λ значення L_s і L_q.

Модель $(M/M/c):(GD/N/\infty)$**,** $c \le N$ відрізняється від моделі $(M/M/c):(GD/\infty/\infty)$ тим, що ємність системи обмежена зверху значенням N (тоді максимальна довжина черги дорівнює $N-c$). Інтенсивності надходження та обслуговування клієнтів дорівнюють λ і μ відповідно. Ефективна інтенсивність надходження заявок у систему обслуговування $\lambda_{еф}$ менше λ в силу обмеження ємності системи значенням N.

Параметри λ_n і μ_n загальної моделі системи обслуговування в даній моделі визначаються наступним чином:

$$\lambda_n = \begin{cases} \lambda, \ 0 \le n \le N, \\ 0, \ n > N, \end{cases}$$

$$\mu_n = \begin{cases} n\mu, \ 0 \le n \le c, \\ c\mu, \ \mathrm{c} \le n \le N. \end{cases}$$

Підставивши λ_n і μ_n в спільний вираз для p_n із підрозділу 2.2 і використавши позначення $\rho = \lambda / \mu$, отримаємо

$$p_n = \begin{cases} \dfrac{\rho^n}{n!} p_0, & 0 \le n \le c, \\ \dfrac{\rho^n}{c! c^{n-c}} p_0, & c \le n \le N, \end{cases}$$

де

$$p_0 = \begin{cases} \left[\sum_{n=0}^{c-1} \dfrac{\rho^n}{n!} + \dfrac{\rho^c \left(1 - \left(\dfrac{\rho}{c} \right)^{N-c+1} \right)}{c! \left(1 - \dfrac{\rho}{c} \right)} \right]^{-1}, & \dfrac{\rho}{c} \ne 1, \\ \left[\sum_{n=0}^{c-1} \dfrac{\rho^n}{n!} + \dfrac{\rho^c}{c!} (N - c + 1) \right]^{-1}, & \dfrac{\rho}{c} = 1. \end{cases}$$

Далі ми обраховуємо L_q для випадку, коли $\dfrac{\rho}{c} \ne 1$:

$$L_q = \sum_{n=c}^{N} (n-c) p_n = \sum_{j=0}^{N-c} j p_{j+c} = \frac{\rho^c \rho}{c! c} \sum_{j=0}^{N-c} j \left(\frac{\rho}{c} \right)^{j-1} p_0 = \frac{\rho^{c+1} d}{c c! d \left(\frac{\rho}{c} \right)} \sum_{j=0}^{N-c} \left(\frac{\rho}{c} \right)^j =$$

$$= \frac{\rho^{c+1}}{(c-1)! (c-\rho)^2} \left\{ 1 - \left(\frac{\rho}{c} \right)^{N-c+1} - (N-c+1) \left(1 - \frac{\rho}{c} \right) \left(\frac{\rho}{c} \right)^{N-c} \right\} p_0.$$

Можна також показати, що для випадку, коли $\dfrac{\rho}{c} = 1$ вираз для L_q має наступний вид:

$$L_q = \frac{\rho^c (N-c)(N-c+1)}{2c!} p_0, \quad \frac{\rho}{c} = 1.$$

Для визначення W_q, W_s і L_s необхідно отримати вираз для $\lambda_{еф}$. Оскільки жоден клієнт не може потрапити в систему після того, як досягнутого ліміт N за її місткістю, то

$$\lambda_{втрат} = \lambda p_N,$$

$$\lambda_{еф} = \lambda - \lambda_{втрат} = (1 - p_N) \lambda.$$

Моделі самообслуговування $(M/M/\infty):(GD/\infty/\infty)$ кількість сервісів є необмеженою, оскільки клієнт виступає одночасно й у ролі сервісу. Типовим прикладом моделі самообслуговування є здавання письмової частини іспиту на право керування автомобілем. Газозаправні станції автомобілів з самообслуговуванням та банківські автомати з 24-годинним режимом роботи не вписуються в розглянуту тут модель, так як обслуговуючі ними пристроями у цих випадках є, по суті, насоси та банківські автомати відповідно.

У даній моделі передбачається, що інтенсивність надходження клієнтів λ є постійною. Інтенсивність обслуговування μ також є постійною. Скориставшись загальною моделлю підрозділу 2.2, маємо

$$\lambda_n = \lambda, \quad n = 0,1,2,...,$$
$$\mu_n = n\mu, \quad n = 0,1,2,....$$

Отже,

$$p_n = \frac{\lambda^n}{n!\mu^n} p_0 = \frac{\rho^n}{n!} p_0, \quad n = 1,2,....$$

Із рівності $\sum_{n=0}^{\infty} p_n = 1$ випливає, що

$$p_0 = \frac{1}{1 + \rho + \frac{\rho^2}{2!} + ...} = \frac{1}{e^{\rho}} = e^{-\rho}.$$

В результаті отримаємо

$$p_n = \frac{e^{-\rho}\rho^n}{n!}, \quad n = 0,1,2,....$$

Ці ймовірності збігаються з ймовірностями розподілу Пуассона з математичним сподіванням $L_s = \rho$. Як і слід очікувати, тут (за принципом самообслуговування) $L_q = W_q = 0$.

2.3.5. Модель $(M/M/R):(GD/K/K)$ при $R < K$

Базовим прикладом для цієї моделі є цех, що нараховує K верстатів. Щоразу, коли верстати виходять з ладу, вдаються до послуг одного з механіків, бригада яких складається з R чоловік. Інтенсивність поломок, віднесена до одного верстату дорівнює λ поломок за одиницю часу. Механік ремонтує верстати, що вийшли з ладу з інтенсивністю μ верстатів за одиницю часу. Передбачається, що моменти часу поломок та час ремонту підпорядковуються розподілу Пуассона [24].

Ця модель відрізняється від усіх розглянутих раніше тим, що потужність джерела, що генерує "клієнтів", скінченна. Наприклад, у моделі цеху джерело може породити скінченну кількість заявок на ремонт. Це стає очевидним, якщо припустити, що всі верстати в цеху зламані, тоді більше не надійде жодна заявка на ремонт. Фактично, лише працюючі верстати можуть зламатися і, отже, генерувати заявки на ремонт.

При заданій інтенсивності λ поломок на один верстат інтенсивність поломок у всьому цеху пропорційна кількості верстатів у робочому стані.

У термінології систем обслуговування наявність n станків у системі означає, що n верстатів зламані. Отже, інтенсивність поломок у всьому цеху обчислюється так:

$$\lambda_n = (K-n)\lambda, \quad 0 \le n \le K.$$

В позначеннях загальної моделі системи обслуговування із підрозділі 2.2 маємо наступне

$$\lambda_n = \begin{cases} (K-n)\lambda, & 0 \le n \le K, \\ 0, & n \ge K, \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 0 \le n \le R, \\ R\mu, & R \le n \le K, \\ 0, & n > K. \end{cases}$$

Тепер із загальної моделі можна отримати наступний вираз:

$$p_n = \begin{cases} \binom{K}{n} \rho^n p_0, & 0 \le n \le R, \\ \binom{K}{n} \dfrac{n!\rho^n}{R!R^{n-R}} p_0, & R \le n \le K, \end{cases}$$

$$p_0 = \left\{ \sum_{n=0}^{R} \binom{K}{n} \rho^n + \sum_{n=R+1}^{K} \binom{K}{n} \frac{n!\rho^n}{R!R^{n-R}} \right\}^{-1}.$$

У цій моделі важко отримати в замкнутій формі вираз для L_s або L_q і, отже, вони повинні обчислюватися відповідно до їх визначення:

$$L_s = \sum_{n=0}^{K} np_n.$$

Значення $\lambda_{еф}$ можна визначити в наступній формі:

$$\lambda_{еф} = M\{\lambda(K-n)\} = \lambda(K - L_s).$$

Використовуючи формули з підрозділу 2.3.2, можна обчислити функціональні показники W_q, W_s і L_q.

2.3.6. Модель $(M/G/1):(GD/\infty/\infty)$. Формула Поллачека-Хінчіна

Аналіз моделей масового обслуговування, в яких вхідні та вихідні потоки не підкоряються пуасонівському розподілу, дуже складний. Взагалі у таких випадках як альтернативний апарат для аналізу моделей обслуговування доцільно використати методи імітаційного моделювання.

У цьому підрозділі розглядається один з небагатьох варіантів системи масового обслуговування, що не підпорядковується пуасонівському розподілу, для якого можуть бути одержані аналітичні результати. Йдеться про той випадок, коли за час обслуговування t має довільний розподіл з математичним сподіванням $M\{t\}$ та дисперсією $D\{t\}$. Для такої моделі відомі аналітичні формули для основних функціональних характеристик обслуговуючої системи, таких як L_s, L_q, W_s і W_q. Однак через аналітичні труднощі неможливо отримати у замкнутій формі вирази для ймовірностей p_n.

Нехай λ – інтенсивність надходження клієнтів у системі обслуговування з одним сервісом. При заданих значеннях $M\{t\}$ та $D\{t\}$ для часу обслуговування та умови $\lambda M\{t\} < 1$ з використанням складного аналізу, пов'язаного із застосуванням апарата ланцюгів Маркова, можна показати, що

$$L_s = \lambda M\{t\} + \frac{\lambda^2\left(M^2\{t\} + D\{t\}\right)}{2\left(1 - \lambda M\{t\}\right)}, \quad \lambda M\{t\} < 1.$$

Ймовірність, що сервіс буде незавантаженим, обчислюється так:

$$p_0 = 1 - \lambda M\{t\} = 1 - \rho.$$

Оскільки $\lambda_{еф} = \lambda$, то інші функціональні характеристики обслуговуючої системи (L_q, W_s і W_q.) можна отримати з формули для L_s як це зроблено у підрозділі 2.3.2 [27].

2.3.7. Модель із вартісними характеристиками

У попередніх пунктах основна увага була зосереджена на пуассонівських системах масового обслуговування. Однак у літературі з теорії масового обслуговування розглядається багато інших моделей. Зокрема, системи масового обслуговування з пріоритетами та системи обслуговування непуассонівського типу становлять істотну частину відповідної наукової літератури.

Рівень обслуговування в системі є функцією інтенсивності обслуговування μ і кількості паралельно працюючих сервісів. Далі розглянемо дві моделі прийняття рішень для визначення «відповідних» рівнів обслуговування для систем масового обслуговування: модель із вартісними характеристиками та модель абсолютним пріорітетом обслуговування. В обох моделях більш високий рівень обслуговування передбачає зменшення часу очікування у системі. У цих моделях для пошуку рівноваги між конфліктуючими факторами (рівнем обслуговування та часом очікування в системі) використовуються функціональні показники обслуговуючої системи, які отримані раніше для різних моделей.

Моделі із вартісними характеристиками прагнуть врівноважити два конфліктуючих вартісних показники:

1) витрати обслуговування;

2) втрати, зумовлені затримками у наданні послуг (час очікування клієнта).

Ці два види витрат конфліктують між собою, тому що збільшення одного з них автоматично веде до зменшення іншого та навпаки.

Нехай рівень обслуговування представляє змінна x, що дорівнює μ або c. Тоді модель зі вартісними характеристиками можна представити у такому вигляді:

$$COC(x) = CTC(x) + CTO(x),$$

де

COC – середня загальна вартість в одиницю часу,

CTC – середня вартість обслуговування в одиницю часу,

CTO – середня вартість очікування в одиницю часу.

Найпростішими функціями CTC і CTO є лінійні функції:

$$CTC(x) = C_1 x,$$

$$CTO(x) = C_2 L_s,$$

де

C_1 – питома вартість на одиницю x за одиницю часу,

C_2 – «ціна» очікування за одиницю часу на клієнта.

2.3.8. Моделі з абсолютним пріорітетом обслуговування

Життєздатність моделі системи обслуговування із вартісними характеристиками залежить від того, наскільки добре ми можемо оцінити параметри вартості. Загалом оцінити ці параметри досить складно, особливо якщо вартість пов'язана з очікуванням клієнта. У моделях з абсолютним пріорітетом обслуговування робиться спроба обійти цю проблему, оперуючи безпосередньо функціональними показниками системи обслуговуваня. Ідея полягає у визначенні прийнятного інтервалу зміни для рівня обслуговування (параметри μ або c) шляхом пошуку розумних меж для конкуруючих економічних показників, що характеризують процес обслуговування. Ці межі представляють собою рівні переважного обслуговування, яких прагне досягти особа, яка приймає управлінське рішення [23].

Проілюструємо застосування цієї процедури для моделі системи обслуговування з декількома сервісами, в якій необхідно визначити «прийнятну» кількість сервісів c. Для цього розглянемо два (конкуруючих) економічних показника процесу обслуговування:

1) середній час очікування у системі W_s.

2) відсоток простою послуг X.

Значення W_s можна обчислити, використовуючи програмне забезпечення TORA для моделі $(M/M/c)$. Відсоток простою засобів обслуговування можна обчислити в такий спосіб.

$$X = \frac{c-\bar{c}}{c}\cdot 100 = \frac{c-(L_s - L_q)}{c}\cdot 100 = \left(1-\frac{\lambda_{еф.}}{c\mu}\right)\cdot 100.$$

Задача зводиться до визначення такої кількості сервісів c, що

$$W_s \le \alpha \text{ і } X \le \beta,$$

де α і β – рівні обслуговування з абсолютним пріорітетом, визначені особою, яка приймає рішення. Наприклад, можна поставити умову, що $\alpha = 3$ і $\beta = 10\%$. Задачу можна розв'язати, побудувавши графіки W_s і X як функції кількості сервісів з (рис. 2.4). Відзначаючи на графіках значення α і β, ми визначаємо прийнятний інтервал зміни для рівня обслуговування c. Якщо обидві згадані вище умови не можна задовольнити одночасно, необхідно послабити один або обидва рівня абсолютного пріорітету доки не буде отримано прийнятний інтервал зміни кількості сервісів [22].

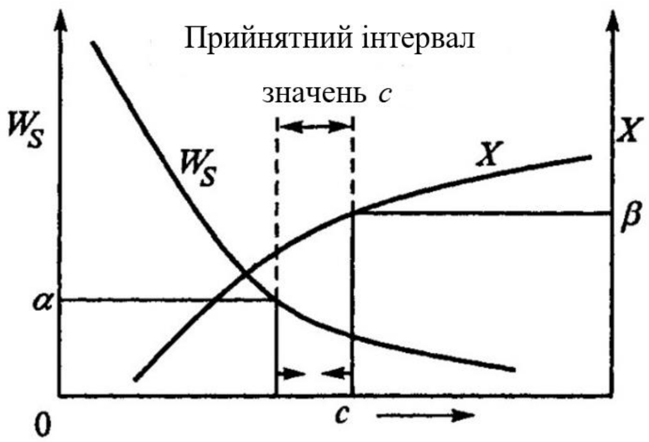

Рис. 2.4. Прийнятний інтервал зміни для рівня обслуговування c

РОЗДІЛ 3. ЗАСТОСУВАННЯМ ТЕОРІЇ МАСОВОГО ОБСЛУГОВУВАННЯ

Приклад 1.

Відвідувачі ресторану швидкого харчування скаржаться на повільне обслуговування. Нині у ресторані працюють три касири. Адміністратор доручив консалтинговій фірмі провести розслідування скарги. В результаті було виявлено наступну залежність між числом касирів та часом очікування обслуговування [2].

Таблиця 3.1

Кількість касирів	1	2	3	4	5	6	7
Середній час очікування (хвилини)	16,2	10,3	6,9	4,8	2,9	1,9	1,3

Наведені дані свідчать про те, що при працюючих в даний час трьох касирах середній час очікування обслуговування приблизно дорівнює 7 хв. Адміністратор хоче зменшити його приблизно до трьох хвилин. З цих же даних випливає, середній час очікування стає менше 3 хв., якщо число касирів більше або дорівнює п'яти.

Результати дослідження системи обслуговування також можна використати для оптимізації моделі із вартісними характеристиками, в якій мінімізуються: сума витрат, пов'язаних з наданням послуг, та втрат, зумовлених затримками у їх наданні. На рис. 1.1 зображено типову вартісну модель системи обслуговування (у грошових одиницях за одиницю часу), де витрати обслуговування зростають зі зростанням його рівня.

У той же час втрати, зумовлені затримками у наданні послуг, зменшуються зі зростанням рівня обслуговування. Головною проблемою, пов'язаною із застосуванням вартісних моделей, є труднощі оцінки втрат в одиницю часу, зумовлених затримками у наданні послуг. Зокрема, це особливо відчутно, коли послуги надаються особі, чия поведінка може не співпадати з інтересами функціонування системи обслуговування.

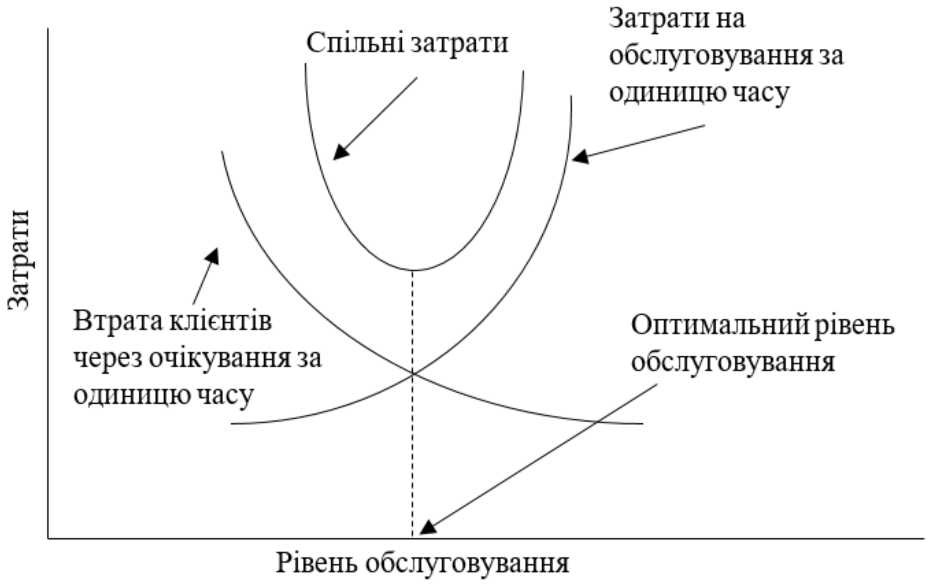

Рис. 3.1. Типова вартість моделей системи обслуговування

Приклад 2.

При обслуговуванні складного агрегату завжди існує запасний блок для негайної заміни у разі поломки. Час виходу з ладу агрегату (або його запасного блоку) є випадковою величиною, розподіленою за експотенціальним законом, та в середньому відбувається кожні 40 хвилин. Оператор, який обслуговує агрегат, стверджує, що агрегат «має звичку» виходити з ладу щовечора близько 20:30. Проаналізуємо затвердження оператора.

Середня інтенсивність відмов агрегату дорівнює $\lambda = 60/40 = 1{,}5$ відмови на годину. Отже, густина експоненціального розподілу часу відмови має вигляд

$$f(t) = 1{,}5e^{-1{,}5t},\ t > 0.$$

Що стосується заяви оператора, то і без обчислень видно, що вона не може відповідати дійсності, тому що не узгоджується з тим, що час між відмовами агрегату розподілено за законом і, отже, є випадковим. Для підтвердження або спростування заяви оператора не можна використовувати ймовірність того, що відмова буде відбуватися о 20:30, оскільки ймовірність такої події залежить від часу дня (щодо 20:30), коли ймовірність обчислюється. Наприклад, якщо обчислення виконуються о 20:20, то ймовірність того, що затвердження оператора виявиться справедливим цим вечором, дорівнює

$$P\left\{t < \frac{10}{60}\right\} = 1 - e^{-1,5\left(10/60\right)} = 0,22,$$

тобто є дуже малою. Якщо обчислення виконуються о 19:00, то ймовірність того, що відмова матиме місце о 20:30, зростає приблизно до 0,9.

Ці два крайні значення ймовірності показують, що достовірність твердження оператора не можна проаналізувати на основі отриманих ймовірностей; у цій ситуації ми повинні покладатися тільки на характеристики експоненціального розподілу (точніше, на його властивість відсутності післядії).

Приклад 3.

У невеликому районі кожні 12 хвилин народжується дитина. Час між народженнями розподілено за експотенціальним законом. Потрібно визначити:

1) середня кількість народжень за рік;

2) ймовірність того, що протягом одного дня не буде жодного народження;

3) ймовірність видачі 50 свідоцтв про народження до кінця третьої години, якщо відомо, що протягом останніх двох годин було видано 40 таких свідоцтв.

Обчислимо інтенсивність народження за день: $\lambda = \frac{24 \times 60}{12} = 120$ народжень на день.

Інтенсивність народжень у районі протягом року дорівнює $\lambda t = 120 \times 365 = 43800$ народжень. Ймовірність того, що протягом одного дня не народиться жодна дитина, обчислюється з використанням пуассонівського розподілу

$$p_0(1) = \frac{(120 \times 1)^0 e^{-120 \times 1}}{0!} \approx 0.$$

Для обчислення ймовірності видачі 50 свідоцтв про народження до кінця третього години за умови, що протягом останніх двох годин було видано 40 таких свідчень, зауважимо, що оскільки розподіл числа народжень є пуассонівським, шукана ймовірність зводиться до ймовірності появи 10 $(= 50 - 40)$ народжень за одну (= 3 – 2) годину. Оскільки $\lambda = 60/12 = 5$ народжень за годину, то

$$p_{10}(1)=\frac{(5\times 1)^{10}e^{-5\times 1}}{10!}\approx 0{,}01813.$$

Формули для обчислення функціональних параметрів систем обслуговування включають громіздкі обчислення, тому їх виконання бажано використовувати програмне забезпечення TORA. На рис. 3.2 показані вихідні дані, отримані за допомогою програми TORA, для моделі чистого народження з інтенсивністю $\lambda t = 5\times 1 = 5$ народжень на день.

Problem title: Example 17.4-1
Scenario 1 — Pure Birth Model

Poisson with Lambda*t = 5.00000

Values of p(n) for n=0 to 17, else p(n) < .00001

0 0.00674	1 0.03369	2 0.08422	3 0.14037	4 0.17547
5 0.17547	6 0.14622	7 0.10444	8 0.06528	9 0.03627
10 0.01813	11 0.00824	12 0.00343	13 0.00132	14 0.00047
15 0.00016	16 0.00005	17 0.00001		

Cumulative values of p(n) for n=0 to 17

0 0.00674	1 0.04043	2 0.12465	3 0.26509	4 0.44049
5 0.61596	6 0.76218	7 0.86663	8 0.93191	9 0.96817
10 0.98630	11 0.99455	12 0.99798	13 0.99930	14 0.99977
15 0.99993	16 0.99998	17 0.99999		

Рис. 3.2. Вихідні дані програми TORA

У моделі абсолютної загибелі передбачається, що система починає функціонувати, коли в момент часу 0 у ній є N клієнтів і не допускається жодного нового надходження клієнта. Обслужені клієнти вибувають із системи з інтенсивністю μ клієнтів за одиницю часу. Нехай $p_n(t)$ – ймовірність того, що після t часових одиниць у системі залишається n клієнтів. Для отримання різницево-диференціальних рівнянь щодо $p_n(t)$ зазвичай слідують логіці міркувань, використаних у моделі чистого народження. Тому маємо

$$p_N(t+h)=p_N(t)(1-\mu h)$$

$$p_n(t+h)=p_n(t)(1-\mu h)+p_{n+1}(t)\mu h,\ 0<n\le N,$$

$$p_0(t+h)=p_0(t)(1)p_1(t)\mu h.$$

При $h\to 0$ отримаємо

$$p'_N(t)=-\mu p_N(t),$$

$$p'_n(t)=-\mu p_n(t)+\mu p_{n+1}(t),\ 0<n<N,$$

$$p_0'(t) = \mu p_1(t).$$

Ці рівняння мають розв'язок

$$p_n(t) = \frac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)!},\ n = 1,2,...,N,$$

$$p_0(t) = 1 - \sum_{n=1}^{N} p_n(t),$$

який називається усіченим розподілом Пуассона.

Приклад 4.

Секція квітів бакалійно-гастрономічного магазину складає 18 дюжин троянд спочатку кожного тижня. У середньому продається 3 дюжини троянд на день (за один раз продається дюжина троянд), але попит підпорядковується розподілу Пуассона. Як тільки рівень запасу знижується до 5 дюжин, робиться нове замовлення на постачання 18 дюжин троянд на початку наступного тижня. Запаси за своєю природою такі, що всі невикористані до кінця тижня троянди приходять у непридатність та ліквідуються [13]. Потрібно обчислити

1) ймовірність розміщення замовлення до кінця кожного дня тижня.

2) середню кількість троянд, які будуть ліквідовані до кінця тижня.

Так як троянди купуються з інтенсивністю $\mu = 3$ дюжини на день, то ймовірність того, що замовлення буде розміщено наприкінці дня t, дорівнює

$$p_{n\le 5}(t) = p_0(t) + p_1(T) + ... + p_5(t) = p_0(t) + \sum_{n=1}^{5} \frac{(3t)^{18-n} e^{-3t}}{(18-n)!},\ t = 1,2,...,7.$$

За допомогою програми TORA отримаємо наступні результати розрахунків.

Таблиця 3.2

t (дні)	1	2	3	4	5	6	7
μt	3	6	9	12	15	18	21
$p_{n\le 5}(t)$	0,0000	0,0088	0,1242	0,4240	0,7324	0,9083	0,9755

Середня кількість троянд, які будуть ліквідовані до кінця тижня $t = 7$, обчислюється (з використанням програми TORA) наступним чином.

$$M\{n|t=7\}=\sum_{n=0}^{18} np_n(7)=0{,}664 \text{ дюжини.}$$

Приклад 5.

Бакалійний магазин працює із трьома касами. Вивіска біля кас сповіщає покупців, що будь-якої миті буде відкрито додаткову каса, як тільки кількість покупців у будь-якій черзі перевищить 3. Це означає, що якщо кількість покупців менше чотирьох, то працюватиме лише одна каса. Якщо кількість покупців від чотирьох до шести, то працюватиме дві каси. Якщо є більше шістьох покупців, будуть відкриті всі три каси. Покупці підходять до кас відповідно до розподілу Пуассона з математичним сподіванням 10 осіб на годину. Час обслуговування одного покупця в касі розподілено за експоненціальним законом із середнім 12 хвилин. Визначимо в режимі ймовірність p_n що n покупців стоять у черзі в касу [7].

З формулювання завдання маємо таке.

$$\lambda_n=\lambda=10 \text{ покупців за годину, } n=0,1,\ldots,$$

$$\mu_n=\begin{cases}\dfrac{60}{12}=5 \text{ покупцівв час}, n=0,1,2,3,\\ 2\times 5=10 \text{ покупцівв час}, n=4,5,6,\\ 3\times 5=15 \text{ покупцівв час}, n=7,8,\ldots\end{cases}$$

Звідси

$$p_1=\left(\frac{10}{5}\right)p_0=2p_0,$$

$$p_2=\left(\frac{10}{5}\right)^2 p_0=4p_0,$$

$$p_3=\left(\frac{10}{5}\right)^3 p_0=8p_0,$$

$$p_4=\left(\frac{10}{5}\right)^3\left(\frac{10}{5}\right)p_0=8p_0,$$

$$p_5 = \left(\frac{10}{5}\right)^3 \left(\frac{10}{5}\right)^2 p_0 = 8p_0,$$

$$p_6 = \left(\frac{10}{5}\right)^3 \left(\frac{10}{5}\right)^3 p_0 = 8p_0,$$

$$p_n = \left(\frac{10}{5}\right)^3 \left(\frac{10}{5}\right)^3 \left(\frac{10}{5}\right)^{n-6} p_0 = 8\left(\frac{2}{3}\right)^{n-6} p_0,\ n = 7,8,...$$

Значить p_0 знаходимо з рівняння

$$p_0 + p_0\left[2 + 4 + 8 + 8 + 8 + 8 + 8\left(\frac{2}{3}\right) + 8\left(\frac{2}{3}\right)^2 + 8\left(\frac{2}{3}\right)^3 + ...\right] = 1$$

або, що рівносильно

$$p_0\left\{31 + 8 + \left[1 + \left(\frac{2}{3}\right) + \left(\frac{2}{3}\right)^2 + ...\right]\right\} = 1.$$

Використаємо формулу суми нескінченної геометричної прогресії

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x},\ |x| < 1,$$

отримаємо наступне

$$p_0\left\{31 + 8 + \left[\frac{1}{1-\frac{2}{3}}\right]\right\} = 1.$$

Отже, $p_0 = \dfrac{1}{35}$.

Знаючи p_0 можна визначити будь-яку ймовірність, що має відношення до завдання. Наприклад, ймовірність того, що працюватиме лише одна каса, обчислюється як ймовірність перебування у системі не більше трьох клієнтів, тобто.

$$p_0 + p_1 + p_2 + p_3 = (1 + 2 + 4 + 8)\left(\frac{1}{35}\right) \approx 0{,}273.$$

Тут передбачається, що в бакалійному магазині буде відкрито щонайменше одну каса, навіть у тому випадку, коли зовсім немає покупців.

Ймовірності p_n можна використовувати визначення чисельних значень функціональних характеристик системи, що розглядається. Наприклад, середня кількість непрацюючих кас дорівнює

$$3p_0 + 2(p_1 + p_2 + p_3) + 1(p_4 + p_5 + p_6) + 0(p_7 + p_8 ...) = 1 \text{ каса.}$$

Приклад 6.

Парковка для відвідувачів коледжу має всього п'ять місць. Автомобілі прибувають на стоянку відповідно до розподілу Пуассона з інтенсивністю шість автомобілів на годину. Час перебування автомобілів на стоянці є експоненціально розподіленою випадковою величиною із середнім часом 30 хв. Відвідувачі, які не можуть знайти вільного місця на стоянці безпосередньо після прибуття, можуть тимчасово очікувати звільнення місця на території стоянки. Таких місць для очікування на стоянці є три. Якщо і стоянка, і всі місця для очікування заповнені, то прибулі автомобілі змушені шукати іншу автостоянку [14]. Потрібно визначити наступне:

a) ймовірність того, що в системі знаходиться автомобілів;

b) ефективну інтенсивність надходження автомобілів на стоянку;

c) середня кількість автомобілів на стоянці;

d) середній час перебування автомобіля в черзі на території стоянки;

e) середня кількість зайнятих місць на автостоянці.

Перш за все, зауважимо, що місце для стоянки в ситуації, що розглядається виступає в ролі сервісу, так що система має всього $c = 5$ засобів обслуговування. Максимальна місткість системи дорівнює 5 + 3 = 8 автомобілів.

Ймовірність може бути визначена як окремий випадок із загальної моделі, розглянутої в підрозділі 2.1. Наприклад, маємо

$$\lambda_n = 6 \text{ автомобілів за годину}, n = 0,1,2,...,8,$$

$$\mu_n = \begin{cases} n\left(\dfrac{60}{30}\right) = 2n \text{ автомобілів за годину}, n = 0,1,2,...,5, \\ 5\left(\dfrac{60}{30}\right) = 10 \text{ автомобілів за годину}, n = 6,7,8. \end{cases}$$

Отже, з співвідношень, отриманих в підрозділі 2.2, отримаємо

$$p_n = \begin{cases} \dfrac{\left(\frac{6}{2}\right)^n}{n!} p_0, n = 0,1,2,...,5, \\ \dfrac{\left(\frac{6}{2}\right)^n}{5!5^{5-n}} p_0, n = 6,7,8. \end{cases}$$

Значення p_0 обчислимо шляхом підстановки значень для p_n, $n = 0,1,2,...,8,$ в рівняння $p_0 + p_1 + ... + p_8 = 1$. В результаті отримаємо

$$p_0 + p_0\left(\frac{3}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!} + \frac{3^6}{5!5^1} + \frac{3^7}{5!5^2} + \frac{3^8}{5!5^3}\right) = 1$$

Розв'язком цього рівняння є $p_0 = 0{,}04812$. Знайдене значення p_0 дозволяє обчислити всі ймовірності від p_1 до p_8:

Таблиця 3.3

n	1	2	3	4	5	6	7	8
p_n	0,14436	0,21654	0,21654	0,16240	0,09744	0,05847	0,03508	0,02105

Ефективну інтенсивність надходження автомобілів на стоянку $\lambda_{еф}$, можна обчислити з використанням схеми (рис. 3.3) відповідно до якої клієнти із джерела надходять з інтенсивністю λ. Автомобіль може надійти на стоянку з інтенсивністю $\lambda_{еф}$ або виїхати в пошуках іншої стоянки з інтенсивністю $\lambda_{втрат}$, тобто $\lambda = \lambda_{еф} + \lambda_{втрат}$. Автомобіль не може в'їхати на стоянку, якщо там вже є 8 автомобілів. Це означає, що частина автомобілів, які не зможуть потрапити на стоянку, пропорційна p_8.

Отже,

$$\lambda_{втрат} = \lambda p_8 = 6 \times 0{,}02105 = 0{,}1263 \text{ автомобілі в за годину,}$$

$$\lambda_{еф} = \lambda - \lambda_{втрат} = 6 - 0{,}1263 = 5{,}8737 \text{ автомобілі в за годину.}$$

Середня кількість автомобілів на стоянці (тих, які займають місця стоянки, і тих, які очікують місця) визначається значенням L_s – середнім числом клієнтів в системі. Значення L_s отримується через p_n наступним чином.

$$L_s = 0p_0 + 1p_1 + ... + 8p_8 = 3{,}1286 \text{ автомобілів.}$$

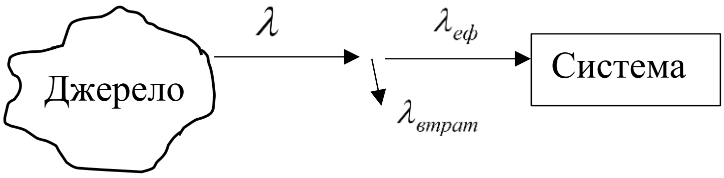

Рис. 3.3. Клієнти із джерела надходять з інтенсивністю λ

Автомобіль, який чекає, поки звільниться місце для стоянки, фактично знаходиться в черзі. Отже, час його очікування дорівнює величині W_q . Для обчислення W_q використовуємо визначення

$$W_q = W_s - \frac{1}{\mu}.$$

Так як

$$W_s = \frac{L_s}{\lambda_{еф}} = \frac{3{,}1286}{5{,}8737} = 0{,}53265 \text{ години,}$$

то

$$W_q = 0{,}53265 - \frac{1}{2} = 0{,}03265 \text{ години.}$$

Середня кількість зайнятих місць на автостоянці дорівнює середньому значенню «зайнятих сервісів» і тому обчислюється таким чином

$$\bar{c} = L_s - L_q = \frac{\lambda_{еф}}{\mu} = \frac{5{,}8737}{2} = 2{,}9368 \text{ місць.}$$

Звідси отримуємо, що коефіцієнт використання місць на стоянці дорівнює $\frac{\bar{c}}{c} = \frac{2{,}9368}{5} = 0{,}58736.$ [15].

Приклад 7.

Автоматична мийка для автомобілів має тільки один мийний бокс. Автомобілі прибувають відповідно до розподілу Пуассона в середньому 4 машини за годину і можуть очікувати обслуговування на стоянці поруч з автомийкою. Час мийки автомобіля є експоненціально розподіленою випадковою величиною з математичним сподіванням 10 хв. Автомобілі, які не поміщаються на стоянці, можуть очікувати на прилеглій до автомийці вулиці. Це означає, що практично немає обмежень на ємність системи обслуговування. Господар автомийки хоче визначити кількість місць на стоянці для автомобілів.

Для даної задачі маємо $\lambda = 4$ автомобіля на годину і $\mu = \frac{60}{10} = 6,$ автомобілів на годину. Так як $\rho = \frac{\lambda}{\mu} < 1,$ то система може функціонувати в стаціонарному режимі [16].

Результати використання програми TORA для розв'язання цієї задачі, представлені на рис. 3.4 отримані шляхом введення даних в наступному порядку: $\lambda = 4$, $\mu = 6$, $c = 1$, ємність системи дорівнює ∞ і ємність джерела також дорівнює ∞.

З цих даних отримуємо, що середня кількість автомашин, які очікують у черзі, дорівнює $L_q = 1{,}33$ автомашини. Ми не можемо розглядати L_q в якості єдиного аргументу при визначенні шуканого кількості місць на стоянці, бо при розрахунку повинна враховуватися максимально можлива довжина черги. Наприклад, можна розрахувати кількість місць на стоянці, при якій, щонайменше, 90% прибулих автомобілів знайде місце на стоянці.

Нехай невідома змінна s представляє шукану кількість місць на стоянці. Тоді система має ємність $s+1$ (чергу плюс місце на мийці). Автомобіль, що прибуває в 90% випадків отримає місце на стоянці, якщо в системі знаходиться максимум автомобілів. Ця умова еквівалентна наступному ймовірнісному твердженню:

$$p_0 + p_1 + \ldots + p_s \geq 0{,}9$$

З лістингу, показаного на рис. 3.4, слідує, що суми ймовірностей p_n рівні 0,86831 і 0,91221 при $n=4$ і $n=5$ відповідно. Це означає, що умова виконується за $s \geq 5$ місць на стоянці.

```
Problem title: Example 2.3.2.1
Scenario 1 – (M/M/1):(GD/*/*)
-----------------------------------------------------------------
Lambda =    4.00000          Lambda eff =     4.00000
Mu =    6.00000              Rho =    0.66667

Ls =    2.00000              Lq =    1.33333
Ws =    0.50000              Wq =    0.33333
-----------------------------------------------------------------
Values of p(n) for n=0 to 25, else p(n)<.00001

0 0.33333     1 0.22222     2 0.14815     3 0.09877     4 0.06584
5 0.04390     6 0.02926     7 0.01951     8 0.01301     9 0.00857
10 0.00578    11 0.00385    12 0.00257    13 000471     14 0.00114
15 0.00076    16 0.00051    17 0.00034    18 0.00023    19 0.00015
20 0.00010    21 0.00007    22 0.00004    23 090003     24 0.00002
25 0.00001

Cumulative values of p(n) for n=0 to 25

0 0.33333     1 0.55556     2`0.70370     3 0.80247     4 0.86831
5 0.91221     6 0.94147     7 0.96098     8 0.97399     9 0.98266
10 0.98844    11 0.99229    12 0.99486    13 0.99657    14 0.99772
15 0.99848    16 0.99898    17 0.99932    18 0.99955    19 0.99970
20 0.99980    21 0.99987    22 0.99991    23 0.39994    24 0.99996
25 0.99997
```

Рис. 3.4 Лістинг

Кількість місць на стоянці s може бути визначено з використанням формули (2.1). Отримуємо

$$(1-\rho)\left(1+\rho+\rho^2+\ldots+\rho^s\right) \geq 0{,}9.$$

Сума усіченого геометричного ряду дорівнює $\dfrac{1-\rho^{s+1}}{1-\rho} \geq 0{,}9.$ Отже, останній вираз зведемо до вигляду

$$1-\rho^{s+1} \geq 0{,}9. \tag{3.1}$$

Спростивши нерівність (3.1), отримаємо

$$\rho^{s+1} \geq 0{,}1. \tag{3.2}$$

Прологарифмувавши обидві частини нерівності (3.2), отримаємо наступне:

$$s \geq \frac{\ln(0{,}1)}{\ln\left(\frac{4}{6}\right)} - 1 = 4{,}679.$$

Таким чином, необхідно $s \geq 5$ місць на стоянці.

Приклад 8.

У моделі роботи автомийки з прикладу 7 є цілком обґрунтованим припущення, що обслуговування виконується відповідно до дисципліни черги $FCFS$. Визначимо надійність використання величини W_s, як оцінки часу очікування у системі. Для цього оцінимо частину клієнтів, час очікування яких перевищує W_s. Використовуючи формулу (2.2), отримуємо

$$P\{\tau > W_s\} = 1 - \int_0^{W_s} w(\tau)d\tau = e^{-\mu(1-\rho)W_s} = e^{-1} \approx 0{,}368.$$

Отже, за дисципліни черги $FCFS$ для 37 % клієнтів час очікування перевищить значення W_s. Цей відсоток здається надмірно великим, особливо якщо врахувати, що поточне значення W_s, для цієї системи обслуговування також є більшим (0,5 години). Зауважимо, що знайдена ймовірність $\left(= e^{-1} \approx 0{,}368\right)$ не залежить від інтенсивностей λ і μ моделі $(M / M / 1):(FCFS / \infty / \infty)$; це, своєю чергою, означає, що її величину не можна зменшити. Отже, якщо ми проектуємо систему на підставі середньої величини W_s, слід очікувати, що для 36,8 % клієнтів час очікування перевищить середній час очікування у системі.

Існує дві можливості покращити ситуацію:

1) збільшити інтенсивність обслуговування клієнтів μ для зменшення значення W_s, до прийнятного рівня;

2) вибрати інтенсивність обслуговування клієнтів таким чином, щоб ймовірність того, що час очікування перевищить заздалегідь певну величину (скажімо, 10 хвилин), була меншою за прийнятно малу величину (наприклад, 10 %). Перша можливість еквівалентна пошуку такого значення μ, що $W_s < \overline{T}$, а

друга – обчислення такого значення μ, що $P\{\tau > \overline{T}\} < \alpha$, де $\overline{T}$ і α повинні визначатися користувачем.

Приклад 9.

Розглянемо ситуацію з автомийкою автомобілів з прикладу 7 Нехай станція має чотири місця для стоянки автомобілів. Якщо всі місця на стоянці зайняті, автомобілі, що знову прибувають, змушені шукати іншу автомийку. Господар хоче визначити вплив обмеженої кількості місць для стоянки автомобілів на втрату клієнтів [14].

В позначеннях, прийнятих у моделі, максимальна місткість системи дорівнює $N = 4 + 1 = 5$. Вхідними даними для програми TORA є числа 4, 6, 1,5
і ∞ відповідно, а вихідні дані представлені на рис. 3.5.

Оскільки ємність системи дорівнює $N = 5$, частка втрачених клієнтів становить $p_s = 0{,}04812$, що при цілодобовій роботі мийної станції еквівалентно втраті $(\lambda p_5) \times 24 = 4 \times 0{,}04812 \times 24 = 4{,}62$ автомобілів на день. Рішення щодо збільшення кількості місць для стоянки автомобілів має ґрунтуватися на сумі втрат автомийки.

```
Problem title: Example 2.3.2.3
Scenario 1 – (M/M/1):(GD/5/*)
-----------------------------------------------------------------
Lambda =    4.00000          Lambda eff =    3.80752
Mu =    6.00000              Rho =    0.66667

Ls =    1.42256              Lq =    0.78797
Ws =    0.37362              Wq =    0.20695
---------------------------------------------------------------------
Values of p(n) for n=0 to 5, else p(n)<.00001

0 0.36541     1 0.24361     2 0.16241     3 0.10827     4 0.07218
5 0.04812

Cumulative values of p(n) for n=0 to 5

0 0.36541     1 0.60902     2`0.77143     3 0.87970     4 0.95188
5 1.00000
```

Рис. 3.5. Вихідні дані отримані за допомогою програми TORA

Аналізуючи ситуацію з іншого боку, зауважуємо, що середній час перебування клієнта в обслуговуючій системі $W_s = 0{,}3736$ (приблизно 22 хв.), тобто менше 30 хв., як це було в прикладі 2.6, коли всім автомобілям, що прибувають дозволялося стати у чергу. Зменшення цього показника обслуговуючої системи приблизно на 25% забезпечено за рахунок втрати близько 4,8% потенційних клієнтів через обмежену кількість місць на стоянці автомобілів [22].

Приклад 10.

У невеликому містечку функціонують дві служби таксі. Кожна з них має в своєму розпорядженні два автомобілі, і за наявною інформацією, замовлення на обслуговування діляться між службами майже порівну. Це підтверджується тим фактом, що замовлення до диспетчерських відділень обох служб надходять з однією і тією ж інтенсивністю рівною 8 викликам на годину. Середній час виконання однієї заявки становить 12 хвилин. Заявки на обслуговування надходять відповідно до розподілу Пуассона, а час обслуговування клієнтів розподілено за експоненціальним законом. Недавно обидві служби були придбані інвестором, який зацікавлений у їхньому об'єднанні з єдиним диспетчерським пунктом для забезпечення швидшого обслуговування клієнтів. Потрібно проаналізувати пропозиції нового господаря.

З точки зору теорії масового обслуговування таксі являють собою обслуговуючі пристрої, а виклик таксі є сервісом. Кожна служба таксі може бути представлена моделлю $(M/M/2):(GD/\infty/\infty)$ з $\lambda = 8$ викликів на годину і $\mu = 60/12 = 5$ поїздок на таксі на годину. Об'єднання служб таксі призведе до моделі $(M/M/4):(GD/\infty/\infty)$ з $\lambda = 2 \times 8 = 16$ викликів на годину та $\mu = 5$ поїздок на одне таксі на годину.

Припустимою мірою для порівняння двох моделей обслуговування є середній час очікування клієнтом таксі з моменту його виклику до моменту прибуття автомобіля, тобто W_q. На рис. 3.6 представлені вихідні дані програми TORA для двох описаних моделей. Результати показують, що час очікування клієнтом приїзду автомобіля дорівнює 0,356 год (приблизно 21 хв.) для моделі обслуговування з двома таксомоторними службами та 0,149 год (приблизно 9 хв.) для моделі обслуговування в об'єднаному варіанті. Значне зменшення (більш ніж 50%) функціонального показника розглянутої обслуговуючої системи робить очевидною доцільність об'єднання двох служб таксі.

Problem title: Example 2.3.4.1
Comparative measures: Nbr of scenarios=2

Nbr	c	Lambda	Mu	l`da_eff	Ls	Ws	Lq	Wq
1	2	8.000	5.000	8.000	4.444	0.556	2.844	0.356
2	4	16.000	5.000	16.000	5.586	0.349	2.386	

Рис. 3.6. Вихідні дані програми TORA для двох описаних моделей

З наведеного аналізу випливає, що об'єднання систем обслуговування завжди забезпечує ефективніший режим роботи. Цей висновок залишається справедливим навіть у тому випадку, коли завантаженість усіх послуг дуже висока.

Приклад 11.

Нехай у задачі, пов'язаній з об'єднанням служб таксі, що розглядалася у прикладі 10 відомо, що об'єднана служба таксі не має фінансових можливостей для покупки нових автомобілів. Друг нового господаря радить інформувати пасажирів про можливі затримки з прибуттям замовленої автомобілів, як тільки список очікуваних клієнтів досягає шести. Цей захід, без сумнівно, змусить нових клієнтів шукати обслуговування в іншому місці, що зменшить час очікування тих клієнтів, які вже очікують у черзі. Необхідно дізнатися, наскільки корисною є порада друга [34].

Обмеження списку тих, хто чекає в черзі, до 6 клієнтів рівносильне тому, що ємність системи стає рівною $N=6+4=10$ клієнтів. Отже, ми маємо справу з системою обслуговування моделі $(M/M/4):(GD/10/\infty)$ з $\lambda=16$ клієнтів на годину та $\mu=5$ поїздок на годину. На рис. 3.7 представлені вихідні дані, отримані за допомогою програми TORA для цієї моделі.

```
Problem title: Example 2.3.4.2
Scenario 1 – (M/M/4):(GD/10/*)
--------------------------------------------------------------------
Lambda =    16.00000        Lambda eff =    15.42815
Mu =    5.00000             Rho =    3.20000

Ls =    4.23984             Lq =    1.15421
Ws =    0.27481             Wq =    0.07481
--------------------------------------------------------------------
Values of p(n) for n=0 to 10, else p(n)<.00001

0 0.03121    1 0.09986    2 0.15977    3 0.17043    4 0.13634
5 0.10907    6 0.08726    7 0.06981    8 0.05584    9 0.04468
10 0.03574

Cumulative values of p(n) for n=0 to 10

0 0.03121    1 0.13106    2`0.29084    3 0.46126    4 0.59760
5 0.70667    6 0.79393    7 0.86374    8 0.91958    9 0.96426
10 1.00000
```

Рис. 3.7 Вихідні дані, отримані за допомогою програми TORA для цієї моделі

Середній час очікування W_q за відсутності обмеження на ємність системи дорівнює 0,149 год (=9 хв.) (див. рис. 3.6), що майже вдвічі більше значення 0,075 год (=4,5 хв.) аналогічного показника за наявності обмеження на ємність системи. Це істотне зменшення функціональної характеристики системи досягнуто за рахунок втрати приблизно 3,6% потенційних клієнтів. Цей результат, однак, не відображає можливу втрату інтересу клієнтів до діяльності служби таксі.

Приклад 12.

Інвестор вкладає 1000 доларів на місяць у спеціальний тип облігацій фондової біржі. Так як інвестор повинен чекати можливості хорошої "купівлі", фактично час здійснення цієї покупки є випадковим. Інвестор зазвичай тримає облігації в середньому три роки, але продасть їх у випадковий момент часу, коли представиться така можливість. Хоча інвестор відомий як хитрий біржовий гравець, досвід минулого показує, що близько 25% облігацій втрачають у ціні приблизно 20% на рік. Інші 75% облігацій підвищуються в ціні приблизно на 12% на рік. Оцінимо середню вартість акцій інвестора протягом тривалого періоду

Цю ситуацію можна представити у вигляді моделі $(M / M / \infty):(GD / \infty / \infty)$, так як інвестор не повинен чекати у черзі, щоб купити або продати облігації. Середній час між розміщеннями замовлення дорівнює 1 місяць, що дає значення $\lambda = 12$ облігацій на рік. Інтенсивність продажу облігацій дорівнює $\mu = 1/3$ облігацій на рік [35].

При зазначених значеннях λ і μ отримуємо

$$L_s = \rho = \frac{\lambda}{\mu} = 36 \text{ облігацій.}$$

Середня річна вартість облігацій інвестора протягом тривалого періоду оцінюється наступною величиною:

$$(0{,}25L_s \times 1000)(1 - 0{,}20) + (0{,}75L_s \times 1000)(1 + 0{,}12) = 63990 \text{ доларів.}$$

Приклад 13.

Компанія має в своєму розпорядженні цех, що налічує 22 верстати. Відомо, що кожен верстат виходить з ладу в середньому один раз на дві години. На його ремонт іде в середньому 12 хв. Як час між поломками верстатів, так і час, необхідний для ремонту, є експоненціально розподіленими випадковими величинами. Компанія зацікавлена у визначенні мінімальної кількості механіків, необхідних забезпечення «плавної» роботи цеху [33].

Описану ситуацію можна проаналізувати, досліджуючи продуктивність верстатів як функцію числа механіків. Така міра продуктивності може бути визначено у такому вигляді:

$$\begin{pmatrix} \text{Продуктивність} \\ \text{верстатів} \end{pmatrix} = \frac{\text{Кількість верстатів} - \text{не працюючі верстати}}{\text{Кількість верстатів}} \times 100 = \frac{22 - L_s}{22} \times 100.$$

На рис. 3.8 представлені отримані за допомогою програми TORA порівняльні характеристики системи обслуговування для $R = 1,2,3,4$ при $\lambda = 0{,}5$ поломки на годину на один верстат і $\mu = 5$ ремонтів на годину. Відповідні значення виготовлення представлені в таблиці 3.4.

Таблиця 3.4

Кількість механіків, R	1	2	3	4

Продуктивність верстатів (у %)	45,44	80,15	88,79	90,45
Зростання продуктивності (у %)	____	34,71	8,64	1,66

Problem title: Example 2.3.5.1
Comparative measures: Nbr of scenarios=4

Nbr	c	Lambda	Mu	l`da_eff	Ls	Ws	Lq	Wq
1	1	0.500	5.000	4.998	12.004	2.402	11.004	2.202
2	2	0.500	5.000	8.816	4.368	0.495	2.604	0.295
3	3	0.500	5.000	9.767	2.466	0.252	0.513	0.052
4	4	0.500	5.000	9.950	2.100	0.211	0.110	0.011

Рис. 3.8. Порівняльні характеристики системи обслуговування, отримані за допомогою програми TORA

Наведені результати показують, що неприйнятною є наявність лише одного механіка, тому що в цьому випадку буде дуже низька продуктивність (45,44%). Якщо кількість механіків збільшується до двох, то продуктивність верстатів зростає на 34,71% і сягає 80,15%. При роботі трьох механіків продуктивність верстатів зростає приблизно на 8,64% і досягає значення 88,79%, у той час як за наявності чотирьох механіків продуктивність зростає лише на 1,66% і досягає значення 90,45%.

Судячи з цих результатів, використання двох механіків може бути виправданим. Ситуація з трьома механіками є не такою переконливою, так як продуктивність при цьому зростає лише на 8,64%. Можливо, що порівняння у грошовому еквівалент змісту третього механіка і прибутку, обумовленого зростанням продуктивності верстатів, на 8,64% можна використовувати для вирішення цього питання. Що стосується прийому на роботу четвертого механіка, то мізерне зростання продуктивності верстатів на 1,66%, який при цьому досягається, не виправдовує цього кроку [11].

Приклад 14.

Нехай у задачі про автомийку з прикладу 10 зроблено додаткове припущення, що встановлено нове обладнання, тому час обслуговування для всіх автомобілів є постійним та рівним 10 хв. Як нове обладнання впливає на функціонування автомийки [12]?

На підставі даних прикладу 10 маємо, що $\lambda_{еф} = \lambda = 4$ автомобілі на годину. Час обслуговування є постійним, тому $M\{t\} = \frac{10}{60} = \frac{1}{6}$ години і $D\{t\} = 0$.

Отже,

$$L_s = 4\left(\frac{1}{6}\right) + \frac{4^2\left[\left(\frac{1}{6}\right)^2 + 0\right]}{2\left(1 - \frac{4}{6}\right)} = 1{,}333 \text{ автомобілів}$$

$$L_q = 1{,}333 - \left(\frac{4}{6}\right) = 0{,}667 \text{ автомобіля}$$

$$W_s = \frac{1{,}333}{4} = 0{,}333 \text{ години}$$

$$W_q = \frac{0{,}667}{4} = 0{,}167 \text{ години.}$$

Цікаво відзначити, що, незважаючи на те, що інтенсивності як надходження клієнтів, так і обслуговування в аналізованої моделі такі ж, як і в пуассонівському випадку з прикладу 10 ($\lambda = 4$ автомобіля на годину та $\mu = \frac{1}{M\{t\}} = 6$ автомобілів на годину), середній час очікування у цьому випадку менший, оскільки час обслуговування є постійним, що свідчить таблиця 3.5.

Таблиця 3.5

	$(M/M/1):(GD/\infty/\infty)$	$(M/D/1):(GD/\infty/\infty)$
W_s (години)	0,5	0,333
W_q (години)	0,333	0,167

У цих результатах є сенс, оскільки постійний час обслуговування передбачає більшу визначеність у функціонуванні системи.

Приклад 15.

Видавнича фірма купує високошвидкісний копіювальний апарат для комерційних цілей. Продавці запропонували чотири моделі копіювальних апаратів, характеристики яких наведені в наступній таблиці 3.6.

Таблиця 3.6

Модель копіювального апарату	Експлуатаційні витрати (долар. за год)	Швидкість друку (стор. за хв.)
1	15	30
2	20	36
3	24	50
4	27	66

Замовлення надходять на фірму відповідно до пуассонівського розподілу з математичним сподіванням чотири роботи протягом 24-годинного дня. Обсяг роботи є випадковою величиною, але в середньому становить приблизно 10000 сторінок. Договори з клієнтами передбачають штраф у сумі 80 долар. (за одну роботу) за затримку виконання замовлення на день. Який копіювальний апарат слід купити фірмі [32]?

Нехай індекс i представляє номер моделі копіювального апарату $(i = 1,2,3,4)$. Загальна очікувана вартість обслуговування в день, пов'язана з використанням копіювального апарату i представлена в наступному вигляді:

$$COC_i = CTC_i + CTO_i = C_{1i} \cdot 24 + C_{2i} L_{si} = 24C_{1i} + 80L_{si}, \quad i = 1,2,3,4$$

Значення величин C_{1i} наведено у постановці завдання. Визначимо значення L_{si} вважаючи, що з практичної точки зору копіювальний апарат може розглядатися, як модель $(M/M/1):(GD/\infty/\infty)$ обслуговуючої системи. При цьому інтенсивність надходження заявок $\lambda = 4$ роботи на день, а інтенсивності μ_i обслуговування, відповідні моделі i-го копіювального апарату, наведені в таблиці 3.7.

Таблиця 3.7

Модель i	Інтенсивність обслуговування μ_i (роботи/день)
1	4,32

2	5,18
3	7,20
4	9,50

Проілюструємо визначення інтенсивності обслуговування з прикладу моделі 1.

Середній час виконання роботи $=\frac{10000}{30}\cdot\frac{1}{60}=5{,}56$ год.

Отже,

$$\mu_1=\frac{24}{5{,}56}=4{,}32 \text{ роботи на день.}$$

Значення L_{si} обчислені за допомогою програми TORA, наведені в наступній таблиці 3.8.

Таблиця 3.8

Модель i	λ_i (роботи/день)	μ_i (роботи/день)	L_{si} (роботи)
1	4	4,32	12,50
2	4	5,18	3,39
3	4	7,20	1,25
4	4	9,50	0,73

Отримано наступні вартісні характеристики (у долар.) для чотирьох моделей копіювальних апаратів, що розглядаються (таблиця 3.9).

Таблиця 3.9

Модель i	CTC_i (у долар.)	CTO_i (у долар.)	COC_i (у долар.)
1	360,00	1000,00	1360,00
2	480,00	271,20	751,20
3	**576,00**	**100,00**	**676,00**
4	648,00	58,40	706,40

Отже, для третьої моделі копіювального апарату вартість мінімальна.

Приклад 16.

На інструментальний склад, де працюють кілька службовців, надходять заявки на заміну ріжучого інструменту відповідно до розподілу Пуассона з середнім значенням 17,5 заявки на годину. Кожен службовець може виконати загалом 10 заявок на годину. Вартість найму нового службовця складає 12 долар. на годину. Втрати, пов'язані з очікуванням верстата, оцінюються приблизно 50 долар. на годину. Необхідно визначити кількість працівників для аналізованої системи обслуговування.

Описана ситуація відповідає моделі $(M/M/c)$ в якій потрібно визначити оптимальне значення c. Отже, $x = c$, і відповідна вартість модель має вигляд

$$COC(c) = C_1 c + C_2 L_s(c) = 12c + 50 L_s(c).$$

Зауважимо, що $L_s(c)$ є функцією кількості працюючих в майстерні.

Використовуємо модель $(M/M/c):(GD/\infty/\infty)$ з $\lambda = 17{,}5$ заявок на годину та $\mu = 10$ заявок на годину. У цьому відношенні модель досягне стійкого стану лише при умові, що $c > \lambda/\mu$, тобто у аналізованому прикладі c має дорівнювати, принаймні, 2. Наведена нижче таблиця 3.10 містить результати обчислень визначення оптимального значення c. Значення $L_s(c)$ обчислені за допомогою програми TORA [25].

Таблиця 3.10

c	$L_s(c)$ (заявки)	$COC(c)$ (у долар.)
2	7,467	397,35
3	2,217	142,35
4	**1,842**	**140,10**
5	1,769	148,45
6	1,754	159,70

Отже, оптимальна кількість працівників дорівнює 4.

Приклад 17

Припустимо, що в прикладі 16 необхідно визначити таку кількість службовців на складі, щоб середній час очікування інструменту був меншим 5 хв. Одночасно потрібно, щоб відсоток часу, протягом якого персонал майстерні залишається вільним, не перевищував 20%.

Зауважимо, що і без обчислень очевидним є той факт, що рівень абсолютного пріорітету у 5 хв. часу очікування інструменту (тобто $W_s \leq 5$) є недосяжним, оскільки з даних завдання випливає, що середній час обслуговування дорівнює 6 хв. Наведена нижче таблиця містить значення W_s і X, як функцій від c і підтверджує це зауваження. Дійсно при $c \geq 5$ маємо $W_s = 6$. Це означає, що час очікування в черзі буде відсутнім. Таблиця 3.11 також показує, що подальше збільшення числа службовців може лише підвищити відсоток їх простою X [20].

Таблиця 3.11

c	2	3	4	5	6	7	8
W_s (хв.)	25,6	7,6	6,3	6,0	6,0	6,0	6,0
X (%)	12,5	41,7	56,3	65,0	70,8	75,0	78,0

У цій ситуації ми фактично нічого не можемо зробити, так як завдання не можна вирішити шляхом збільшення кількості службовців на складі. Необхідно або зменшити час обслуговування в системі, або визнати, що в реальній ситуації, якій відповідає математична модель, інтенсивність запитів на інструменти надзвичайно висока ($\lambda = 17,5$ запитів на годину) . Очевидно, що саме на останній факт має бути звернена увага. Наприклад, можна провести дослідження причин такого високого попиту на заміну інструменту. Чи може бути причиною цього неадекватна конструкція самого інструменту? Або причиною є те, що оператори верстатів цілеспрямовано намагаються підірвати виробництво, таким чином висловлюючи своє невдоволення?

ВИСНОВКИ

В ході виконання магістерської роботи було здійснено огляд наявної літератури з теми, з’ясовано як можна застосувати теорію масового обслуговування до побудові ймовірнісних моделей.

Прикладні задачі теорії масового обслуговування пов'язані з дослідженням будь-яких процесів, що складаються з багатьох однорідних елементарних операцій, на здійснення яких впливають випадкові чинники. Наведено приклади.

Теорія систем масового обслуговування – це область прикладної математики, що є відокремленою частиною теорії випадкових процесів з комплексом завдань для самостійного дослідження. Вона базується на теорії ймовірностей та математичній статистиці. Теорія систем масового обслуговування в даний час має широку область застосування не тільки в економіці, а й в інших значущих галузях для життя.

Дослідження СМО можна звести до побудови і вивчення деякого випадкового процесу, який описує еволюцію системи. Щоб математично описати СМО, необхідно описати властивості вхідного потоку однорідних подій – потоку замовлень, структуру системи, дисципліну і характеристики обслуговування, а також зазначити ті характеристики, які треба визначити.

На основі дослідження розроблено методичний посібник, який можна використовувати студентам при вивченні дисципліни "Математичне моделювання».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бєлий Є К. Введення в теорію масового обслуговування : навч. посіб. Петрозаводськ : ПетрГУ, 2014. 76 с.

2. Боровков А. А. Ймовірнісні процеси теорії масового обслуговування. Москва : Наука, 1972. 428 с.

3. Вішнєвській В. М. Теоретичні основи проектування комп'ютерних мереж. Москва : Техносфера, 2003. 512 с.

4. Гіхман І. І., Скороход А. В. Введення в теорію випадкових процесів. Москва : Наука, 1965. 656 с.

5. Гнеденко Б. В. Бесіди про теорію масового обслуговування. Москва : Ліброком, 2017. 72 с.

6. Гнеденко Б. В., Коваленко І.Н. Введення в теорію масового обслуговування. Москва : УРСС 2021. 400 с.

7. Дослідження операцій і методи оптимізації / Корольова М. Є., Павленка В. І., Савіна О. В., Тимошенка А. Г. Київ : Україна, 2007. 177 с.

8. Жерновий Ю. В. Марковські моделі масового обслуговування. Львів : ЛНУ ім. Івана Франка, 2004. 154 с.

9. Кармелюк Г. І. Теорія ймовірності та математична статистика. Київ : Центр навчальної літератури. 2019. 576 с.

10. Карташевський В. Г. Основи теорії масового обслуговування. Москва : Гаряча лінія – Телеком, 2013. 130 с.

11. Кірпічніков А. П. Методи прикладної теорії масового обслуговування. Москва : УРСС, 2018. 224 с.

12. Клейнрок Л. Теорія масового обслуговування. Москва : Машинобудування, 1979. 432 с.

13. Коваленко І. Н., Каштанов В. А., Івченко Г. І. Теорія масового обслуговування. Москва : Ліброков, 2012. 304 с.

14. Коломієць С. В. Теорія випадкових процесів : навч. посіб. Вид. 2-ге, Суми : ДВНЗ "УАБС НБУ" 2013. – Ч. ІІ. – 103 с.

15. Косолапов А.А. Аналітичні моделі масового обслуговування в задачах проектування інформаційних систем : навч. посіб. Дніпропетровськ : «LikePrint», ФОП Гечка Т.О., 2015. 186 с.

16. Кофман А. Методи та моделі дослідження операцій. Москва : Світ, 1966. 432 с.

17. Крилов В. В., Сомохвалов С. С. Теорія телетрафіка і її застосування. Спб : ПХБ-Петербург, 2005. 288 с.

18. Лаврусь О. В., Миронов Ф. С. Теорія масового обслуговування. Самара, 2002. 38 с.

19. Ложковський А.Г. Теорія масового обслуговування в телекомунікаціях. Одеса : ОНАЗ ім. О.С. Попова, 2010. 112 с.

20. Мур Дж., Уедерфорд Л. Економічне моделювання в Microsoft Excel. Москва: Вільямс, 2004. 1019 с.

21. Рижіков Ю. І. Численні методи теорії черг. Санг-Петербург : Лань. 2019. 512 с.

22. Сааті Т. Елементи теорії масового обслуговування та її застосування. Москва : Ліброком, 2010. 520 с.

23. Солнішкіна І. В. Теорія систем масового обслуговування : навч. посіб. Комсомольськ-Амуре : ФГБОУ ВПО «КнАГТУ», 2015. 76 с.

24. Таха Х. А. Введення в використання операцій. Київ : Вільямс, 2005. 912 с.

25. Томашевський В. М. Моделювання систем. Київ : Група BHV, 2005. 352 с.

26. Уолренд Дж. Введення в теорію мереж масового обслуговування. Москва : Мир, 1993. 336 с.

27. Хінчин А. Я. Роботи з математичної теорії масового обслуговування. Москва : УРСС, 2019. 240 с.

28. Шимко, П.Д., Власов М. П. Оптимальне управління економічними системами. Москва : ІНФРА-М, 2014. 320 с.

29. Hall R. Queueing Methods for Service and Manufacturing, Prentice Hall, Upper Saddle River, N.J., 1991.

30. Lipsky L. Queueing Theory, A Linear Algebraic Approach, Macmillan, New York, 1992.

31. Morse P. Queues, Inventories, and Maintenance, Wiley, New York, 1958.

32. Parzen E. Stochastic Processes, Holden-Day, San Francisco, 1962.

33. Saaty T. Elements of Queueing Theory with Applications, Dover, New York, 1983.

34. Tijms H.C. Stochastic Models - An Algorithmic Approach, Wiley, New York, 1994.

35. Obretenov A., Dimitrov B. Teoria masowej obsługi. PWN, 1989.