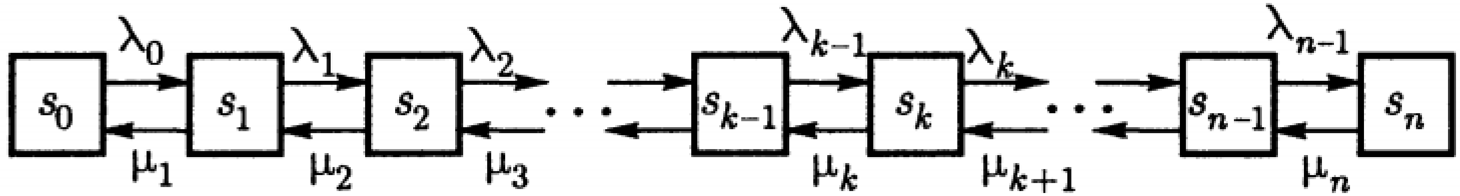
Глава 3. Системы массового обслуживания

В презентации использованы копии со страниц учебника «Е.С.Вентцель, Л.А. Овчаров. Задачи и упражнения по теории вероятности» и иллюстрации из «Л. Клейнрок. Теория массового обслуживания» , а также копии из «И.В. Солнышкина. Теория систем массового обслуживания» и

«Н.В.Кошуняева,Н.Н.Патронова.Теория массового обслуживания» .

При подготовке также использован материал учебника «Е.С.Вентцель. Исследование операций: задачи, принципы, методология».



- Процесс Маркова, схема (граф состояний) которого имеет вид, показанный на рис. называются ***процессом гибели и размножения***. Это название заимствовано из биологических задач, где состояние популяции s_k означает наличие в ней k единиц. Переход в право связан с «размножением» единиц, а влево — с их «гибелью».
  «Интенсивности размножения» -(λ_0 ,…,λ_{n-1}). «Интенсивности гибели» - (μ_1,…, μ_n)

- Из уравнений Колмогорова можно найти финальные вероятности состояний:

$$p_1^* = \frac{\lambda_0}{\mu_1}\, p_0^*; \quad p_2^* = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\, p_0^*;\ldots;$$

$$p_k^* = \frac{\lambda_0\lambda_1\ldots\lambda_{k-1}}{\mu_1\mu_2\ldots\mu_k}\, p_0^* \quad (k = 0, \ldots, n);\ldots;$$

- $$p_n^* = \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}\, p_0^*; \quad p_0^* = \left\{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \ldots + \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}\right\}^{-1}$$

Кроме того интерес представляют ещё моментные характеристики процесса: усреднение по вероятности для функции популяции получается из: $\frac{dEX(t)}{dt} = \Sigma(\lambda_k-\mu_k)p_k(t)$

а мера отклонений от него(дисперсия): $\frac{d\,DX(t)}{dt} = \Sigma(\lambda_k+\mu_k+2k(\lambda_k-\mu_k)-2EX(t)\,(\lambda_k-\mu_k))p_k(t)$

- Причём уравнение для математического ожидания величины популяции следует непосредственно из уравнений Колмогорова.
  Действительно, если расписать (k-1)-ый, k-ый и (k+1)-ый члены суммы
- $\frac{d\,EX(t)}{dt} = \sum_{k=0}^{n} k\, p'_k(t)$ через уравнения Колмогорова :
- $(k-1)\,[\,-(\lambda_{k-1}+\mu_{k-1})\,p_{k-1} + \lambda_{k-2}\,p_{k-2} + \mu_k p_k\,]$
- $k\,[\,-(\lambda_k+\mu_k)\,p_k + \lambda_{k-1}\,p_{k-1} + \mu_{k+1}p_{k+1}\,]$
- $(k+1)\,[\,-(\lambda_{k+1}+\mu_{k+1})\,p_{k-1} + \lambda_k\,p_k + \mu_{k+2}p_{k+2}\,]$
- то легко видеть, что остальные члены суммы не содержат p_k и поэтому
- коэффициент при p_k можно найти из суммы выделенных слагаемых:
- $(k-1)\,\mu_k - k(\lambda_k+\mu_k) + (k+1)\lambda_k = \lambda_k - \mu_k$, т.е.
-

$$\frac{d\,EX(t)}{dt} = \sum_{k=0}^{n} (\lambda_k - \mu_k)\, p_k(t)$$

- **Система массового обслуживания (СМО)** – это любая система, предназначенная для обслуживания каких-либо заявок (требований, клиентов), поступающих на нее в случайные моменты времени. Состоит из входящего потока (заявок) и одного, либо нескольких потоков обслуживания (каналов, линий обслуживания).
- **Примерами** СМО могут служить:
- расчетно-кассовые узлы в банках, на предприятиях;

  персональные компьютеры (или сервера в сети), обслуживающие поступающие заявки или требования на решение тех или иных задач;
- станции технического обслуживания автомобилей; АЗС;
- аудиторские фирмы;
- отделы налоговых инспекций, занимающиеся приёмкой и проверкой текущей отчетности предприятий;
- телефонные станции и т. д.
-

  Открытая СМО – когда входящий поток требований не зависит от состояния системы, поступая в неё извне.

- *Замкнутая СМО* – когда поток требований находится внутри системы, зависит от состояния системы. Пример:
  Компьютерный класс, или сложная вычислительная система в которой узлы периодически требует ремонта, перезагрузки или переустановки программного обеспечения (т.е. создают заявки на обслуживание).

  Различают СМО с отказами (потерями) и СМО с очередью. В СМО с отказами заявка, пришедшая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем в процессе работы не участвует. В СМО с очередью заявка, пришедшая в момент занятости всех каналов, не покидает СМО, а становится в очередь и ждет, пока не освободится какой-нибудь канал. Число мест в очереди может быть как ограниченным, так и неограниченным. Очередь может иметь ограничения не только по количеству стоящих в ней заявок (длине очереди), но и по времени ожидания (такие СМО называются «системами с нетерпеливыми клиентами)
-

Дисциплина очереди определяет принцип, в соответствии с которым поступающие на вход обслуживающей системы требования подключаются из очереди к процедуре обслуживания. Чаще всего используются дисциплины очереди, определяемые следующими правилами:

- ***Упорядоченный тип*: первым пришел – первый обслуживаешься**; first in first out (**FIFO, FF**) - самый распространенный тип очереди.
- ***С приоритетами***: тот же принцип, но требования в потоке делятся на приоритетные и не приоритетные, и образует две очереди (вторая обслуживается только во время, когда нету первой) **PR** (Priority)
- ***Стековый или магазинный:*** *пришел последним — обслуживаешься первым* (**LIFO, LF**) (обойма для патронов, тупик на железнодорожной станции, поездка на лифте).
- ***Равновероятный*** выбор заявки, **SP** (Same Probability)

  Бывают и другие дисциплины о.)

  Механизм обслуживания *(Дисциплина обслуживания*) - определяется характеристиками самой процедуры обслуживания и структурой обслуживающей системы, в том числе:
- количество каналов обслуживания ;
- продолжительность процедуры обслуживания (вероятностное распределение времени обслуживания заявок)–задаёт **интенсивность обслуживания канала** (т.е. интенсивность потока на канале обслуживания при идеальной, максимальной загруженности – когда канал всегда занят);

Различие (либо равенство) интенсивностей среди каналов обслуживания и возможность задания предпочтений выбора между ними в последнем случае;

- количество требований, удовлетворяемых в результате процедуры обслужив-я;
- вероятность выхода из строя обслуживающего канала;
- возможность\невозможность участия в обслуживании заявки сразу нескольких каналов;
- все каналы могут обслуживать любые заявки – тогда СМО называют ***полнодоступной*** , либо некоторые каналы обслуживают лишь определённые категории заявок (***неполнодоступные СМО***, в которых рассматривают несколько входящих потоков заявок)

• Для краткой записи СМО приняты т.н. **символы (обозначения) Кендалла**:

Обычно используют **4** символа: первый характеризует входной поток требований, второй — распределение длительностей обслуживания и третий — число приборов в обслуживающей системе. **а 4й** символ - число мест в очереди (0, ∞ или конечн. число).

Приведем перечень общепринятых символов, характеризующих распределения вероятностей, которые ставятся в соответствие моделям массового обслуживания:

M — экспоненциальное распределение продолжительностей интервалов между поступлениями требований или длительностей обслуживания (от определяющего слова «марковский»);

D — детерминированное (или регулярное) распределение длительностей интервалов между поступлениями требований или длительностей обслуживания;

E_n — n-фазное распределение Эрланга

(возникает в потоке Эрланга)

(т.е. пальмовский)

(GU) GI — рекуррентный характер входного потока [1]) без каких-либо специальных предположений относительно функции распределения;

G — общий вид распределения длительностей обслуживания (т. е. не делается никаких конкретизирующих предположений относительно функции распределения).

- Все символы записываются подряд через черту. Например, простейшая СМО с одним каналом и отказами (потерями, нет очереди) :
  M / M / 1 / 0
  Остальные символы (начиная с пятого – не обязательны, характеризуют специфические особенности устройства СМО).
-
  Например, далее могут добавить ещё символ, обозначающий ограничение на источники нагрузки (т.е. если в СМО поступает конечное число требований):
- M / M / 1 / 0 / K (K – целое положит. число)

- И после этого ещё символы для обозначения дисциплины очереди и особого устройства механизма обслуживания (если отличны от FF| FM), например:

  D / M / 1 / 0 / LF / FM

  Также в последнем разряде можно было бы увидеть:

  FM (Full Matrix)– полнодоступная система,
  G – неполнодоступная система (Grading)
  и другое (например, определённая разновидность неполнодоступной системы)

Задачи теории массового обслуживания — нахождение вероятностей различных состояний СМО, а также установление зависимости между заданными параметрами (числом каналов n, интенсивностью потока заявок λ, распределением времени обслуживания и т.д.) и *характеристиками эффективности* работы СМО. В качестве таких характеристик могут рассматриваться, например, следующие:

среднее число заявок A, обслуживаемое СМО в единицу времени, или *абсолютная пропускная способность* СМО; (= интенсивность потока обслуживания)

вероятность обслуживания поступившей заявки Q или *относительная пропускная способность* СМО; $Q = A / \lambda$;

вероятность отказа $P_{\text{отк}}$, т.е. вероятность того, что поступившая заявка не будет обслужена, получит отказ; $P_{\text{отк}} = 1 - Q$;

среднее число заявок в СМО (обслуживаемых или ожидающих в очереди) $\bar{z}$;

среднее число заявок в очереди $\bar{r}$; (также распространено обозн. $L_{\text{оч.}}$)

среднее время пребывания заявки в СМО (в очереди или под обслуживанием) $\bar{t}_{\text{сист}}$;

среднее время пребыван заявки в очереди $\bar{t}_{\text{оч}}$; среднее число занят каналов $\bar{k}$.

В общем случае все эти характеристики зависят от времени. Но многие СМО работают в неизменных условиях достаточно долгое время, и поэтому для них успевает установитьс режим, близкий к стационарному. (Тогда соотв. процесс Маркова эргодичен) повсюду, не оговаривая этого каждый раз специально, будем вычислять финальные вероятности состояний и финальные характеристики эффективности СМО, относящиеся к предельному, стационарному режиму ее работы.

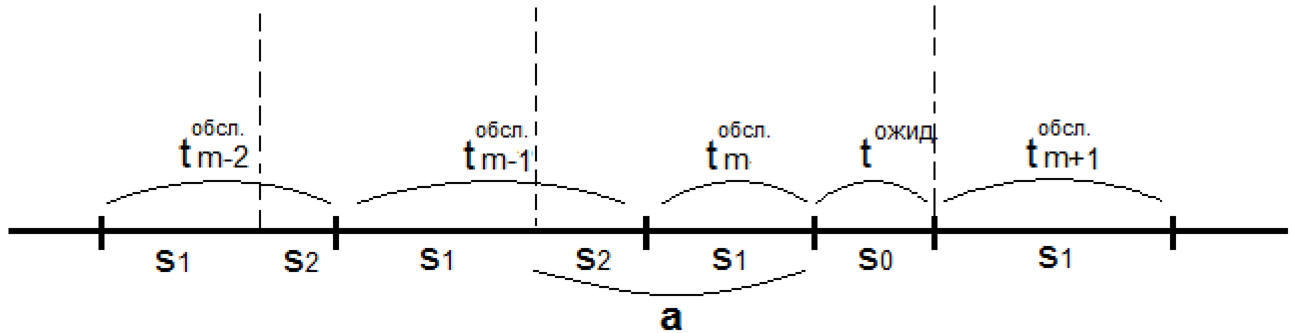
- **ОПР.**
  СМО, в которой поток требований простейший, а время обслуживания на всех каналах распределено по показательному закону будем называть *простейшей.*

- **ТЕОРЕМА**
- Любая СМО стандартного механизма обслуживания (полнодоступная, с равными интенсивностями каналов обслуживания) с пуассоновским потоком требований и определённым распределением времени обслуживания моделируется процессом Маркова.

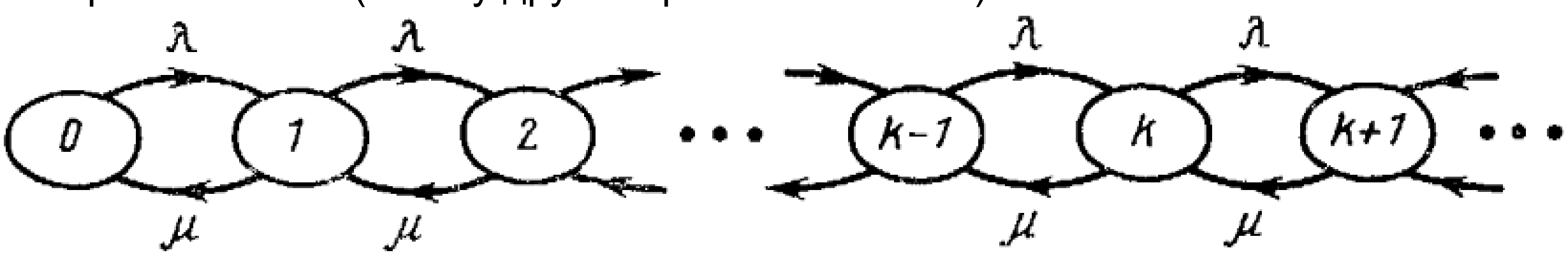
- Причём, в случае стационарности потока требований и экспоненциального распределения времени обслуживания имеем однородный процесс Маркова.

  СЛЕДСТВИЕ
  Любая простейшая СМО моделируется процессом Маркова.

- **Д-ВО** Теорему можно доказать, начав с рассмотрения схемы работы (поток событий) канала для одноканальной СМО с неограниченной очередью:



- в которой пунктиром обозначены моменты поступления новых заявок и подписаны периоды нахождения системы в состояниях
- S_2 – канал работает, одна заявка в очереди
- S_1 – канал работает, нет заявок в очереди
- S_0 – канал свободен (простаивает)
- Схема моделирующего процесса будет частным случаем схемы гибели и размножения (как и у других простейших СМО):



- Покажем, что все интенсивности переходов именно такие: λ и μ (и что эти величины представляют в самой СМО). Далее можно увидеть, что в других простейших СМО схема составляется по аналогии.
- (Подробнее см.в приложении)

Простейшая СМО с отказами (задача Эрланга). На n-канальную СМО с отказами поступает простейший поток заявок с интенсивностью λ, время обслуживания — показательное с параметром $\mu = 1 / \bar{t}_{обсл}$. Состояния СМО нумеруются по числу заявок, находящихся в СМО (в силу отсутствия очереди, оно совпадает с числом занятых каналов):

s_0 — СМО свободна;

s_1 — занят один канал, остальные свободны; …;

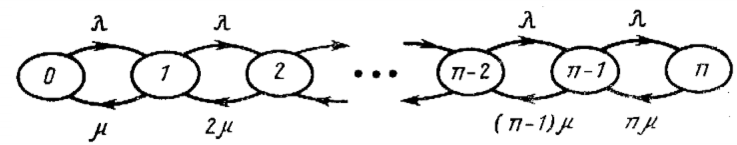
s_k — занято k каналов, остальные свободны ($1 \le k \le n$); …;

s_n — заняты все n каналов.

Финальные вероятности состояний выражаются формулами Эрланга:

$$p = \left\{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!}\right\}^{-1}; \quad p_k = \frac{\rho^k}{k!} p_0 \;(k = 1, 2, \ldots, n),$$

где $\rho = \lambda / \mu$.



Характеристики эффективности:

$$A = \lambda (1 - p_n); \quad Q = 1 - p_n; \quad P_{отк} = p_n; \quad \bar{k} = \rho (1 - p_n).$$

- Финальная вероятность последнего состояния выражается соответственно т.н. формулой потерь Эрланга:

$$p_n = \frac{(\lambda/\mu)^n / n!}{\sum_{k=0}^{n} (\lambda/\mu)^k / k!}$$

В обозначениях Кендалла эта задача запишется как

M / M / n / 0. Её частным случаем является простейшая одноканальная СМО с отказами (M / M / 1 / 0)

Задача 1.

Станция наведения истребителей имеет 3 канала. Каждый канал может одновременно наводить один истребитель на одну цель. Среднее время наведения истребителей на цель равно 2 мин. Поток целей простейший с плотностью *75* самолета *в час* . Найдите среднюю долю целей, проходящих через зону действия не обстрелянными, если цель, по которой наведение не началось в момент, когда она вошла в зону действия истребителей, вообще остается неактивной. Сделайте вывод о работе станции.

Решение: В задаче дано количество каналов обслуживания $n = 3$;

$\lambda = 75$ супост $/ч$; $t_{обсл} = 2$ *мин* $\Rightarrow \mu = \frac{1}{2} = 0,5$ супост $/мин$ *или* $60 * 0,5 =$

$= 30$ супост $/ч$; $\rho = \frac{75}{30} = 2,5$.

Находим вероятность того, что обслуживанием не занят ни один канал: $P_0 = \left(\sum_{n=0}^{n} \frac{\rho^n}{n!}\right)^{-1} = \left(1 + 2,5 + \frac{2,5^2}{1*2} + \frac{2,5^3}{1*2*3}\right)^{-1} = 0,11 \approx 11\,\%$;

$$P_{отк} = P_n = \frac{\rho^n}{n!} * P_0 = \frac{2,5^3}{3!} * 0,11 = 0,282 \approx 28\,\%;$$

$$Q = P_{обсл} = 1 - P_{отк} = 1 - 0,282 = 0,718;$$

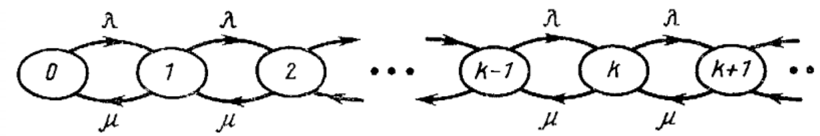
$$\overline{k} = \rho * P_{обсл} = \frac{A}{\mu} = 2,5 * 0,718 = 1,795 \approx 2 \text{ канала.}$$

Вывод: нужно добавить ещё один канал наведения.

- **Упражнение.** Для каждой из описанных СМО и для тех, что будут далее, найдите и изучите пример из методического пособия «Н.В.Кошуняева,Н.Н.Патронова. Теория массового обслуживания».

Система M | M | 1 | ∞

Простейшая одноканальная СМО с неограниченной очередью. На одноканальную СМО поступает простейший поток заявок с интенсивностью λ. Время обслуживания — показательное с параметром $\mu = 1 / \bar{t}_{\text{обсл}}$. Длина очереди не ограничена. Финальные вероятности существуют только при $\rho = \lambda / \mu < 1$ (при $\rho \geq 1$ очередь растет неограниченно). Состояния СМО нумеруются по числу заявок в СМО, находящихся в очереди или обслуживамых:



s_0 — СМО свободна;
s_1 — канал занят, очереди нет;
s_2 — канал занят, заявка стоит в очереди; …;
s_k — канал занят, $k - 1$ заявок стоят в очереди; … .

Финальные вероятности состояний выражаются формулами:

$$\rho_0 = 1 - \rho, \quad \rho_k = \rho^k \, (1 - \rho) \quad (k = 1, 2, \ldots),$$

где $\rho = \lambda / \mu < 1$.

Характеристики эффективности СМО:

$$A = \lambda; \; Q = 1; \; P_{\text{отк}} = 0; \quad \bar{z} = \frac{\mu}{1 - \rho}; \; \bar{r} = \frac{\rho^2}{1 - \rho}; \; \bar{t}_{\text{сист}} = \frac{\rho}{\lambda \, (1 - \rho)}; \; \bar{t}_{\text{оч}} = \frac{\rho^2}{\lambda \, (1 - \rho)};$$

среднее число занятых каналов (или вероятность того, что канал занят)

$$\bar{k} = \lambda / \mu = \rho.$$

- Заметим, что бесконечная сумма $1 + \rho + \rho^2 + \rho^3 + \ldots = \lim\limits_{k\to\infty} \dfrac{1-\rho^k}{1-\rho} = \dfrac{1}{1-\rho}$
- Средняя длина очереди легко получается из выражения

$$\overline{L}_{оч.} \equiv \overline{r} = \sum_{k=1}^{\infty} (k-1)\, p_k = \sum_{k=1}^{\infty} (k-1)\, \rho^k (1-\rho) = \rho^2 (1-\rho)\,(1 + 2\rho + 3\rho^2 + \ldots)$$

- если внимательно посмотреть на эти строки:

- $1 + \rho + \rho^2 + \rho^3 + \ldots$
- $\rho + \rho^2 + \rho^3 + \ldots = \rho\,(1 + \rho + \rho^2 + \rho^3 + \ldots)$
- $\rho^2 + \rho^3 + \ldots$
- ……….
- (подробнее см. конспект практического занятия)

А среднее число заявок в СМО $\overline{z} = \sum_{k=1}^{\infty} k\, p_k = \rho\,(1-\rho)\,(1 + 2\rho + 3\rho^2 + \ldots)$

В этой СМО в предельном стационарном режиме среднее время пребывания заявки в системе $\bar{t}_{сист}$ выражается через среднее число заявок в системе с помощью *формулы Литтла*:

$\bar{t}_{сист} = \bar{z} / \lambda$, где λ — интенсивность потока заявок. И $\bar{t}_{оч} = \bar{r} / \lambda$.

- Найдём выражение для среднего числа занятых каналов $\overline{k}$
- $\overline{k}$ = P(что канал занят) · n =

= n (t $^{обслуж.}$/ [длина интервала м\у заявками на канале]) = n · $\lambda^{\text{обслуживания на канале}}$ · t $^{обслуж.}$ =

= A t $^{обслуж.}$ = A / μ

ТЕОРЕМА (формула Литтла)

В открытой СМО среднее время нахождения заявки в системе и среднее время нахождения заявки в очереди (т.е. среднее время ожидания обслуживания) связаны со средним количеством заявок в СМО и со средней длиной очереди соответственно через интенсивность потока обслуживания (т.е. через абсолютную пропускную способность СМО, а в случае неограниченной очереди – таким же образом и через интенсивность входящего потока) следующими выражениями:

$\overline{t\ \text{оч.}} = \overline{r} / \lambda$; $\overline{t\ \text{сист.}} = \overline{z} / \lambda$ (неогр. очередь)

$\overline{t\ \text{оч.}} = \overline{r} / A$; $\overline{t\ \text{сист.}} = \overline{z} / A$ (огр. очередь)

Д-ВО

Пусть: X(t) число заявок, прибывших в СМО до момента t,

Y(t) число заявок, покинувших СМО до момента t.

И та, и другая функции являются случайными и меняются скачком (увеличиваются на единицу) в моменты приходов заявок и уходов заявок Вид функций показан на рисунке: обе линии — ступенчатые, верхняя — X(t), нижняя — Y(t). Очевидно, что для любого момента t разность Z(t) = X(t)-Y(t) есть не что иное, как число заявок, находящихся в СМО. Когда линии сливаются, в системе нет заявок.

- Рассмотрим очень большой промежуток времени T и вычислим для него среднее число заявок, находящихся в СМО. Оно будет равно:

$$\overline{z} = \frac{1}{T}\int_0^T Z(t)\,dt$$

- Но этот интеграл - не что иное, как площадь фигуры, заштрихованной на рисунке. Фигура состоит из прямоугольников с высотой, равной единице, и основанием, равным времени пребывания в системе соответствующей заявки (первой, второй и т. д.). Обозначим эти времена t_1 , t_2 , … . Тогда

$$\overline{z} = \frac{1}{T}\sum t_i \quad \rightarrow \quad \overline{z} = \frac{1}{AT}\sum t_i\, A$$

Но AT - это средн. число обслужен. за время T заявок, поэтому:

$\overline{z} = \overline{t_{\text{сист}}}\ A$ (а при неогр. очереди A = λ)

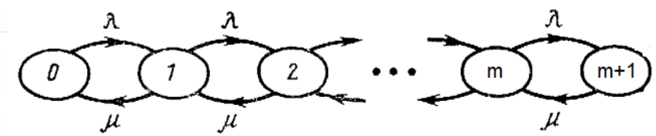
Простейшая одноканальная СМО с ограничением по длине очереди. На одноканальную СМО поступает простейший поток заявок с интенсивностью λ; время обслуживания — показательное с параметром $\mu = 1 / \bar{t}_{обсл}$. В очереди m мест. Если заявка приходит в момент, когда все эти места заняты, она получает отказ и покидает СМО. Состояния СМО:

s_0 — СМО свободна; s_1 — канал занят, очереди нет;
s_2 — канал занят, одна заявка стоит в очереди; …;
s_k — канал занят, $k-1$ заявок стоят в очереди; …;
s_{m+1} — канал занят, m заявок стоят в очереди.

Финальные вероятности состояний существуют при любом $\rho = \lambda / \mu$

$$p_0 = \frac{1-\rho}{1-\rho^{m+2}}; \quad p_k = \rho^k p_0 \ (k = 1, \ldots, m+1).$$

Характеристики эффективности СМО:



$$A = \lambda\,(1 - p_{m+1}); \quad Q = 1 - p_{m+1}; \quad P_{отк} = p_{m+1}.$$

Среднее число занятых каналов (вероятность того, что канал занят)

$\bar{k} = 1 - p_0.$ Среднее число заявок в очереди

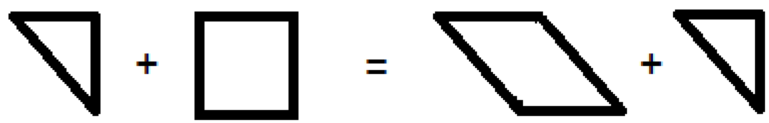
$$\bar{r} = \frac{\rho^2\,[1 - \rho^m\,(m + 1 - m\rho)]}{(1-\rho^{m+2})\,(1-\rho)}.$$

Среднее число заявок в СМО $\bar{z} = \bar{r} + \bar{k}.$

По формуле Литтла $\bar{t}_{сист} = \bar{z} / A$; $\bar{t}_{оч} = \bar{r} / A$.

- В этой задаче средняя длина очереди найдена из следующего выражения:
- $$L_{оч.} \equiv r = \sum_{k=1}^{m+1} (k-1)\, p_k = \sum_{k=1}^{m+1} (k-1)\, \rho^k \frac{1-\rho}{1-\rho^{m+2}} = \rho^2 \frac{1-\rho}{1-\rho^{m+2}} (1 + 2\rho + 3\rho^2 + \dots + m\rho^{m-1}) \qquad (*)\ ,$$

  - которое не трудно упростить в смысле практического удобства, увидев, что сумма элементов в зелёном параллелограмме из таблицы внизу равна $(1 + \rho + \rho^2 + \dots + \rho^{m-1})^2$, сумма в правом красном треугольнике в ρ^m больше, чем в левом, а сумма в квадратной части этой таблицы, остающаяся при отбрасывании левого треугольника, равна $m\rho^{m-1}(1 + \rho + \rho^2 + \dots + \rho^{m-1})$. Тогда, сопоставив, что



  - можно найти сумму в левом треугольнике, которая отличается от суммы в последней скобке выражения (*) лишь пределом суммирования.

1	ρ	ρ^2	…	ρ^{m-2}	ρ^{m-1}	ρ^m	ρ^{m+1}	…	ρ^{2m-2}
	ρ	ρ^2	…	ρ^{m-2}	ρ^{m-1}	ρ^m	ρ^{m+1}	…	ρ^{2m-2}
		ρ^2	…	ρ^{m-2}	ρ^{m-1}	ρ^m	ρ^{m+1}	…	ρ^{2m-2}
			…	…	…		…		
					ρ^{m-1}	ρ^m	ρ^{m+1}	…	ρ^{2m-2}

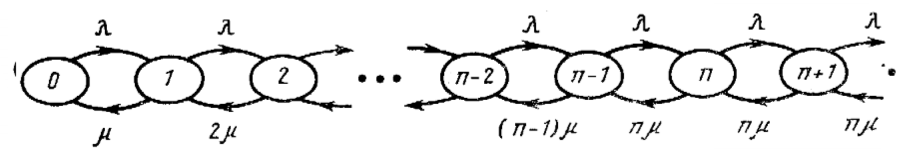
Можно найти эту сумму проще, дифференцируя степенной ряд ρ^k.

Простейшая многоканальная СМО с неограниченной очередью. На n-канальную СМО поступает простейший поток заявок с интенсивностью λ; время обслуживания одной заявки — показательное с параметром $\mu = 1/\bar{t}_{\text{обсл}}$. Финальные вероятности существуют только при $\rho/n = \chi < 1$, где $\rho = \lambda/\mu$. Состояния СМО нумеруются по числу заявок в СМО:

$$\left.\begin{array}{l} s_0 - \text{СМО свободна;} \\ s_1 - \text{занят один канал; \ldots;} \\ s_k - \text{занято } k \text{ каналов } (1 \le k \le n);\ \ldots; \\ s_n - \text{заняты все } n \text{ каналов;} \end{array}\right\} \text{очереди нет;}$$

s_{n+1} — заняты все n каналов, одна заявка стоит в очереди; …;
s_{n+r} — заняты все n каналов, r заявок стоят в очереди; … .
Финальные вероятности состояний выражаются формулами:

$$p_0 = \left\{1 + \frac{\rho}{1!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n \cdot n!}\frac{1}{1-\chi}\right\}^{-1}$$



$$p_k = \frac{\rho^k}{k!}p_0 \ (1 \le k \le n); \quad p_{n+r} = \frac{\rho^{n+r}}{n^r \cdot n!}p_0 = \chi^r p_n \ (r = 1, 2, \ldots).$$

Характеристики эффективности СМО:

$$\bar{r} = \rho^{n+1}p_0 / [n \cdot n!\,(1-\chi)^2] = \chi\, p_n / (1-\chi)^2; \qquad \bar{k} = \rho$$

$$\bar{z} = \bar{r} + \bar{k} = \bar{r} + \rho; \qquad \bar{t}_{\text{сист}} = \bar{z}/\lambda; \quad \bar{t}_{\text{оч}} = \bar{r}/\lambda.$$

Простейшая многоканальная СМО с ограничением по длине очереди. Условия и нумерация состояний те же, что и с неогр.оч, с той разницей, что число m мест в очереди ограничено. Финальные вероятности состояний существуют при любых λ и μ и выражаются формулами:

$$p_0 = \left\{1 + \frac{\rho}{1!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n \cdot n!}\,\frac{1-\chi^m}{1-\chi}\right\}^{-1}$$

когда $\chi = 1$

вместо $\frac{1-\chi^m}{1-\chi}$ будет $(1 + \chi + \chi^2 + \ldots + \chi^{m-1})$

$$p_k = \frac{\rho^k}{k!} p_0 \ (1 \le k \le n); \qquad p_{n+r} = \frac{\rho^{n+r}}{n^r \cdot n!} p_0 = \chi^r p_n \ (1 \le r \le m),$$

где $\chi = \rho / n = \lambda / (n\mu)$.

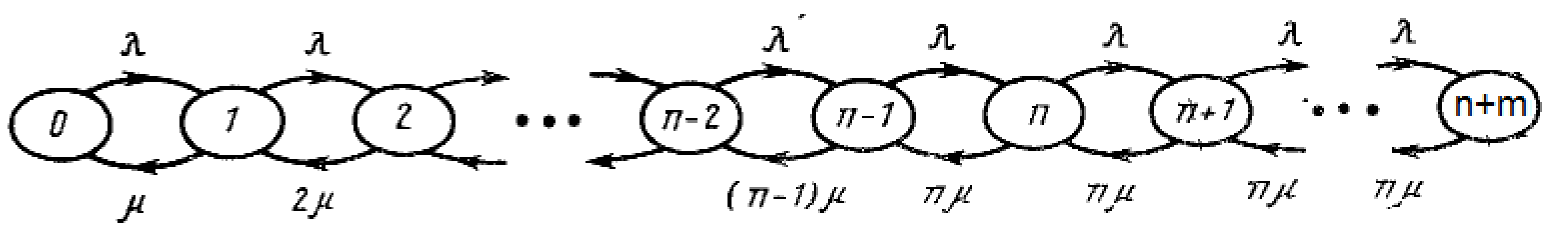
Характеристики эффективности СМО:

$$A = \lambda(1 - p_{n+m}); \quad Q = 1 - p_{n+m}; \quad P_{\text{отк}} = p_{n+m}; \quad \bar{k} = \rho(1 - p_{n+m});$$

$$\bar{r} = \frac{\rho^{n+1} p_0}{n \cdot n!}\,\frac{1 - (m+1)\chi^m + m\chi^{m+1}}{(1-\chi)^2};$$

$$\bar{z} = \bar{r} + \bar{k};$$

$$\bar{t}_{\text{оч}} = \bar{r} / A \qquad \bar{t}_{\text{сист}} = \bar{z} / A$$



Пример

Автозаправочная станция (АЗС) имеет две колонки ($n = 2$); площадка возле нее допускает одновременное ожидание не более четырех автомобилей ($m = 4$). Поток автомобилей, прибывающих на станцию, простейший с интенсивностью $\lambda = 1$ авт/мин. Время обслуживания автомобиля — показательное со средним значением $\bar{t}_{\text{обсл}} = 2$ мин. Найти финальные вероятности состояний АЗС и ее характеристики: A, Q, $P_{\text{отк}}$, $\bar{k}$, $\bar{z}$, $\bar{r}$, $\bar{t}_{\text{сист}}$, $\bar{t}_{\text{оч}}$.

Решение. $\lambda = 1$, $\mu = 1/2 = 0{,}5$; $\rho = 2$; $\chi = \rho / n = 1$.

$$p_0 = \left\{1 + 2 + \frac{2^2}{2!} + \frac{2^2}{2!} \cdot 4\right\}^{-1} = \frac{1}{13}; \; p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{2}{13};$$

$$P_{\text{отк}} = 2/13; \quad Q = 1 - P_{\text{отк}} = 11/13;$$

$$A = \lambda Q = 11/13 \approx 0{,}85 \text{ авт/мин};$$

$$\bar{k} = A/\mu = 22/13 \approx 1{,}69 \text{ колонки};$$

$$\bar{r} = \frac{2^2}{2!} \frac{4\,(4+1)}{2} \cdot \frac{1}{13} \approx 1{,}54 \text{ автомобиля};$$

$$\bar{z} = \bar{r} + k \approx 3{,}23 \text{ автомобиля}.$$

С простейшими замкнутыми СМО и системами с приоритетом ознакомимся вкратце.

- В *замкнутой СМО* поток требований находится внутри системы и полностью зависит от состояния системы (в противоположность открытой СМО) .
- Пример:
- Компьютерный класс, или вычислительная система, в которой узлы (компьютеры, устройства) периодически выходят из строя (отказ) и требуют восстановления: ремонта или переустановки программного обеспечения, и таким образом создают заявки на обслуживание.

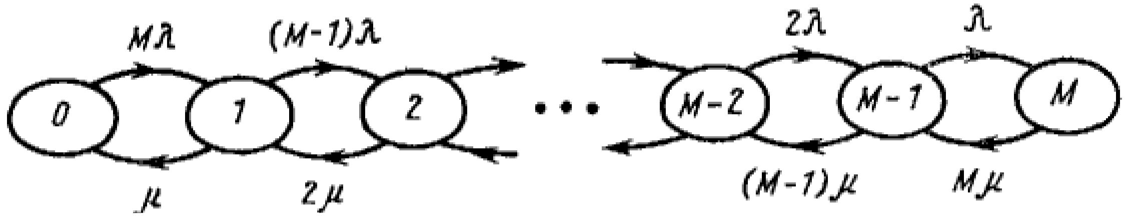
- Для таких замкнутых СМО в литературе можно встретить такое название, как например «СМО с конечным числом источников нагрузки» (или же с «ограничен-ным числом [заявок]или[источников заявок]»), что означает, что каждый источник создаёт новую только если выполнены все предыдущие.

- Дадим обзор простейших моделей замкнутых СМО, в которых потоки отказов устройств будем рассматривать простейшими, а время восстановления каждого устройства - экспоненциальным.

Задачу про компьютерный класс(или про выч.систему) можно рассмотреть в нескольких вариациях (все они «с неограниченной очередью»):

- устройства восстанавливаются по одному;
  (замкнутая одноканальная СМО с неограниченной очередью)
- могут восстанавливаться одновременно сразу все отказавшие устройства;
  (замкнутая СМО моментального обслуживания (или же: «бесконечноканальная»))
- число одновременно восстанавливающихся устройств ограничено
  (замкнутая многоканальная СМО с неограниченной очередью)

- **Простейшая замкнутая СМО моментального обслуживания** (когда ремонтируем одновременно все узлы-устройства)



$$\lambda_k = \begin{cases} \lambda(M-k), & 0 \leqslant k \leqslant M; \\ 0 \text{ в остальных случаях;} \end{cases}$$

$$\mu_k = k\mu, \quad k = 1, 2, 3, \ldots$$

(Процесс гибели и размножения с такими интенсивностями)

M – количество узлов системы

$$p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda(M-i)}{(i+1)\mu} = p_0 \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k}, \quad 0 \leqslant k \leqslant M,$$

$$p_0 = \left[\sum_{k=0}^{M} \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k}\right]^{-1} = \frac{1}{(1+\lambda/\mu)^M}$$

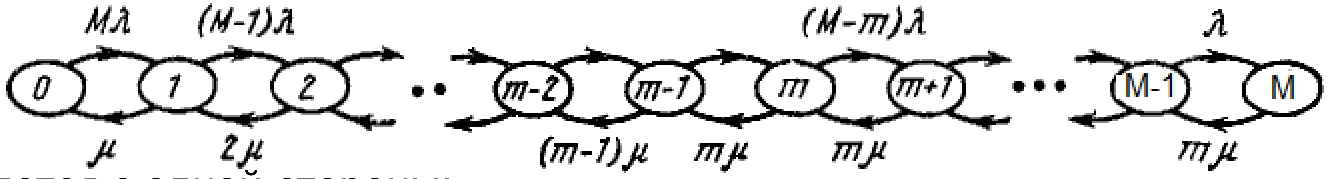
Таким образом,

$$p_k = \begin{cases} \dfrac{\left(\frac{\lambda}{\mu}\right)^k \binom{M}{k}}{(1+\lambda/\mu)^M}, & 0 \leqslant k \leqslant M; \\ 0 \text{ в остальных случаях.} \end{cases}$$

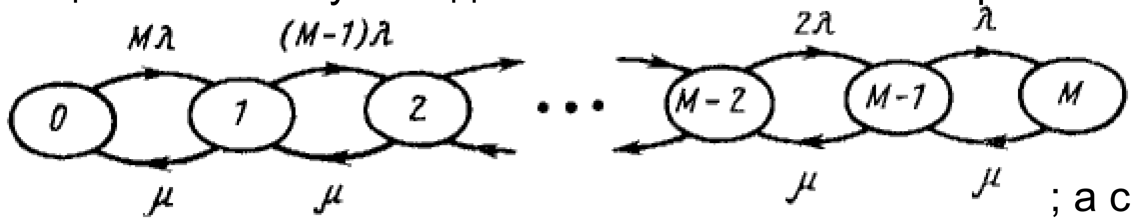
Важнейшей характеристикой эффективности будет среднее число работающих узлов системы:

$$\overline{z} = \sum_{k=0}^{M} k p_k = \frac{\sum_{k=0}^{M} k \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k}}{(1+\lambda/\mu)^M} = \frac{\rho\left((1+\rho)^M\right)'_\rho}{(1+\rho)^M} = \frac{M\lambda/\mu}{1+\lambda/\mu}$$

- **Простейшая замкнутая многоканальная СМО с неограниченной очередью** (для задачи об отказах и восстановлении узлов выч.системы – случай когда можно восстанавливать одновременно ограниченное число узлов) моделируется процессом Маркова вида

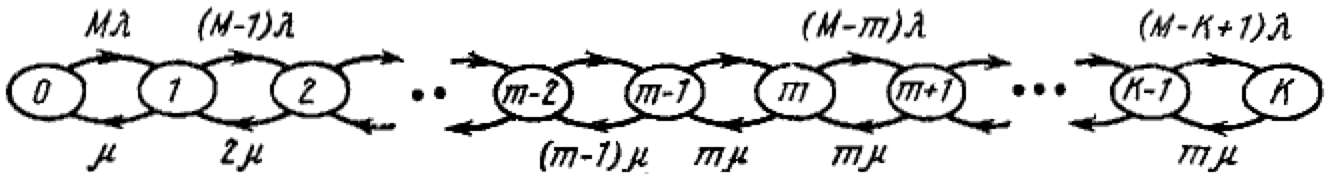


- и является с одной стороны:
  1) обобщением замкнутой одноканальной СМО с неограниченной очередью:



- ; а с другой стороны

2)частным случаем **простейшей замкнутой многоканальн. СМО с огранич.очередью**:



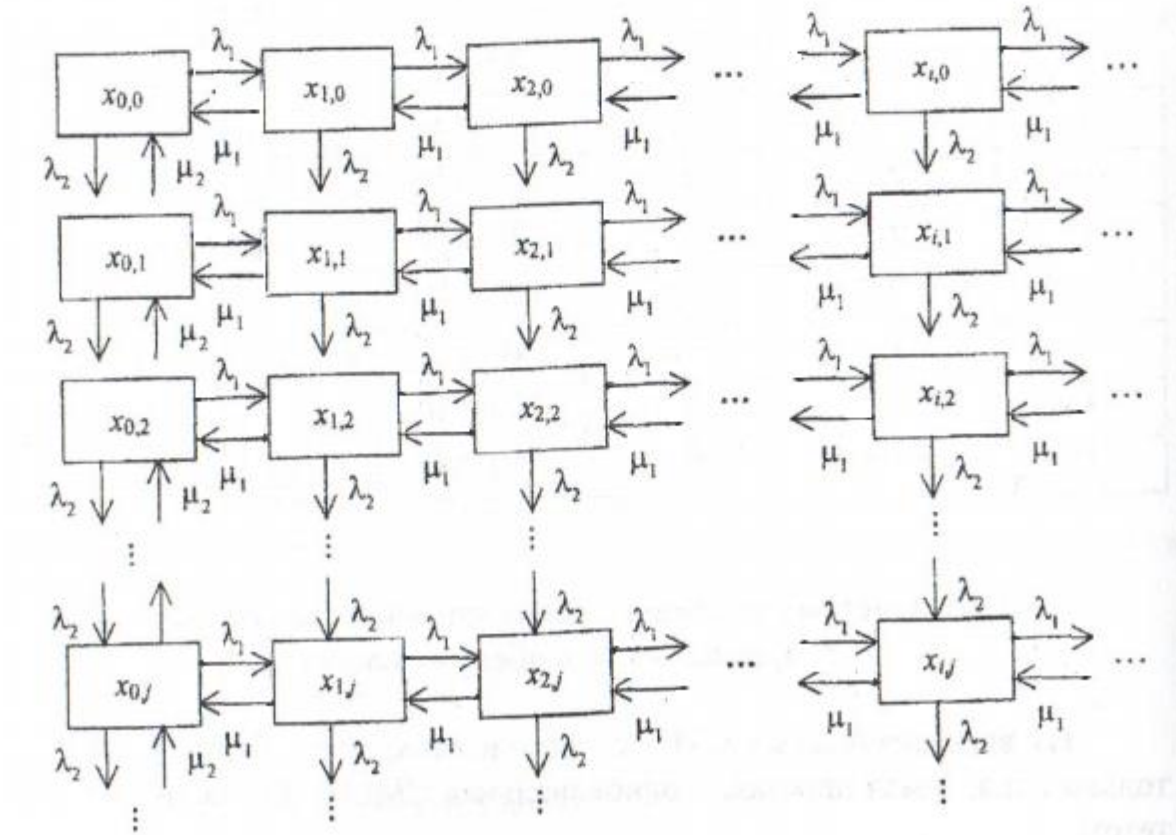
- $$p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda (M-i)}{(i+1)\mu} = p_0 \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k}, \quad 0 \leqslant k \leqslant m-1.$$

В области $m \leqslant k \leqslant K$ имеем $p_k = p_0 \prod_{i=0}^{m-1} \frac{\lambda (M-i)}{(i+1)\mu} \prod_{i=m}^{k-1} \frac{\lambda (M-i)}{m\mu} =$

$$= p_0 \left(\frac{\lambda}{\mu}\right)^k \binom{M}{k} \frac{k!}{m!} m^{m-k}, \quad m \leqslant k \leqslant K.$$

Формула для p_0 получается весьма сложной и здесь не приводится, хотя она может быть получена непосредственными вычислениями.

Схема процесса Маркова для **простейшей одноканальной СМО с приоритетами и неограниченной очер**.



Первый номер состояния – к-во приоритетных заявок в очереди, а второй – неприоритетных.

λ_1 – интенсивность потока приоритетных заявок;

λ_2 – интенсивность потока неприоритетных заявок

μ_1 и μ_2– параметры экспоненциальных распределений времени обслуживания приоритетной и неприоритетной заявки соответственно

Система Колмогорова:

$$\frac{dP_{0,0}(t)}{dt} = -(\lambda_1+\lambda_2)P_{0,0}(t)+\mu_1P_{1,0}(t)+\mu_2P_{0,1}(t),$$

$$\frac{dP_{0,j}(t)}{dt} = -(\lambda_1+\lambda_2+\mu_2)P_{0,j}(t)+\lambda_2P_{0,j-1}(t)+\mu_2P_{0,j+1}(t)+\mu_1P_{1,j}(t), \quad j>0,$$

....

$$\frac{dP_{i,0}(t)}{dt} = -(\lambda_1+\lambda_2+\mu_1)P_{i,0}(t)+\lambda_1P_{i-1,0}(t)+\mu_1P_{i+1,0}(t), \quad i>0,$$

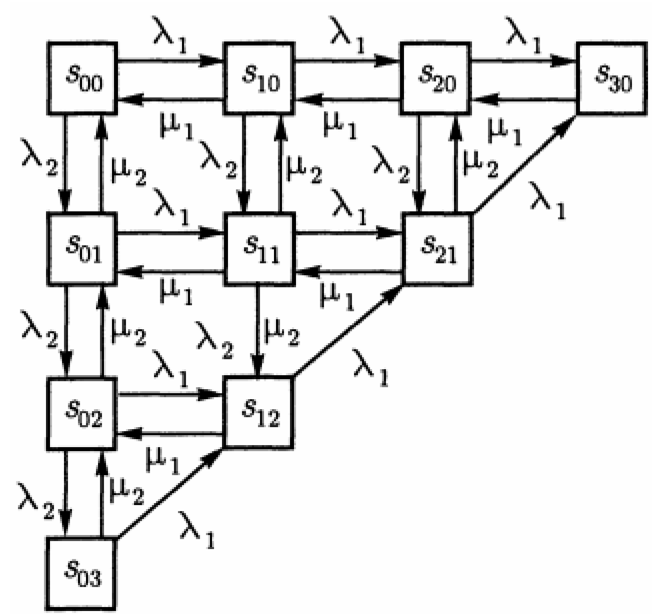
$$\frac{dP_{i,j}(t)}{dt} = -(\lambda_1+\lambda_2+\mu_1)P_{i,j}(t)+\lambda_2P_{i,j-1}(t)+\lambda_1P_{i-1,j}(t)+\mu_1P_{i+1,j}(t), \quad i>0,\ j>0.$$

....

при начальных условиях $P_{0,0}(0)=1;\ P_{i,j}(0)=0$ (при $i\neq 0$ или $j\neq 0$)

- **Схема для простейшей одноканальной СМО с приоритетами и ограниченной очередью**

- (в этой задаче приоритетные заявки вытесняют неприоритетные из очереди, если она была полностью заполнена)



- **Схема для простейшей многоканальной СМО с приоритетами и отказами**

  (в этой задаче приоритетные заявки вытесняют неприоритетные с устройств обслуживания, если все они заняты)