Министерство образования Республики Беларусь

Учреждение образования
«Гомельский государственный технический
университет имени П. О. Сухого»

Кафедра «Технология машиностроения»

В. С. Мурашко

МЕТОДЫ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В МАШИНОСТРОЕНИИ

ПОСОБИЕ
по курсу «Математическое моделирование и методы исследования операций» для студентов специальности 1-53 01 01 «Автоматизация технологических процессов и производств (по направлениям)» дневной формы обучения

Гомель 2018

УДК 519.8(075.8)
ББК 22.18я73
М91

Рекомендовано научно-методическим советом машиностроительного факультета ГГТУ им. П. О. Сухого (*протокол № 9 от 07.05.2018 г.*)

Рецензент: доц. каф. «Информационные технологии» ГГТУ им. П. О. Сухого канд. техн. наук, доц. *В. В. Комраков*

Мурашко, В. С.
М91 Методы исследования операций в машиностроении : пособие по курсу «Математическое моделирование и методы исследования операций» для студентов специальности 1-53 01 01 «Автоматизация технологических процессов и производств (по направлениям)» днев. формы обучения / В. С. Мурашко. – Гомель : ГГТУ им. П. О. Сухого, 2018. – 115 с. – Систем. требования: PC не ниже Intel Celeron 300 МГц ; 32 Mb RAM ; свободное место на HDD 16 Mb ; Windows 98 и выше ; Adobe Acrobat Reader. – Режим доступа: https://elib.gstu.by. – Загл. с титул. экрана.

Рассмотрены вопросы разработки и использования математических моделей для описания, исследования и оптимизации процессов в машиностроении.

Для студентов специальности 1-53 01 01 «Автоматизация технологических процессов и производств (по направлениям)» дневной формы обучения.

УДК 519.8(075.8)
ББК 22.18я73

Введение

Известно, что моделирование (в широком смысле) является основным методом исследований во всех областях знаний и научно обоснованным методом оценок характеристик сложных систем, используемых в различных сферах инженерной деятельности, в частности в машиностроении.

В настоящее время в машиностроении интенсификация процессов создания новых конкурентоспособных изделий требует сокращения сроков и повышения качества проектно-конструкторских работ. Эти требования можно обеспечить только применяя новые технологии проектирования, основанные на использовании методов математического моделирования и вычислительной техники. Современные технологии основываются как на опыте инженерной практики, так и на научных теоретических и экспериментальных исследованиях. Поэтому инженер должен уметь практически решать задачи, требующие применения современных математических методов.

Настоящее пособие предназначено для студентов специальности 1-53 01 01 «Автоматизация технологических процессов и производств (по направлениям» и соответствует программе курса «Математическое моделирование и методы исследования операций» и является продолжением пособия «Математическое моделирование технологических задач».

В настоящем пособии рассматриваются методы оптимизации: классические методы оптимизации, методы поиска экстремума унимодальных функций, методы направленного и случайного поиска; экспериментальные методы построения математических моделей и технических систем. В пособии также уделяется особое внимание на использование матричных игр и игр с природой, систем массового обслуживания при решении технологических задач.

1 ЗАДАЧИ И МЕТОДЫ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

Разнообразие практических задач исследования операций настолько велико, что для их решения требуются различные методы исследования: аналитические, численные и методы случайного поиска. Каждый из методов исследования включает большое число методов оптимизации, которые можно разделить на две группы: точные методы, обеспечивающие нахождение оптимума за конечное число шагов, и приближенные, приводящие за конечное число шагов к результату, незначительно отличающемуся от оптимального.

1.1 Цель и задачи исследования операций

Исследование операций – это прикладное направление кибернетики, изучающее способы совершенствования и повышения эффективности организации, планирования и управления в различных системах на основе количественных методов.

Цель исследования операций – определить, как необходимо организовать работу системы (объекта, процесса), чтобы ее (его) функционирование обеспечивало максимальный эффект, выработать соответствующие научно обоснованные решения и рекомендации по оптимальному функционированию системы на основе использования практически всех существующих методов анализа и синтеза (оптимизации) систем и ЭВМ.

Система – это совокупность взаимосвязанных, взаимодействующих элементов (людей, машин, ...), выполняющих определенную задачу.

Термин «исследование операций» появился в период второй мировой войны, когда для решения военно-технических задач создавались смешанные группы ученых разных специальностей.

Дадим несколько определений исследования операций, приведенных в различных работах:

- это методология научного поиска и вместе с тем инструмент для решения важных практических задач
- это научный подход к решению задач организационного управления, который предполагает:

1. построение математических, экономических или статистических моделей для задач принятия решений и управления в сложных ситуациях или в условиях неопределенности;

2. изучение взаимосвязей, определяющих возможные последствия принимаемых решений, а также установление критериев эффективности, позволяющих оценивать относительное преимущество того или иного варианта действий;

- это применение математических, количественных методов для обоснования решения во всех областях целенаправленной человеческой деятельности;
- это математические методы, позволяющие установить закономерности и оценить ожидаемую эффективность процессов, протекающих в производственной, экономической, военных сферах, и получить рекомендации, которые учитываются при обосновании решения на управление этими процессами;
- это научная дисциплина, которая наблюдает реальные явления, связанные с функциональными системами, разрабатывает модели, предназначенные для объяснения данных явлений, использует эти модели для описания того, что произойдет при изменении условий, и проверяет предсказания новыми наблюдениями.

Ни одно из приведенных определений, по-видимому, нельзя считать полным, исчерпывающим. Каждое из них раскрывает ту или иную сторону, тот или иной аспект исследования операций – этой новой, быстро прогрессирующей отрасли знаний. Точнее было бы связать все определения союзом «и».

Несмотря на короткую историю становления исследования операций, в настоящее время уже выделился определенный класс задач, эффективно решаемых в рамках исследования операций: распределительные; управления запасами; замены оборудования; упорядочения и согласования; выбора оптимальных режимов движения; массового обслуживания; состязательные; поиска и др.

Решение задач исследования операций предполагает выполнение следующих основных этапов:

- постановки задачи и выбора критерия оптимизации;
- выявления основных особенностей, взаимосвязей и количественных закономерностей исследуемой системы;
- построения математической модели системы (процесса);
- исследования математической модели с помощью специализированных алгоритмов и программ, поиска оптимальных параметров системы (процесса, операции).

Каждый из этих этапов важен в процессе решения задач исследования операций. Так, если неправильно поставлена задача или выбран не тот критерий оптимизации, то, как бы точно ни была построена математическая модель, какой бы эффективный алгоритм исследования или ЭВМ ни использовались, решение будет неточным. В то же самое время при правильно поставленной задаче, выбранном критерии оптимизации, адекватно построенной математической модели, если использован не тот алгоритм исследования, решение будет также неточным.

Исследование математической модели может быть произведено с использованием различных методов. Применение того или иного метода исследования зависит от вида математической модели, ее размерности и других факторов.

Под математической моделью понимается совокупность соотношений (например, формул, уравнений, неравенств, логических условий, операторов и т. д.), описывающих исследуемую систему, ограничения, критерий оптимизации (целевую функцию).

В практике решения задач исследования операций используются различные методы исследования математической модели. При этом ни один из них не является универсальным с точки зрения безусловной применимости к решению той или иной задачи исследования операций.

Выделяют следующие основные методы исследования математической модели:

- аналитический;
- исследование системы (процесса) с помощью численных методов и ЭВМ;
- исследование системы (процесса) методами случайного поиска.

Аналитический метод, как правило, дает наглядную картину исследуемой системы (процесса) и характеризующих ее (его) параметров. Однако построение математической модели, удобной для аналитического исследования, – трудная задача, но, несмотря это, данный метод широко используется при решении многих практических задач исследования операций. Искусство исследователя как раз и состоит в том, чтобы в процессе построения математической модели обеспечить возможность исследования ее аналитическими методами (метод дифференциального исчисления, метод множителей Лагранжа и т. д.). Если аналитическое исследование затруднительно, то используют другие способы.

Исследование с помощью численных методов и ЭВМ менее наглядно по сравнению с аналитическим, но класс моделей, пригодных для исследования численными методами, значительно шире. Результаты исследования (таблицы, графики и т. д.) менее компактны по сравнению с аналитическим исследованием и требуют большой вычислительной работы при наличии многих переменных и изменяемых условий функционирования исследуемой системы (процесса). В настоящее время этот метод находит все более широкое применение на практике в связи с интенсивным внедрением современных ЭВМ и соответствующего программного обеспечения.

Среди численных методов наибольшее распространение получили:

- для унимодальных функций одной переменной методы дихотомии, Фибоначчи и золотого сечения;
- для функций нескольких переменных методы поочередного изменения параметров, градиента, наискорейшего спуска, математического программирования и др.

Исследование систем методами случайного поиска предполагает воспроизведение (имитацию) происходящих явлений с сохранением их логической структуры и расположения по времени с намеренным использованием случайных величин и процессов. Моделирование процесса функционирования системы (процесса) производится на быстродействующих ЭВМ по специально разработанным программам.

К методам случайного поиска (статистическим методам) относятся

- ненаправленный случайный поиск (метод Монте-Карло);
- направленный случайный поиск без самообучения (поиск с парными пробами, линейный поиск, нелинейный поиск и т. д.);
- направленный поиск с самообучением.

Повышение эффективности использования методов исследования операций неразрывно связано с широким использованием ЭВМ, которые требуют наличия соответствующих программ, комплекса программ.

1.2 **Классические методы оптимизации**

1.2.1 Метод прямого перебора. Если известна функциональная связь целевой функции Y и искомой переменной X, то можно последовательно вычислить значения целевой функции для некоторых значений искомой переменной.

Вычисления повторяются до тех пор, пока не будет найдено минимальное (максимальное) значения целевой функции

$$Y = f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m),$$

$$x_i = x_{0i} + \Delta x_i k \ (k = 0, 1, .., i)$$

Этот метод может быть использован для решения задач исследования операций, если имеются одна искомая переменная или несколько с небольшим диапазоном изменения искомых переменных.

Особенность и преимущества метода прямого перебора заключаются:

- в независимости поиска от вида и характера целевой функции;
- в цикличности поисковой процедуры;
- в определении глобального экстремума целевой функции;
- в простоте алгоритма и программы оптимизации;
- в малом объеме необходимо машинной памяти.

Главным недостатком метода является продолжительное время работы ЭВМ. В случае большой области изменения искомой переменной и (или) наличия более чем одного экстремума исследуемой функции использование этого метода неэффективно.

1.2.2 Классический метод дифференциального исчисления. Если известна функциональная связь целевой функции с искомыми переменными $Y = f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m)$, которая обладает непрерывными первыми частными производными, то, определив частные производные по своим искомым переменным, приравняв частные производные от Y по x_i нулю и решив совместно систему уравнений

$$\frac{\partial Y}{\partial x_i} = 0, (i = 1, .., n), \qquad (1.1)$$

найдем значения x_{io}, дающие стационарные значения целевой функции, среди которых находятся оптимальные.

На первый взгляд кажется, что использование этого метода позволяет достаточно просто решать задачу определения оптимума нелинейной функции нескольких переменных. Однако это не так.

Существует ряд трудностей при его реализации и ограничений по применению:

- при большом числе оптимизируемых параметров рассматриваемый метод становится весьма сложным в части решения системы уравнений

$$\frac{\partial Y}{\partial x_i} = 0, (i = 1, .., n);$$

- условие определения экстремума, выраженное системой уравнений (1.1), является необходимым, но не достаточным для решения задачи. Так как выражения (1.1) определяют положение стационарных точек внутри области, среди которых, кроме экстремальных, могут быть особые точки типа «седла», учет достаточных условий нахождения экстремумов функции многих переменных является сложным как в алгоритмическом, так и в вычислительном плане. Так, достаточное условие существования минимума функции одной переменной – это положительная величина производной четного порядка $\frac{\partial^{2k} Y}{\partial x} > 0, (k = 1, .., n)$. При наличии функции двух переменных достаточным условием существования минимума функции является положительная определенность матрицы

$$\begin{vmatrix} \frac{\partial^2 Y}{\partial^2 x^2{}_1} & \frac{\partial^2 Y}{\partial x_1 \partial x_2} \\ \frac{\partial^2 Y}{\partial x_1 \partial x_2} & \frac{\partial^2 Y}{\partial^2 x_2^2} \end{vmatrix};$$

- рассматриваемый метод дает возможность найти экстремум только в том случае, если он лежит внутри, а не на границе области возможных значений переменных, следовательно, требуется дополнительный анализ значений функции $f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m)$ на границах допустимой области изменения $(x_1, ..., x_i, ..., x_n)$;
- оптимизируемая функция $f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m)$ должна быть непрерывной и иметь первые и вторые производные по оптимизируемым параметрам;
- необходимо, чтобы оптимизируемые параметры, определяющие значения минимума или максимума функции, были независимы, т. е. не должно быть дополнительных уравнений, связывающих между собой часть параметров.

1.2.3 Метод множителей Лагранжа. С помощью этого метода можно определить экстремальные точки функции многих переменных при наличии дополнительных связей между оптимизируемыми параметрами.

Пусть имеется целевая функция $Y = f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m)$, экстремум которой требуется найти, причем существуют дополнительные условия

$$\varphi_k(x_1, ..., x_i, ..., x_n, u_1, ..., u_m) = 0, \ (k = 1, 2, .., p).$$

Введя p дополнительных множителей λ_1, λ_2,.., λ_p построим новую

$$L(x_1, ..., x_i, ..., x_n, u_1, ..., u_m, \lambda_1, ..\lambda_p) =$$
$$f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m) - \sum \lambda_k f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m),$$

где λ_k множитель Лагранжа.

Необходимые условия экстремума состоят в равенстве нулю всех первых частных производных от L по x_1, …,x_i, ..,x_n, λ_1, ..,λ_k,.., λ_p :

$$\frac{\partial L}{\partial x_i} = 0, (i = 1, .., n)$$
$$\frac{\partial L}{\partial \lambda_k} = 0, (k = 1, .., p).$$

В результате получается $n+p$ уравнений с $n+p$ неизвестными (x_1, …,x_i, ..,x_n), (λ_1, ..,λ_k,.., λ_p). Решение этих уравнений относительно переменных x и λ дает возможность определить положение стационарной точки.

Использование вспомогательной функции $L(x, \lambda)$ позволяет заменить задачу с дополнительными условиями задачей без них.

Недостатком метода множителей Лагранжа является введение p дополнительных переменных, которые должны быть исключены с помощью p дополнительных уравнений. Кроме того, для этого метода сохраняют свою силу недостатки и трудности классического метода дифференциального исчисления.

Принципиальным недостатком метода множителей Лагранжа является невозможность решения с его помощью задач, имеющих ограничения в форме неравенств. Нелинейное программирование как новая математическая дисциплина возникла главным образом в связи с указанной ограниченностью метода множителей Лагранжа.

Выбор того или иного метода решения экстремальных задач исследования операций зависит в основном от следующих факторов:

- от принадлежности задачи исследования операций к тому или иному классу;
- от способа задания критерия оптимизации;
- от линейности или нелинейности математической модели решаемой задачи;
- от времени и средств, имеющихся в распоряжении ответственного за выработку оптимальных решений, отличных склонностей и уровня подготовки вырабатывающего оптимальное решение;
- от качества и количества информации об объекте оптимизации.

1.3 Методы поиска экстремума унимодальных функций

Функция одной переменной, имеющая в интервале исследования горб (впадину), называется унимодальной, или: функция Y – унимодальная,

если $x_1<x_{0пт}$ ($x_1> x_{0пт}$), $x_2<x_{0пт}$ ($x_2>x_{0пт}$) и $x_1<x_2$, то $Y(x_1)<Y(x_2)$ ($Y(x_1)>Y(x_2)$) [18].

Унимодальная функция не обязательно должна быть гладкой (рис. 1.1, а) и даже непрерывной (рис. 1.1, б) – она может быть изломанной (недифференцируемой), разрывной и даже может в некоторых точках интервала быть неопределенной.

Предположение унимодальности не связано с жесткими ограничениями и выполняется во многих практических задачах поиска оптимума. Предыдущие методы нельзя применять к функциям с изломами и к разрывным функциям.

Если целевая функция унимодальна, то можно сузить интервал исследования функции на оптимум путем определения значений целевой функции в двух точках интервала задания функций $Y(x_1)$ $Y(x_2)$ и последующего поинтервального сравнения. При этом возможны три случая (рис. 1.2):

- если $Y_1>Y_2$, то $x_{0пт}<x_2$, т. е. оптимум не может лежать правее, иначе нарушается предположение об унимодальности функции (интервал $[x_2, x]$ из дальнейшего рассмотрения исключается);
- если $Y_1<Y_2$, то $x_{0пт} > x_1$;
- если $Y_1 = Y_2$, то $x_1< x_{0пт} <x_2$.

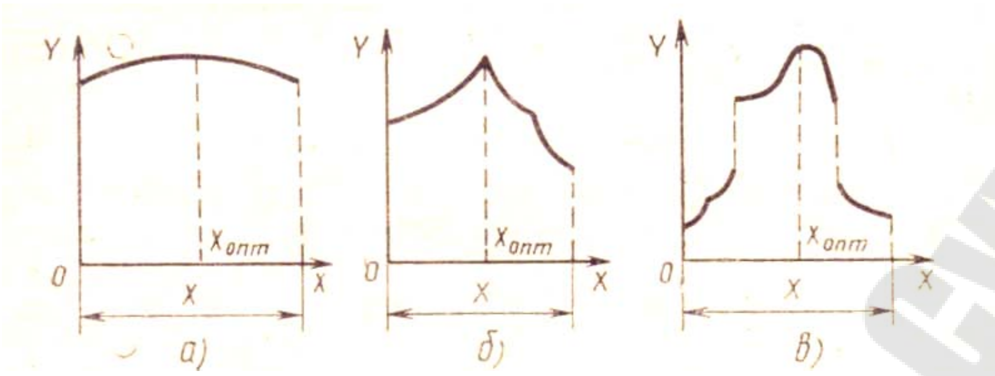

Рисунок 1.1 – Разновидности унимодальной функции:
а –выпуклая; б – непрерывная; в – произвольная

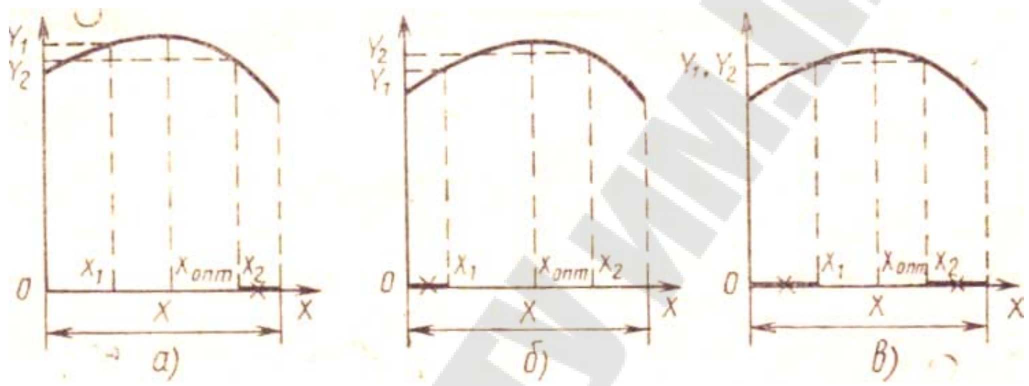

Рисунок 1.2 – Возможные случаи сужения интервала исследования

Как правило, задачи исследования операций имеют унимодальную целевую функцию. Последовательно сужая интервал исследования, в котором находится оптимальное значение искомой управляющей переменной, можно с достаточной степенью точности найти оптимальное значение искомой переменной. Для этого необходимо выработать такую стратегию поиска, чтобы за заданное число шагов (этапов) определить минимальный интервал, в котором лежит искомый оптимум, или свести исходный интервал до области заданной длины за минимальное число шагов (расчетов).

К последовательным детерминированным методам поиска экстремума унимодальных функций (учитывающим результаты предыдущих шагов) относятся методы:

- дихотомии;
- Фибоначчи;
- золотого сечения.

1.3.1 Метод дихотомии (половинное деление). В этом методе (рис. 1.3) искомая длина интервала исследования x_N, в котором лежит искомый оптимум, уменьшается с каждым шагом N почти в два раза.

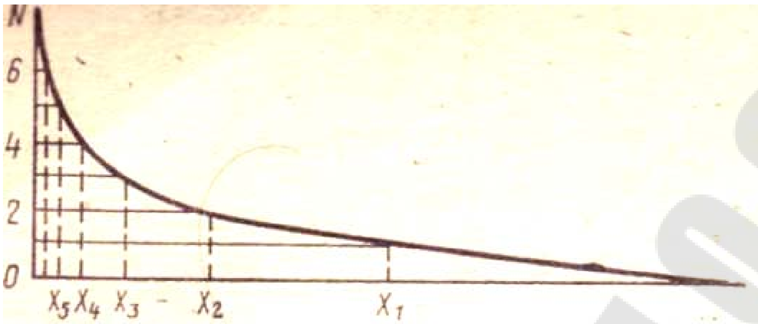

Рисунок 1.3 – Зависимость длины интервала от числа шагов поиска

Алгоритм метода состоит из следующих этапов:

- делят исходный интервал исследования пополам;
- вблизи точки деления (по разные ее стороны) подсчитывается дважды значение целевой функции $Y(x'_{N+1})$, $Y(x''_{N+1})$:

$$x'_{N+1}=x_{N+1}/2-\Delta x/2;\ x''_{N+1}=x_{N+1}/2+\Delta x/2,$$

где Δx – наименьший интервал изменения управляющей переменной, при котором становится возможным обнаружить оптимум между $Y(x'_{N+1})$ и $Y(x''_{N+1})$;

- используя свойство унимодальных функций, определяют интервал, в котором находится экстремальное значение целевой функции, за тем процесс расчета повторяется по аналогичной схеме до тех-пор, пока не будет найден $x_{опт}$.

В отличие от метода прямого перебора, в котором эффективность поиска прямо пропорциональна числу расчетов, в методе дихотомии эффективность возрастает с ростом N экспоненциально. Так, если для определения оптимального значения искомой переменной методом прямого перебора требуется около 1000 вычислений, то методом дихотомии – 20.

Интервал, в котором находится оптимум, после N шагов определяется по формуле:

$$x_N=\left[2^{-N/2}+\left(1-2^{-N/2}\right)\cdot\Delta x\right]\cdot x_0,$$

где x_0 – первоначальный интервал исследования; N – число шагов; $x_0/x_N\approx 2^{N/2}$.

1.3.2 Метод Фибоначчи. Как ни эффективен метод дихотомии, существует еще более совершенный. В методе Фибоначчи точка деление интервала исследования определяется с каждым новым расчетов (в методе дихотомии необходимо на каждом шаге выполнять два расчета). В интервал исследования попадает предыдущий расчет, для продолжения поиска достаточно лишь произвести расчет симметрично имеющемуся.

Пусть задано число расчетов (шагов) N. Необходимо их так произвести, чтобы интервал, в котором лежит оптимум, был минимальным. Для этих целей наиболее подходят числа Фибоначи:

$$F_N=F_{N-1}+F_{N-2},\ F_0=F_1=1.$$

Алгоритм метода состоит из следующих этапов.

1. Изменяют масштаб исходного интервала, в котором лежит оптимум. В качестве единицы измерения принимают $l=x_0/F_N$ или если задана длина интервала l, в котором лежит оптимум, находят его на исходном интервале длиной x_0. Для этого, разделив x_0 на l находят близлежащее большее число Фибоначчи F_N, а по нему определяют N – число необходимых расчетов для определения интервала.
2. Расставляют первые две точки x'_1 и x''_1 на интервале исследования x_0 на расстоянии F_{N-2} от концов;
3. Вычисляют значение целевой функции в этих точках для сужения интервала исследования. Пусть $Y'_1>Y''_1$ тогда интервал $[x'_1, F_N]$ исключается из рассмотрения.
4. На новом интервале исследования снова расставляют две точки x'_2 и x''_2, но в одной из них уже известно значение целевой функции $Y''_2=Y'_1$.
5. Переходят к этапу 3 и так далее, пока не достигают искомого интервала, в котором лежит значение переменной, максимизирующее ее целевую функцию.

На рис. 1.4 видно, что интервал, в котором находится оптимум сокращается до $x''_1=F_{N-1} = F_5$. Затем интервал, в котором находится оптимум, сократится до $x''_2=x'_1$ при $Y'_2> Y''_2$ и т. д. Последний N-й расчет определяет интервал длиной l, в котором находится экстремум целевой функции. Если при методе прямого счета требуется около 1000 вычислений, то при методе дихотомии – 20,. методе Фибоначчи – 15. Грубо говоря, $x_0/x_N\approx F_N$.

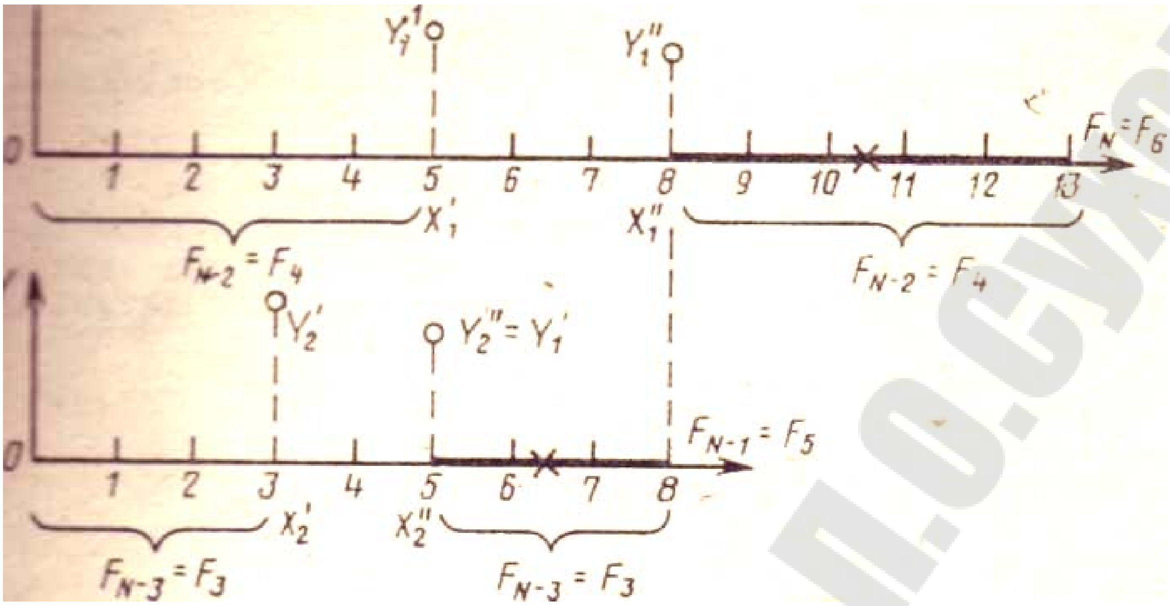

Рисунок 1.4 – Процесс сужения интервала исследования

1.3.3 Метод золотого сечения. Метод Фибоначчи требует дополнительных расчетов, так как заранее неизвестно число предполагаемых вычислений. Однако существует еще один метод, который совершенно не зависит от числа готовящихся вычислений и почти столь же эффективен, как метод Фибоначчи. Правило – каждый раз ставить точку расчета симметрично на расстоянии, равном 1/1,62≈0,62 от концов интервала исследования.

Золотое сечение, как известно, проводит деление отрезка АВ на две неравные части так, чтобы отношение всего отрезка АВ к большей части СВ равнялось отношению большей части СВ к меньшей АС.

1.4 Регулярные методы оптимизации. Методы направленного поиска

Под регулярными (детерминированными) методами оптимизации подразумевается такая направленная система действий, которая строго предопределяется сложившейся ситуацией. В настоящее время существует много методов регулярной оптимизации. Характерной чертой их является использование в процессе решения задачи результатов каждого данного шага (иногда также и предыдущих шагов) поиска оптимальной точки для определения направления изменения оптимизирующих параметров на каждом следующем шаге. При этом значение целевой функции систематически уменьшается (увеличивается).

1.4.1 Метод поочередного изменения параметров (метод покоординатного спуска, подъема, метод Гаусса – Зейделя). Суть его в поочередной оптимизации последовательно по каждому управляющему параметру.

Сначала оптимизация производится по одному параметру x_1, затем переходят ко второму x_2 и так далее, пока значения целевой функции не перестанут уменьшаться (возрастать) (рис. 1.5).

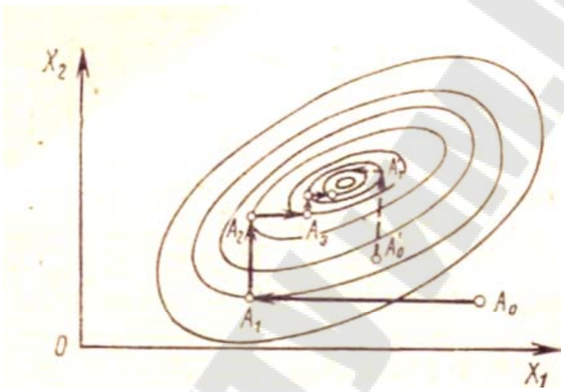

Рисунок 1.5 – Графическая интерпретация метода поочередного изменения параметров

Недостатки:

- обеспечивает получение только локального оптимума или особой точки типа седла;
- позволяет оптимизировать только непрерывно изменяющиеся параметры;
- результаты поиска существенно зависят от удачного выбора первого направления движения, начальной точки;
- эффективность метода существенно снижается при наличии ограничений.

Преимущество метода – простота алгоритма и программы.

1.4.2 Метод градиента. Один из самых распространенных регулярных методов поиска. Процесс оптимизации по методу градиента заключается в определении направления наибольшего изменения целевой функции $Y = f(x_1, ..., x_i, ..., x_n, u_1, ..., u_m)$ и некотором перемещении по этому направлению. Направление наибольшего изменения функции определяется направлением градиента оптимизируемой функции.

Для нахождения составляющих градиента необходимо: вычислить частные производные целевой функции по оптимизируемым параметрам $\frac{\partial f}{\partial x_i}\,(i=1,..,n)$.

Градиент функции $grad\ f(x)=\left(\frac{\partial f}{\partial x1},...,\frac{\partial f}{\partial x_n}\right)$, вычисленный в точке $N(x^{(N)}{}_1,\ ..x^{(N)}{}_n)$, характеризует направление быстрейшего возрастания функции, а антиградиент – направление быстрейшего убывания. Градиент всегда перпендикулярен к поверхности равных значений целевой функции.

Для определения оптимума необходимо при минимизации целевой функции для каждой точки поиска N определить градиент и сделать шаг по направлению антиградиента (т. е. в направлении, обратном градиентному) $x^{N+1}{}_i = x^N{}_i - a\frac{\partial f}{\partial x_i}\,(i=1,..,n)$, где a>0 – длина рабочего шага, значение которого может определяться двояко:

1. если a=const, то модуль рабочего шага для метода градиента оказывается пропорциональным модулю градиента:

$$\left|\Delta x^{N+1}\right| = \left|x_i{}^{N+1} - x_i{}^N\right| = a\sqrt{\left(\frac{\partial f}{\partial x1}\right)^2 + .. + \left(\frac{\partial f}{\partial xn}\right)^2},$$

т.е. не является постоянным;

2. если же $a = b\left|grad\ f(x)^N\right|^{-1}$, то модуль рабочего хода окажется постоянным, так как

$$\Delta x^{N+1} = b\left|grad\ f(x)^N\right|^{-1}\cdot\left|grad\ f(x)^N\right| = b = const.$$

Выбор того или иного значения a зависит от априорных представлений о форме гиперповерхностей равного уровня целевой функции. Существенное ускорение процесса поиска оптимума при сохранении простоты определения шага может быть достигнуто в случае использования комбинированного способа: шаг принимается постоянным и достаточно большим при движении вдали от оптимума, а после входа в зону оптимума предусматривается возможность уменьшения шага в 2, 4, ... раз (рис. 1.6).

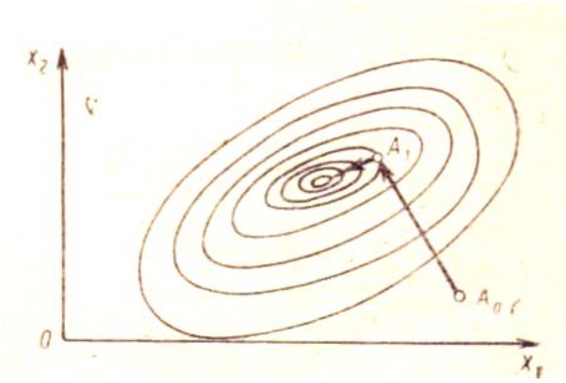

Рисунок 1.6 – Графическая интерпретация метода градиентов

Недостатки:

- необходимо перед каждым рабочим шагом производить довольно сложный предварительный анализ;
- при наличии ограничений применение метода затрудняется;
- необходимо определять частные производные целевой функции по оптимизируемым параметрам;
- невысокая скорость достижения оптимума;
- трудности в достижении конечной точки (оптимума);
- большая чувствительность к масштабным преобразованиям.

Преимущество – точность, так как на каждом рабочем шаге производится попытка двигаться в наилучшем направлении.

1.4.3 Метод наискорейшего спуска (подъема). Это разновидность метода градиента, обеспечивающая наименьшее число шагов в определении оптимума.

Суть его сводится к следующему. После определения градиента функции $Y=f(x)$ делается шаг в направлении, обратном градиенту. Если значение целевой функции при этом уменьшилось по сравнению с исходным, то делается очередной рабочий шаг в том же направлении, а не определяется заново градиент, как при методе градиента. Если же после очередного рабочего шага значение целевой функции увеличилось по сравнению с предыдущим значением, то движение прекращается, заново определяется направление градиента и т. д. Повышенная скорость сходимости является существенным преимуществом этого метода.

1.5 Методы случайного поиска

Методы случайного поиска отличаются от регулярных (детерминированных) методов оптимизации намеренным введением элемента случайности. Регулярные методы оптимизации более тонко настроены на специфику исследуемого процесса. Методы случайного поиска эффективны при решении сложных задач больших размерностей с произвольно заданными целевыми функциями, ограничениями, тогда как регулярные методы, как правило, неприменимы.

Под поиском понимается процесс отыскания такого значения критерия оптимизации (целевой функции), которое близко к оптимальному и в то же время удовлетворяет всем ограничениям.

Различают:

- ненаправленный случайный поиск,
- направленный случайный поиск без самообучения,
- направленный случайный поиск с самообучением.

1.5.1 Ненаправленный случайный поиск (или метод статистических испытаний, метод Монте-Карло) заключается в многократном моделировании независимых случайных вариантов решений из области допустимых, вычислении в каждом из них критерия оптимизации и запоминании наиближайшего к экстремуму.

Метод Монте-Карло относится к числу универсальных, поскольку позволяет решать многоэкстремальные задачи общего вида с отысканием глобального экстремума.

Основной недостаток метода заключается в необходимости проведения большого числа испытаний для получения решения, достаточно близкого к оптимальному, т. е. наличие медленной сходимости к экстремуму.

1.5.2 Направленный случайный поиск без самообучения является модернизацией направленного случайного поиска. В основе метода лежит использование результатов поиска предыдущих шагов.

1.5.3 Направленный случайный поиск с самообучением заключается в добавлении алгоритмов самообучения, которые корректируют вектор предыстории (случайный вектор E) преимущественно в направлении удачного предыдущего шага или в направлении, обратном предыдущему неудачному шагу, в случае нелинейной целевой функции, например путем перестройки вероятностей P_1 ..., P_i, ... P_n составляющих случайного вектора E=(ε_1 ..., ε_i..., ε_n). Так, при целевой функции, близкой линейной, вероятность шага в удачном направлении увеличивается, и наоборот.

На базе широкого использования различных методов случайного поиска развилось новое направление в исследовании самых разнообразных процессов – имитационное моделирование. Вообще под имитацией понимается численный метод исследования системы, процесса, операции с помощью математических моделей, описывающих функционирование исследуемых систем, на ЭВМ для анализа синтеза интересующих исследователя функциональных характеристик. Имитационное моделирование позволяет проводить широкие исследования случайных факторов, которыми изобилуют реальные системы, процессы, операции. Но в ряде случаев оно требует больших затрат на программирование, отладку моделей и проведение исследования, а также усложняется процесс интерпретации полученных результатов.

Использование методов случайного поиска определяется в основном следующими ситуациями:

- неприемлемостью или отсутствием соответствующих аналитических и численных методов исследования модели;
- возможностью создания модели функционирования системы (процесса), получением необходимого количеств информации о моделируемой системе (процессе);
- наличием, большого числа случайных факторов в исследуемой системе (процессе);
- наличием современной вычислительной техники и соответствующего программного обеспечения.

2 ОПЕРАТИВНО – КАЛЕНДАРНОЕ ПЛАНИРОВАНИЕ В ТЕХНОЛОГИЧЕСКИХ СИСТЕМАХ НА ОСНОВЕ ТЕОРИИ РАСПИСАНИЙ

2.1 Задачи теории расписаний

2. 1. 1 Анализ задач теории расписаний

Теория расписаний занимается изучением специфических моделей, имеющих место в оперативно-календарном планировании, и разработкой математических методов построения наилучших в смысле заданного критерия эффективности календарных планов. Модели, изучаемые в теории расписаний, могут быть отнесены к классу детерминированных. Это значит, что обстановка в которой приходится вырабатывать наилучшее решение, характеризуется полной определенностью.

Задачи теории расписаний могут быть условно разделены на три группы [5]:

1. задачи согласования (задачи сетевого планирования и управления);
2. задачи распределения (задачи балансирования сборочной линии, задача о назначениях и др.);
3. задачи упорядочения.

Задачи упорядочения носят самый общий характер. Они возникают повсюду, где существует возможность выбора той или иной очередности выполнения работ: при распределении работ на производстве, составлении расписания приземления самолетов, обслуживании клиентов в банках, формировании очередности выполнения программ вычислительным центром и т.д.

Формальная схема процесса упорядочения примерно такова:

- α есть совокупность последствий при выполнении сначала задачи *A,* а затем *B,*
- β есть совокупность последствий при выполнении сначала задачи *B,* а затем *A,*
- если α предпочтительное, чем β, то выбирается последовательность: «*A*, затем *B*».

В дальнейшем рассматриваются задачи упорядочения при условии, что решены все вопросы, относящиеся к тому, что и каким образом должно быть выполнено. При этом предполагается, что не существует зависимости между характером этих решений и устанавливаемым порядком. Кроме того, предполагается следующее.

1. Подлежащие выполнению работы определены и известны полностью.
2. Однозначно определены устройства, выделяемые для выполнения заданных работ.
3. Задана совокупность всех элементарных действий, связанных с выполнением каждой из работ, и ограничений, налагаемых на порядок их выполнения.

Существует определенное различие между упорядочением и составлением расписания. Упорядочение подразумевает формирование очередности операций, выполняемых одной машиной, в то время как составление расписания означает задание последовательности действий нескольким машинам. Однако такое различие не вносит существенной четкости в постановку задач, поэтому оба термина будут рассматриваться как синонимы.

Основным понятием теории расписаний является понятие *операции.* Операцию можно рассматривать как элементарную задачу, подлежащую выполнению.

Каждая операция характеризуется:

- индексом принадлежности к определенной *работе;*
- индексом принадлежности к определенной *машине;*
- числом, представляющим собой *длительность* операции.

По первому индексу все множество операций разбивается на полную систему непересекающихся подмножеств, называемых *работами.*

Разбиение исходного множества по второму индексу приводит к взаимно непересекающимся подмножествам операций, относящихся к определенным *машинам.*

Для каждой работы задается последовательность составляющих ее операций (определяемая технологическим процессом).

Машиной будем называть устройство, способное выполнить все, что связано с некоторой операцией; *системой обслуживания* – множество всех машин, используемых для выполнения некоторого подмножества операций. Совокупность машин, работ (операций) и дисциплин назначения операций соответствующим машинам назовем *процессом обслуживания.*

С другой стороны, расписание может рассматриваться как задача *упорядочения операций, выполняемых каждой машиной.*

Почти вся теория, разработанная в настоящее время, относится к весьма ограниченному числу моделей *простого процесса обслуживания.* Под последним понимается процесс, для которого существенны следующие ограничения.

1. Каждая машина может быть назначена в любой момент времени. Машины не могут выходить из строя и, следовательно, быть недоступными из-за неисправности или ремонта. Они не могут быть недоступны также из-за пересменок, типичных для производства. Каждая машина формально представляет собой интервал (0, T), где T есть произвольно большое число.
2. Работы представляют собой строго упорядоченные последовательности операций. Из отдельных операций этих последовательностей не может формироваться никакая новая последовательность.
3. Каждая операция выполняется только одной машиной.
4. Существует только по одной машине каждого типа.
5. Отсутствуют прерывания операций. Это означает, что если некоторая машина начала выполнять операцию, то эта операция должна быть завершена прежде, чем на этой же машине начнется выполнение какой-либо другой операции.
6. Интервалы выполнения последовательных операций одной и той же работы не пересекаются. Одновременно не может выполняться двух операций одной и той же работы.
7. В каждый момент времени машина может выполнять не более одной операции.

Перечисленные ограничения, с одной стороны, упрощают формализацию, а с другой, делают ее более абстрактной. Последнее приводит к тому, что модель становится неадекватной большинству практических случаев, где требуется ослабление одного или нескольких из приведенных ограничений. Тем не менее, такая модель сохраняет в основном структуру большинства практических задач, а при ее исследовании вырабатывается интуиция, полезная в многочисленных приложениях. Во всяком случае, в настоящее время описанная формализация применяется в большинстве исследований. Если говорить о более сложных моделях, то для них пока существуют лишь отдельные результаты.

2.1.2 Классификация задач теории расписаний

Задача теории расписаний считается заданной, если определены:

- подлежащие выполнению работы и операции;
- количество и типы машин, выполняющих операции;
- порядок прохождения машин;
- критерии оценки расписаний.

Задачи теории расписаний различаются числом выполняемых в системе работ, характером поступления их в систему и порядком участия отдельных машин в выполнении конкретной работы. В зависимости от характера поступления работ различают два вида задач: *статические* и *динамические.* В статических задачах, если система свободна, в нее одновременно поступает определенное число работ. После этого новые работы не поступают и расписание составляется для вполне определенного и известного заранее числа работ. В динамических задачах выполнение работ происходит непрерывно. Работы поступают в систему в некоторые моменты времени, которые можно предсказать только в статистическом смысле. Поэтому моменты будущих поступлений не определены. Можно ожидать, что упорядочение в динамических и статических задачах потребует совершенно различных методов решения.

Порядок выполнения машинами операций одной работы определяет, является ли система машин *конвейерной.* В конвейерной системе последовательность прохождения машин одинакова для каждой из работ. Согласно принятой терминологии это означает, что существует такая нумерация машин, что для одной и той же работы номер машины, выполняющей операцию x, меньше номера машины, выполняющей операцию y, если x предшествует y. В противоположность этому существуют системы со *случайным порядком выполнения работ машинами.* Здесь отсутствует описанная направленность прохождения машин, и в этом случае любая операция любой работы может выполняться равновероятно любой машиной [5].

В качестве примера задачи упорядочения рассмотрим одну из задач оперативно-календарного планирования работы производственного участка, обеспечивающего выпуск некоторого количества деталей различных типов. Для каждого типа деталей предполагается известными технологическая последовательность обработки деталей на станках и время обработки каждой детали на каждом из станков. Требуется принять решения, направленные на эффективную организацию работы участка, т.е. определить такой порядок запуска деталей в производство, при котором общее время пребывания их на обработке было бы минимальным.

Покажем, что общее время пребывания деталей на обработке (участке) зависит от порядка запуска деталей в производство [18].

Пример 2.1. Пусть участок состоит из двух станков №1 и №2. По плану должны быть изготовлены три различные детали, каждая из которых сначала обрабатывается на станке №1, затем на станке №2. Время обработки деталей на каждом из станков задано в таблице 2.1.

Таблица 2.1 –Исходные данные

Станок	**Детали**		
	Д1	**Д2**	**Д3**
№1	3	1	2
№2	2	4	1

Существуют шесть различных способов запуска деталей в производство. Рассмотрим и сравним два из них: (Д2, Д1, Д3) и (Д3, Д1, Д2).

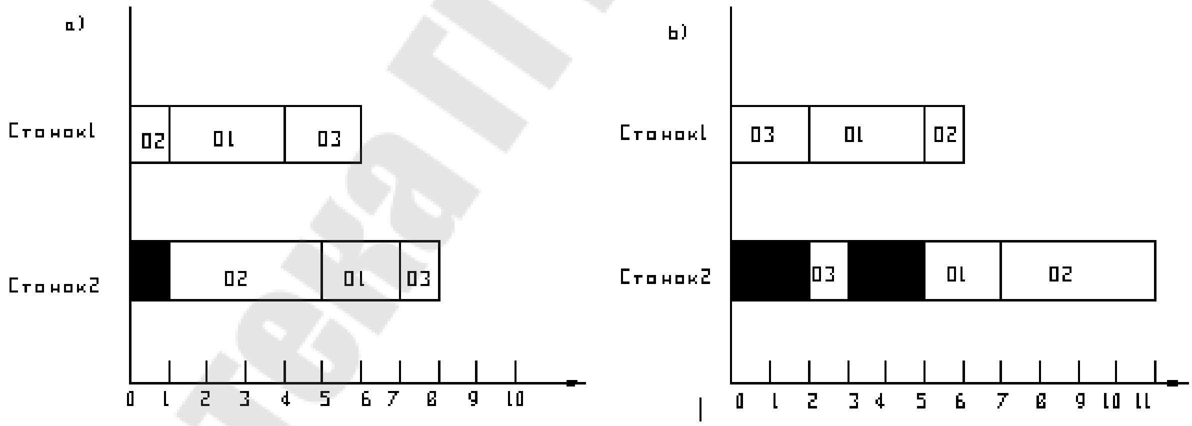

Рисунок 2.1 – Два способа запуска деталей в производство

Так как на станке №2 деталь может обрабатываться лишь после того как она обработана на станке №1, график обработки деталей показан на рис2.1. Из рисунка 2.1 видно, что при первом способе запуска деталей в производство для выполнения всего плана требуется 8 ед. времени, в то время как при втором – 11 ед., т.е. первый способ оформления работы является более выгодным.

2.1.3 Формы представления расписаний

Формы представления расписаний могут быть разделены на три группы: графические, аналитические, табличные [18].

Там, где требуется наглядность часто используются графические способы представления расписаний (графики Ганта, Гант-карты, планировочные графики, учетные графики и т.д.).

График-Ганта представляет упорядочивание занятости различных устройств, механизмов, рабочих мест и т.д. Каждая операция на графике изображается масштабированным отрезком прямой, соответствующим длительности выполнения операции. Моменты начала и конца операций изображаются ⌈ ⌉ соответственно.

Планировочный график – это схематическое изображение процесса, состоящего из ряда работ. При построении планировочного графика каждой работе отводится отдельная строка, в которой указывается названия работы, и проводится линия, соответствующая продолжительности работы в выбранном масштабе времени (табл. 2.2).

Таблица 2.2 – Пример планировочного графика

Работа	**Недели**							
	1	2	3	4	5	6	7	8
Ознакомление с заданием								
Выбор и обоснование метода								
Разработка блок-схемы алгоритма								
Написание программы								
Отладка программы								
Оформление отчета								

Ленточный график служит для изображения выполнения операций, составляющих одну работу. Так как операции во времени не пересекаются, на графике они изображаются последовательно с учетом времени, необходимого для их выполнения (табл. 2.3).

Таблица 2.3 – Пример ленточного графика

Работа	**Время, мин**						
	5	10	15	20	25	30	35
Токарная обработка							
Сверление отверстий							
Шлифование							
Окраска							
Сушка							

Среди аналитических способов задания расписания можно выделить способ задания с помощью вектор- функции времени. Для рис.10.1 достаточно задать вектор-функцию $S(t) = (S_1(t), S_2(t))$, где $S_1(t)$ описывает загрузку станка №1, а $S_2(t)$ – станка №2. Если $S_i(t') = 0, i = 1, 2$, то это означает, что станок i в момент времени t' не занят работой и простаивает. Если же $S_i(t') = R, i = 1, 2$, то это означает, что в момент времени t' на станке i выполняется операция *R*. Так, расписание, представленное на рисунке 2.1а в виде вектор-функции может быть записано следующим образом:

$$S(t) = \begin{pmatrix} S_1(t) = \begin{cases} 2, \textit{при } 0 \le \text{t} < 1, \\ 1, \textit{при } 1 \le \text{t} < 4, \\ 3, \textit{при } 4 \le \text{t} < 6, \\ 0, \textit{в остальных случаях} \end{cases} \\ S_2(t) = \begin{cases} 2, \textit{при } 1 \le \text{t} < 5, \\ 1, \textit{при } 5 \le \text{t} < 7, \\ 3, \textit{при } 7 \le \text{t} < 8, \\ 0, \textit{в остальных случаях} \end{cases} \end{pmatrix}.$$

Для выработки оптимальных решений в задачах теории расписаний широко применяются различные методы исследования операций комбинаторные, методы математического программирования: методы динамического и дискретного программирования и т.д.

2.2 Задачи теории расписаний с одним обслуживающим устройством

2.2.1 Постановка задачи и критерии эффективности

Дадим постановку задачи с одним обслуживающим устройством в терминах «машина-работа (операция)». Пусть для выполнения на одной машине одновременно поступает множество $N = \{1, 2, ..., n\}$ работ. Также предположим, что продолжительность выполнения каждой работы на машине известно, и обозначим ее через t_i, $i \in N$.

Задача построения расписания выполнения множества работ одной машиной состоит в том, чтобы определить такой порядок выполнения работ, при котором некоторый заданный критерий эффективности будет принимать оптимальное решение [18].

Введем следующие обозначения, необходимые в дальнейшем:

- $\underline{t_i}$ – время начала выполнения работы $i \in N$;
- $\overline{t}_i$ – время окончания выполнения работы $i \in N$;
- d_i – директивное время, в течение которого должно быть завершено выполнение работы $i \in N$;
- α_i – штраф за ожидание работы i в единицу времени (или некоторая стоимость, связанная с выполняемой работы) до момента начала ее обработки;
- β_i – стоимость выполнения работы $i \in N$.

Время начала и окончания выполнения работ связано следующей зависимостью: $\overline{t}_i = \underline{t_i} + t_i$.

Предполагается, что перерывы в выполнении работ не допускаются. Это значит, что если машина приступила к выполнению некоторой работы, то она продолжает выполнять ее до тех пор, пока не закончит. Время T, необходимое для выполнения всех работ множества N, не зависит от порядка выполнения работ и равно сумме времен выполнения всех работ $T = \sum_{i \in N} t_i$.

Можно записать, что задержка z_i (превышение директивного срока пребывания в системе) работы i составляет

$$z_i = \max(0, \overline{t}_i - d_i), \quad i \in N.$$

Таким образом, критерий, позволяющий вычислять максимальный штраф, связанный с опозданием в выполнении работ, имеет вид:

$$\Phi_1 = \max_{i \in N}(\alpha_i \cdot z_i).$$

Часто встречается критерий, представляющий сумму штрафов, связанных с ожиданием работ в системе, которую необходимо минимизировать

$$\Phi_2 = \sum_{i \in N} \alpha_i \cdot \underline{t_i}.$$

Рассмотрим задачу минимизации суммы связанных средств на производственном участке. Стоимость, связанная с выполняемой работой, после ее выполнения будет равна $\gamma_i = \alpha_i + \beta_i$. Время, в течение которого выполненная работа i будет находиться на участке, ожидая окончания всех работ, составит $T - \overline{t_i}$

Можно записать критерий, который можно использовать для минимизации суммы связанных средств

$$\Phi_3 = \sum_{i \in N} \overline{\gamma_i} \cdot \underline{t_i},$$

где $\overline{\gamma_i} = -\gamma_i$.

2.2.2 Алгоритмы решения задач с одним обслуживающим прибором

Рассмотрим задачу построения оптимального расписания по критерию Φ_2. Для него может быть сформулировано следующее правило [18].

1. Для всех работ вычислить отношения $\dfrac{t_i}{\alpha_i}$.
2. Упорядочить работы по возрастанию этого отношения.

Учитывая, что $\overline{\gamma_i} = -\gamma_i$, получим правило для построения расписания, минимизирующего сумму связанных средств по критерию Φ_3:

1. Для всех работ вычислить отношения $\dfrac{t_i}{\gamma_i}$.
2. Упорядочить работы по убыванию этого отношения.

Приведем алгоритм, построения оптимального расписания по критерию Φ_1 [18].

Предварительный шаг. Вычисляем время окончания всех работ T. Переходим к шагу1.

Шаг1. Среди всех неупорядоченных работ находим такую работу l, для которой

$$\alpha_l \cdot z_l = \min_i \alpha_i \cdot z_i,$$

где $z_i = \max(0, T - d_i)$. Здесь минимум по i вычисляется по множеству индексов только неупорядоченных работ. Переходим к шагу 2.

Шаг 2. Работу с номером l выполняем последней среди рассматриваемого множества. Исключаем работу l из рассмотрения. Если множество работ пусто, то задача решена. Если нет, то заменяем T на $T - t_l$ и переходим к Шагу1.

После повторения n раз Шага1 и Шага2 будет построено расписание, оптимальное по критерию Φ_1.

Пример 2.2. Определить оптимальный порядок обработки деталей, минимизирующий сумму штрафов, связанных с пролежанием деталей в ожидании обслуживания для исходных данных (табл. 2.4), приведенных в таблице и построить оптимальное расписание в виде кусочно непрерывной функции $S(t)$.

Таблица 2.4 – Исходные данные

Характеристики	**Номер детали**					
	Д1	Д2	Д3	Д4	Д5	Д6
t_i	3	5	4	3	2	6
α_i	2	2	1	3	1	2

Для построения оптимального расписания в соответствии с критерием Φ_2 вычислим отношения $\dfrac{t_i}{\alpha_i}$ и пронумеруем детали в порядке возрастания отношения.

Результаты вычислений представлены в таблице 2.5 а график функции $S(t)$ на рисунке 2.2.

При этом значение критерия эффективности равен:

$$\varPhi_2 = \sum_{i \in N} \alpha_i \underline{t_i} = 3 \cdot 0 + 2 \cdot 3 + 1 \cdot 6 + 2 \cdot 8 + 2 \cdot 13 + 1 \cdot 19 = 73,$$ где $\underline{t_i}$ – начало работы.

Таблица 2.5 –Результаты вычислений

Параметры	**Номер детали**					
	Д1	Д2	Д3	Д4	Д5	Д6
t_i / α_i	1.5	2.5	4	1	2	3
Оптимальный порядок	2	4	6	1	3	5

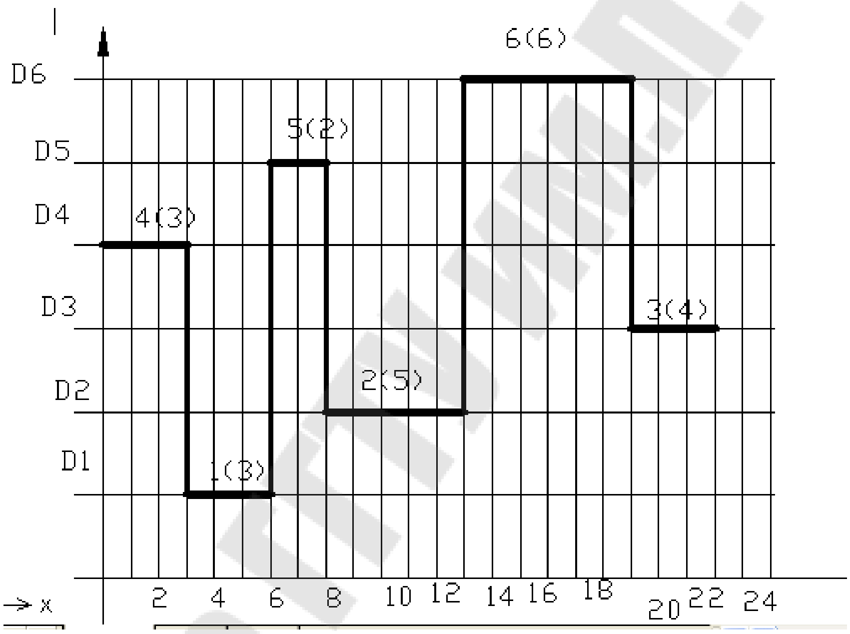

Рисунок 2.2 – Расписание в виде кусочно –непрерывной функции

2.3 Задача теории расписаний с двумя последовательными обслуживающими устройствами (Задача Джонсона)

2.3.1 Постановка задачи и алгоритм Джонсона

Дадим формулировку в общем виде.

Имеется множество $N = \{1, 2, ..., n\}$ работ, которые должны быть выполнены на m машинах. Время работы i на машине j обозначим через $t_{i,j}$ $(i = 1 \div n, \quad j = 1 \div m)$, предполагая его заранее известным. Порядок выполнения операций, составляющих работу, может быть как одним и тем же, так и различным для разных работ. Задача построения расписания состоит в указании порядка, в котором должны выполняться работы, чтобы суммарное время простоя всех машин было минимальным. При построении любого расписания в том числе и оптимального должны учитываться следующие условия.

1. В любой момент времени на машине не может выполняться больше одной работы.
2. Одна работа в фиксированный момент времени может занимать только одну машину.

Сначала рассмотрим случай, когда число машин равен двум (M_1 и M_2). Каждая работа состоит из двух операций, которые выполняются сначала на первой машине, затем на второй. Время работы i на первой, а затем на второй машине равно $t_{i,1}$ и $t_{i,2}$ соответственно.

Считая, что порядок выполнения операций на первой и второй машине один и тот же, приведем следующий алгоритм построения оптимального расписания, который называется *алгоритмом Джонсона* [18].

Предварительный шаг. Записываем матрицу $\left\| t_{j,i} \right\|$ $(i = 1 \div n, \quad j = 1, 2)$ времени выполнения операций. Переходим к первому шагу.

Первый шаг. Выбираем в матрице $\left\| t_{j,i} \right\|$ минимальный элемент. Если он находится в первой строке (соответствующей первой машине), то данную работу выполняем первой, если во второй строке – то последней. Переходим к шагу два.

Второй шаг. Исключаем из рассмотрения время выполнения операции, относящееся к упорядоченной работе. Если множество элементов матрицы $\left\| t_{j,i} \right\|$ пусто, то задача решена. Если нет, переходим к первому шагу.

Таким образом, для построения оптимального расписания Шаг1 и Шаг2 должны быть повторены *n* раз. Если же случится, что $t_{1,i} = t_{2,i}$, то эта работа может быть упорядочена как по $t_{1,i}$, так и по $t_{2,i}$.

Описанный алгоритм минимизирует суммарное время простоя обеих машин.

Пример 2.3. Имеется пять работ, каждая из которых состоит из двух операций, которые выполняются сначала на первой, затем на второй машине. Время выполнения операций приведено в таблице 2.6.

Таблица 2.6 – Исходные данные

	P1	**P2**	**P3**	**P4**	**P5**
M1	5	3	4	1	2
M2	2	1	3	2	3

По алгоритму Джонсона оптимальный порядок выполнения работ (P4, P5, P3, P1, P2).

Для определения времени простоя второй машины построим график Ганта.

Из графика Ганта (рис. 2.3) нетрудно заметить, что если машины начинают выполнять работы одновременно, то время простоя составит 5 единиц. Для ликвидации простоя необходимо обеспечить дополнительный уровень работы.

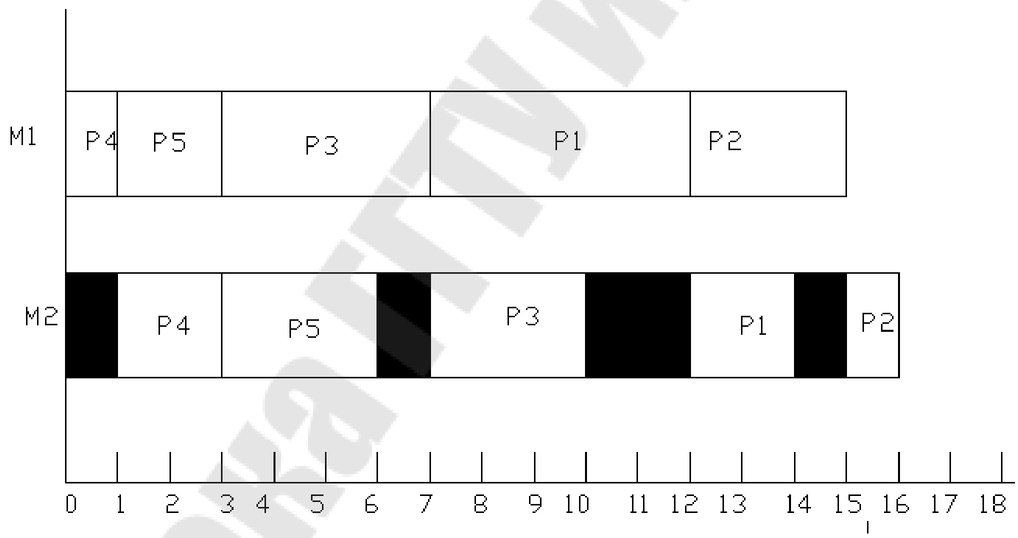

Рисунок 2.3 – График Ганта для двух машин

2.3.2 Смешанный вариант задачи Джонсона

Рассмотрим ситуацию, когда множество работ, подлежащих выполнению может быть разделено на 4 подмножества:

- N_1 – подмножество работ, которые выполняются только на первой машине;
- N_2 – подмножество работ, которые выполняются только на второй машине;

- $N_{1,2}$ – подмножества работ, которые выполняются сначала на первой, затем на второй машине;
- $N_{2,1}$ – подмножества работ, которые выполняются сначала на второй, затем на первой машине.

Для того чтобы оптимально упорядочить множество работ $N = N_1 \cup N_2 \cup N_{1,2} \cup N_{2,1}$ необходимо сначала определить оптимальный порядок выполнения работ, принадлежащих подмножествам $N_{1,2}$ и $N_{2,1}$, в соответствии с алгоритмом Джонсона. Чтобы суммарное время простоя было минимальным (в данном случая может простаивать и первая машина) необходимо:

- на первой машине выполнить сначала работы из подмножества $N_{1,2}$, затем из N_1, а потом – из $N_{2,1}$;
- на второй машине выполнить сначала работы из подмножества $N_{2,1}$, затем из N_2, а потом – из $N_{1,2}$.

Рассмотренный общий случай задачи Джонсона для двух машин часто называют *смешанным вариантом* задачи Джонсона [18].

Пример 2.4. Условие смешанного варианта задачи Джонсона для двух машин задается следующей таблицей 2.7.

Таблица 2.7 Исходные данные для смешанной задачи Джонсона

N	$N_{1,2}$				$N_{2,1}$			N_1		N_2	
$M \backslash P$	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}
M_1	3	1	2	4	1	3	2	5	2	0	0
M_2	4	3	1	2	5	2	1	0	0	2	1

здесь p_j – имя работы, M_i – имя машины.

По алгоритму Джонсона: оптимумы для $N_{1,2}$ и $N_{2,1}$, соответственно: (p_2, p_1, p_4, p_3) и (p_7, p_6, p_5).

Оптимальные режимы работ:

- Машина №1: $(p_2, p_1, p_4, p_3, (p_8, p_9), p_7, p_6, p_5)$;
- Машина №2: $(p_7, p_6, p_5, (p_{10}, p_{11}), p_2, p_1, p_4, p_3)$.

Записи (p_8, p_9) и (p_{10}, p_{11}) означают, что порядок выполнения этих работ произвольный.

2.4 Задача теории расписаний с тремя и более последовательными обслуживающими устройствами

2.4.1 Частные случаи решения задачи для трех машин

В общем виде точное решение задачи Джонсона найдено только для двух машин. Во всех остальных случаях приходится использовать специальные математические методы, которые весьма трудоемкие в вычислительном плане.

Рассмотрим случай, когда число машин равно трем, предполагая, что порядок выполнения работ на машинах один и тот же [18].

Рассмотрим частные случаи, когда решение задачи может быть получено достаточно просто.

Перейдем к случаю трех машин.

Пусть m=3 – число машин; $N = \{1,2,...,n\}$ – множество работ; $t_{i,j}$ – время выполнения i-ой работы на j-ой машине. Предполагается, что у всех работ одна и та же последовательность прохождения по машинам.

В этой ситуации справедливы нижеследующие два утверждения, которые приводятся без доказательства.

<u>Утверждение №1</u>. Пусть работы можно пронумеровать так, что окажутся выполненными одновременно следующие неравенства:

$$t_{1,1} \le t_{2,1} \le \ldots \le t_{n,1},$$
$$t_{1,2} \le t_{2,2} \le \ldots \le t_{n,2},$$
$$t_{1,3} \le t_{2,3} \le \ldots \le t_{n,3},$$
$$t_{i,1} \le t_{i,2} \le t_{i,3}, \forall i.$$

Тогда суммарный простой машин будет минимальным при следующем порядке запуска работ на исполнение: 1–2–…–n.

<u>Утверждение №2</u>. Пусть для матрицы $(t_{i,j})$ выполнено хотя бы одно из двух следующих условий:

$$\begin{matrix} \min\limits_i t_{i,1} \ge \max\limits_i t_{i,2} \\ \min\limits_i t_{i,3} \ge \max\limits_i t_{i,2} \end{matrix}.$$

Построим новую матрицу $(\tau_{i,j})$, в которой i=1, 2,..., n, j=1, 2 и $\tau_{1,i} = t_{1,i} + t_{2,i}$ $\tau_{i,2} =_{,1,i} +_{,1,i}$ и будем считать ее матрицей времен задачи Джонсона для двух машин в последовательном варианте и множества тех же работ $N=\{1, 2, ..., n\}$. Тогда суммарное время простоя исходных трех машин при выполнении исходных работ будет минимальным, если эти работы направлять на исполнение в том порядке, который является оптимальным в задаче с матрицей $(\tau_{i,j})$.

Вопросы для самопроверки

1. Какие используются формы представления расписаний?
2. Дайте постановку задачи с одним обслуживающим устройством.
3. Алгоритмы решения задач с одним обслуживающим прибором
4. Алгоритм Джонсона.
5. Смешанный вариант задачи Джонсона.
6. Частные случаи решения задач теории расписаний с тремя обслуживающими устройствами.

3 МОДЕЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ НА ОСНОВЕ ТЕОРИИ ИГР

3.1 Использование матричных игр при решении технологических задач

3.1.1 Основные понятия теории игр

В процессе целенаправленной человеческой деятельности возникают ситуации, в которых интересы отдельных лиц (участников, групп, сторон) либо прямо противоположны (антагонистичны), либо не будучи непримиримыми, все же не совпадают. Простейшими и наиболее наглядными примерами таких ситуаций являются спортивные игры, арбитражные споры и т. п. Здесь каждый из участников сознательно стремится добиться наилучшего результата за счет другого участника. Подобного рода ситуации встречаются и в различных сферах производственной деятельности.

Для указанных ситуаций (будем называть их, конфликтными), характерно, что эффективность решений, принимаемых в ходе конфликта каждой из сторон, существенно зависит от действий другой стороны. При этом ни одна из сторон не может полностью контролировать положение, так как и той и другой стороне решения приходится принимать в условиях неопределенности. Так, при определении объема выпуска продукции на одном предприятии нельзя не учитывать размеров выпуска аналогичной продукции на других предприятиях. В реальных условиях нередко возникают ситуации, в которых антагонизм отсутствует, но существуют противоположные тенденции. Так, для нормального функционирования производства необходимо наличие запасов разнообразных ресурсов, но с другой стороны стремление к чрезмерному увеличению этих запасов вызывает дополнительные затраты по их хранению. В приведенных примерах конфликтные ситуации возникают в результате сознательной деятельности людей. Однако на практике встречаются и такие неопределенности, которые порождаются не сознательным противодействием другой стороны, а недостаточной информированностью об условиях проведения планируемой операции.

Раздел математики, изучающий *конфликтные* ситуации, т. е. ситуации, в которых интересы участников (игроков) противоположны или не совпадают, называется *теорией игр.* Теория игр – это математическая теория конфликтных ситуаций, определяющая рекомендации по рациональному образу действий каждого из участников в ходе конфликтной ситуации, т. е. таких действий, которые обеспечивали бы ему наибольший выигрыш (наименьший проигрыш). Игровую схему можно придать многим ситуациям в экономике. Здесь выигрышем может быть эффективность использования дефицитных ресурсов, производственных фондов, величина прибыли, себестоимость и т. д.

Чтобы проанализировать конфликтную ситуацию с помощью математических методов, ее необходимо упростить, учитывая лишь важнейшие факторы, существенно влияющие на ход конфликта. Отсюда *игра* – это упрощенная модель конфликтной ситуации, отличающаяся от реального конфликта тем, что ведется по определенным правилам. Поэтому можно сказать, что *игра* – совокупность правил, определяющих возможные действия участников игры. Суть игры в том, что каждый из участников принимает такие решения, которые, как он полагает, могут обеспечить ему наилучший результат (исход). *Исход* игры – это значение некоторой функции, называемой *функцией выигрыша (платежной функцией).*

Будем рассматривать только такие игры, в которых выигрыш выражается количественно. Величина выигрыша зависит от стратегии, применяемой игроком. *Стратегия* – это совокупность правил, однозначно определяющих последовательность действий игрока в каждой конкретной ситуации, складывающейся в процессе игры. Всякая игра состоит из отдельных партий. *Партией* называют каждый вариант проведения игры определенным образом. В свою очередь в партии игроки совершают определенные ходы. *Ход* – это выбор и реализация игроком одного из возможных вариантов поведения. Ходы бывают *личные* и *случайные.* При *личном ходе* игрок сознательно выбирает и реализует ту или иную стратегию. При *случайном ходе* выбор стратегии производится с использованием механизма случайного выбора, например, с применением таблицы случайных чисел [7].

Конфликтные ситуации, встречающиеся в практике, порождают различные виды игр. Классифицировать игры можно по разным признакам. Различают, например, игры по количеству игроков. Если при этом игроки объединяются, например, в две группы, преследующие противоположные цели, то имеет место *игра двух «лиц» (парная игра).* В зависимости от количества стратегий в игре они делятся на *конечные* и *бесконечные.* В зависимости от взаимоотношений участников различают игры *бескоалиционные* (участники не имеют права заключать соглашения), или *некооперативные,* и *коалиционные,* или *кооперативные* [7].

По характеру выигрышей игры делятся на игры с *нулевой суммой* и *ненулевой суммой.* В первых – общий капитал игроков не меняется, а лишь перераспределяется в ходе игры, в связи с чем сумма выигрышей равна нулю. В играх с ненулевой суммой сумма выигрышей участников отлична от нуля [7].

По виду функции выигрыша игры делятся на *матричные, биматричные, непрерывные, выпуклые, сепарабельные* и др. В матричных играх (при двух участниках) выигрыши первого игрока задаются матрицей, в биматричных выигрыши каждого игрока задаются своей матрицей. Другие типы таких игр различаются видом аналитического выражения функции выигрышей. По количеству ходов игры делятся на *одноходовые* (выигрыш распределяется после одного хода каждого игрока) и *многоходовые* (выигрыш распределяется после нескольких ходов). Многоходовые игры в свою очередь делятся на *позиционные, стохастические, дифференциальные* и др. В зависимости от состояния информации различают игры с полной и неполной информацией [7].

3.1.2 Матричные игры. Решение матричных игр в чистых стратегиях

Будем рассматривать игры, в которых у каждого из двух игроков *A* и *B* конечное число возможных действий – *чистых стратегий.* Допустим, что игрок *A* располагает *m* чистыми стратегиями $A_1, ..., A_m$, а игрок *B* – *n* чистыми стратегиями $B_1, ..., B_n$. Чтобы игра была полностью определенной, необходимо указать правило, сопоставляющее каждой паре чистых стратегий A_i и B_j число $a_{i,j}$ – выигрыш игрока *A* за счет игрока *B* или проигрыш игрока *B*. При $a_{i,j} < 0$ игрок *A* платит игроку *B* сумму $\left|a_{i,j}\right|$. Если игра состоит только из личных ходов, то выбор пары чистых стратегий (A_i, B_j) единственным образом определяет исход (результат) игры. Если же в игре используются и случайные ходы, то исход, игры обусловлен средним значением выигрыша (математическим ожиданием). Таким образом, мы рассматриваем *парные игры с нулевой суммой,* в которых выигрыш одного игрока равен проигрышу другого [7].

Если известны значения $a_{i,j}$ выигрыша для каждой пары (A_i, B_j) стратегий, то можно составить *матрицу игры* – платежную матрицу (табл. 3.1). *Платежная матрица* – это табличная запись функции выигрыша. Описанные игры называют *матричными* или *прямоугольными.* Отдельная партия в такой игре реализуется следующим образом. Игрок *A* выбирает одну из строк платежной матрицы (одну из своих чистых стратегий). Не зная результата его выбора, игрок *B* выбирает один из столбцов (свою чистую стратегию). Элемент матрицы, стоящий на пересечении выбранных строки и столбца, определяет *выигрыш* игрока *A (проигрыш* игрока *B*).

Целью игроков является выбор наиболее выгодных стратегий, доставляющих игроку *A* максимальный выигрыш, а игроку *B* минимальный проигрыш. В теории игр исходят из предположения, что каждый игрок считает своего противника разумным и стремящимся помешать ему достичь наилучшего результата.

Стратегию игрока *A* называют *оптимальной,* если при ее применении выигрыш игрока *A* не уменьшается, какими бы стратегиями не пользовался игрок *B*. Оптимальной для игрока *B* называют стратегию, при использовании которой проигрыш игрока *B* не увеличивается, какие бы стратегии ни применял игрок *A*. С учетом этого игрок *A* анализирует матрицу выигрышей следующим образом: для каждой своей чистой стратегии *A*, – он определяет минимальное значение α_i, – выигрыша в зависимости от применяемых игроком *B* чистых стратегий B_j, т.е. $\alpha_i = \min a_{i,j}$ (i=1, …, m).

Таблица 3.1 – Платежная матрица

A_i	B_j				α_i
	B_1	B_2	...	B_n	
A_1	$a_{1,1}$	$a_{1,2}$	...	$a_{1,n}$	α_1
A_2	$a_{2,1}$	$a_{2,2}$	...	$a_{2,n}$	α_2
...	...	...	...	...	...
A_m	$a_{m,1}$	$a_{m,2}$	...	$a_{m,n}$	α_m
β_j	β_1	β_2	...	β_n	

Затем по минимальным выигрышам α_i он отыскивает такую чистую стратегию A_0, при которой этот минимальный выигрыш будет максимальным, т. е. находит

$$\alpha = \max_i \alpha_i = \max_i \min_j a_{i,j} \tag{3.1}$$

Число α, определяемое по формуле (3.1), называется *нижней чистой ценой игры (максимином).* Оно показывает, какой минимальный выигрыш может получить игрок *A,* применяя свои чистые стратегии при любых действиях игрока *B*. Соответствующая стратегия A_0 игрока *A* называется *максиминной* [7].

Игрок *B* при наилучшем своем поведении максимально уменьшает проигрыш. Поэтому для каждой чистой стратегии B_j он отыскивает $\beta_j = \max_i a_{i,j}$ $(j = 1, \cdots n)$, а затем по β_j находит такую свою стратегию B_{j_0}, при которой его проигрыш будет минимальным, т. е.

$$\beta = \min_j \beta_j = \min_j \max_i a_{i,j}. \tag{3.2}$$

Число β, определяемое по формуле (3.2), называется *верхней чистой ценой игры (минимаксом).* Оно показывает, какой максимальный проигрыш вследствие использования чистых стратегий может быть у игрока *B*. Соответствующая чистая стратегия B_{j_0} игрока *B* называется *минимаксной* [7].

Таким образом, используя чистые стратегии, игрок *A* обеспечивает выигрыш не меньше α, а игрок *B* в результате применения своих чистых стратегий может не позволить игроку *A* выиграть больше, чем β. Принцип осторожности, диктующий игрокам выбор максиминной и минимаксной стратегий, называют *принципом минимакса.* Максиминную и минимаксную стратегии игроков для краткости иногда называют *минимаксными.*

Пример 3.1. Найти максиминную и минимаксную стратегии в игре с матрицей:

$$\begin{pmatrix} 0 & 4 & -1 & 3 \\ 1 & 0 & 2 & 2 \\ 3 & 1 & -2 & -1 \end{pmatrix}.$$

Решение. В последнем столбце табл. 3.2 выписаны минимальные по строкам элементы α_i. Наибольшим из них будет 0. Итак, максимин

$\alpha = \max_i \alpha_i = \max_i \min_j a_{i,j}$ =0, а максиминной чистой стратегией игрока *A* является A_2. В последней строке таблицы приведены максимальные элементы соответствующих столбцов β_j. Наименьшим из них является 2, значит минимакс $\beta = \min_i \beta_j = \min_j \max_i a_{i,j} = 2$, а минимаксной для игрока *B* является стратегия B_3.

Таблица 3.2 – Поиск минимаксной стратегии

	B_1	B_2	B_3	B_4	α_i
A_1	0	4	-1	3	-1
A_2	1	0	2	2	0
A_m	3	1	-2	-1	-2
β_j	3	4	2	3	

Теорема 1. В матричной игре нижняя чистая цена игры не превосходит верхней чистой цены игры, т. е. $\alpha \le \beta$.

Если в матричной игре нижняя и верхняя чистые цены игры совпадают, т. е. $\alpha = \beta$, то говорят, что эта игра имеет *седловую точку* в чистых стратегиях и *чистую цену игры* $v = \alpha = \beta$.

Обозначим i_0 и j_0 через номера чистых стратегий, при которых имеет место равенство $\alpha = \beta$. Пару чистых стратегий $(A_{i_0}; B_{j_0})$ игроков *A* и *B*, при которых достигается это равенство, называют *седловой точкой матричной игры,* а элемент $a_{i,j}$ матрицы, стоящий на пересечении *i*-строки и *j*-столбца, – *седловым элементом платежной матрицы.*

3.1.3 Решение матричных игр в смешанных стратегиях

Если игра имеет седловую точку, то оптимальными для игроков будут их минимаксные стратегии, а чистой ценой игры – седловой элемент матрицы игры. Если игра седловой точки не имеет, то решение игры затрудняется.

Обратимся к общему случаю матричной игры (табл. 3.3). Обозначим через $p_1; ...; p_m$ вероятности, с которыми игрок *A* использует чистые стратегии $A_1; ...; A_m$. Ясно, что

$$p_i \geq 0, \quad \sum_{i=1}^{m} p_i = 1 \qquad (3.3)$$

Упорядоченное множество $p = (p_1; ...; p_m)$, элементы которого удовлетворяют условиям (3.3), полностью определяет характер игры игрока *A* и называется его *смешанной стратегией.* Таким образом, смешанной стратегией игрока *A* является полный набор вероятностей применения его чистых стратегий. Ясно, что механизм случайного выбора чистых стратегий, которым пользуется игрок *A,* обеспечивает ему множество смешанных стратегий. Любая его чистая стратегия A_i может рассматриваться как частный случай смешанной стратегии, *i-я* компонента которой равна 1, а остальные равны 0, т. е. $p = (0; ...; 1; ...; 0)$.

Таблица 3.3 – Общий случай матричной игры

	B_1	B_2	...	B_n	p_i
A_1	$a_{1,1}$	$a_{1,2}$	...	$a_{1,n}$	p_1
A_2	$a_{2,1}$	$a_{2,2}$	...	$a_{2,n}$	p_2
...	...	...	...	...	...
A_m	$a_{m,1}$	$a_{m,2}$	...	$a_{m,n}$	p_m
q_j	q_1	q_2	...	q_n	

Аналогично упорядоченное множество $q = (q_1; ...; q_n)$, элементы которого удовлетворяют соотношениями ($q \geq 0 \quad \sum_{j=1}^{n} q_j = 1$ *j*=1, ..., *n*) является *смешанной стратегией игрока B*. Игрок *B* также располагает множеством смешанных стратегий.

Итак, пусть игроки *A* и *B* применяют смешанные стратегии *p* и *q*. Это означает, что игрок *A* использует стратегию A_i с вероятностью p_i а игрок *B* – стратегию B_j с вероятностью q_j. Поскольку игроки выбирают свои чистые стратегии случайно и независимо друг от друга, то вероятность выбора комбинации стратегий $(A_i; B_j)$ будет равна произведению вероятностей p_i и q_j, т. е. $p_i q_j$. При использовании смешанных стратегий игра приобретает случайный характер, случайной становится и величина выигрыша игрока *A* (проигрыша игрока *B*). Поэтому теперь можно вести речь лишь о средней величине (математическом ожидании) выигрыша. Ясно, что эта величина является функцией от смешанных стратегий *p*, *q* и определяется по формуле

$$f(p;q)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}p_iq_j\,.$$

Функция $f(p;q)$ называется *платежной функцией* игры с матрицей $\|a_{i,j}\|$.

Игрок *A,* изменяя свои смешанные стратегии *p*, стремится максимизировать средний выигрыш $f(p;q)$, а игрок *B,* изменяя свои смешанные стратегии *q*, старается сделать этот его выигрыш как можно меньше. Учитывая разумность действий противника в условиях противоположности целей, для решения игры с точки зрения игрока *A* необходимо найти такие смешанные стратегии *p* и *q*, при которых ему обеспечивался бы средний выигрыш, равный $\min\limits_q\max\limits_p f(p;q)$. Эту величину назовем *верхней ценой игры* и обозначим через β, т. е. $\beta=\min\limits_q\max\limits_p f(p;q)$

Аналогичной должна быть ситуация и относительно игрока *B:* Нижняя цена игры α должна равняться $\max\limits_p\min\limits_q f(p;q)$, т.е. $\alpha=\max\limits_p\min\limits_q f(p;q)$.

По аналогии с играми, имеющими седловые точки в чистых стратегиях, назовем *оптимальными* смешанные стратегии p^* и q^* игроков *A* и *B,* удовлетворяющие равенству [7]

$$\min_q\max_p f(p;q)=\max_p\min_q f(p;q)=f(p^*,q^*)\qquad(3.4)$$

Величину $f(p^*;q^*)$, полученную по формуле (3.4), называют *ценой игры.* Обозначим ее через ν. Итак, $\nu=f(p^*;q^*)$.

Назовем *оптимальными* смешанные стратегии p^* и q^* соответственно игроков *A* и *B,* если они образуют седловую точку для платежной функции $f(p;q)$, т. е. удовлетворяют неравенству.

$$f(p;q^*)\le f(p^*;q^*)\le f(p^*;q)\qquad(3.5)$$

Из неравенства (3.5) следует, что в седловой точке $(p^*;q^*)$ платежная функция $f(p;q)$ достигает максимума, по смешанным стратегиям p игрока A и минимума по смешанным стратегиями q игрока B.

3.1.4 Свойства оптимальных смешанных стратегий

Теорема 2. В смешанных стратегиях любая конечная матричная игра имеет седловую точку [7].

Теорема 3. Для того чтобы смешанные стратегии $p^* = (p_1^*; ...; p_m^*)$ и $q^* = (q_1^*; ...; q_n^*)$ были оптимальными для игроков A и B в игре с матрицей $\| a_{i,j} \|$ и ценой ν, необходимо и достаточно выполнение неравенств [7]

$$\sum_{i=1}^{m} a_{i,j} p_i^* \geq \nu, \tag{3.6}$$

$$\sum_{j=1}^{n} a_{i,j} q_j^* \leq \nu. \tag{3.7}$$

Итак, чтобы проверить является ли набор $(p^*; q^*; \nu)$ решением матричной игры надо найти неотрицательное решение $(p_1; ...; p_m; q_1; ...; q_n; \nu)$ системы n+m линейных неравенств (3.6)-(3.7) и уравнений (3.8)

$$\sum_{i=1}^{m} p_i = 1, \quad \sum_{j=1}^{n} q_j = 1 \tag{3.8}$$

Теорема 4. Если один из игроков придерживается своей оптимальной смешанной стратегии, то его выигрыш остается неизменным и равным цене игры независимо от того, какую стратегию применяет другой игрок, если только не выходит за пределы своих активных стратегий [7].

Решение игры можно упростить, если выявить в матрице игры доминирование одних стратегий над другими, это позволит сократить размерность матрицы.

Если в матрице игры элементы k-й не меньше соответствующих элементов s-й строки, то выигрыш игрока A при стратегии A_k будет больше, чем при стратегии A_s. Говорят, что стратегия A_k доминирует над стратегией A_s. A_k –доминирующая стратегия, A_s – доминируемая стратегия. Аналогично и для стратегии игрока B.

Теорема 5**.** Пусть I – игра, в матрице которой k-я стратегия игрока A доминирует над s-й, а I' – игра, матрица которой получена из матрицы игры I исключением s-й строки. Тогда: а) цена игры I' равна цене игры I; б) оптимальная смешанная стратегия $q^* = (q_1^*; ...; q_n^*)$ игрока B в игре I' является также его оптимальной смешанной стратегией и в игре I; в) если $p'^* = (p_1^*; ...; p_{s-1}^*; p_{s+1}^*; ...; p_m^*)$ оптимальная смешанная стратегия игрока A в игре I', то его смешанная стратегия $p^* = (p_1^*; ...; p_{s-1}^*; 0; p_{s+1}^*; ...; p_m^*)$ является оптимальной в игре I [7].

Аналогичная теорема доказывается для случая доминирования стратегий игрока B.

Теорема 6. Пусть p^* и q^* – оптимальные стратегии игроков A и B в игре I с матрицей $\| a_{i,j} \|$ с ценой игры ν. Тогда p^* и q^* будут оптимальными и в игре I' с матрицей $\| ba_{i,j} + c \|$ (где b > 0) и ценой $\nu' = b\nu + c$ [7].

3.1.5 Численные методы решения матричных игр. Связь теории игр с линейным программированием

Пусть имеем игру размерности mxn с матрицей $\| a_{i,j} \|$. Обозначим через $p^* = (p_1^*; ...; p_m^*)$ и $q^* = (q_1^*; ...; q_n^*)$ оптимальные смешанные стратегии игроков A и B. Стратегия p^* игрока A гарантирует ему выигрыш не меньше ν, независимо от выбора стратегии B_j игрока B. Это можно записать так [7]:

$$\begin{cases} a_{1,1}p_1 + a_{2,1}p_2 + ... + a_{m,1}p_m \geq \nu \\ a_{1,2}p_1 + a_{2,2}p_2 + ... + a_{m,2}p_m \geq \nu \\ .... \\ a_{1,n}p_1 + a_{2,n}p_2 + ... + a_{m,n}p_m \geq \nu \end{cases}, \qquad (3.9)$$

где $p_1 + p_2 + ... + p_m = 1; \quad p_i \geq 0 \quad (i = 1 ... m)$.

Аналогично стратегия q^* игрока B гарантирует ему проигрыш не больше ν, независимо от выбора стратегии A_i игрока A, т.е.

$$\begin{cases} a_{1,1}q_1 + a_{1,2}q_2 + \ldots + a_{1,n}q_n \le \nu \\ a_{2,1}q_1 + a_{2,2}q_2 + \ldots + a_{2,n}q_n \le \nu \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m,1}q_1 + a_{m,2}q_2 + \ldots + a_{m,n}q_n \le \nu \end{cases}, \qquad (3.10)$$

где $q_1 + q_2 + \ldots + q_n = 1; \quad q_j \ge 0 \quad (j = 1 \ldots n)$.

Поскольку элементы платежной матрицы всегда можно сделать положительными (теорема 6), то и цена игры $\nu > 0$.

Преобразуем системы (3.9) и (3.10), разделив обе части каждого неравенства на положительное число ν и введем новые обозначения: $\dfrac{p_i}{\nu} = x_i, \quad \dfrac{q_j}{\nu} = y_j \quad (i = 1 \ldots m; \quad j = 1 \ldots n)$.

Получим

$$\begin{cases} a_{1,1}x_1 + a_{2,1}x_2 + \ldots + a_{m,1}x_m \ge 1 \\ a_{1,2}x_1 + a_{2,2}x_2 + \ldots + a_{m,2}x_m \ge 1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1,n}x_1 + a_{2,n}x_2 + \ldots + a_{m,n}x_m \ge 1 \end{cases}, \qquad (3.11)$$

где $x_1 + x_2 + \ldots + x_m = \dfrac{1}{\nu}; \quad x_i \ge 0 \quad (i = 1 \ldots m)$ (3.12)

и

$$\begin{cases} a_{1,1}y_1 + a_{1,2}y_2 + \ldots + a_{1,n}y_n \le 1 \\ a_{2,1}y_1 + a_{2,2}y_2 + \ldots + a_{2,n}y_n \le 1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m,1}y_1 + a_{m,2}y_2 + \ldots + a_{m,n}y_n \le 1 \end{cases} \qquad (3.13)$$

где $y_1 + y_2 + \ldots + y_n = \dfrac{1}{\nu}; \quad y_j \ge 0 \quad (j = 1 \ldots n)$. (3.14).

Так как игрок A стремится максимизировать цену игры ν, то обратная величина $\dfrac{1}{\nu}$ будет минимизироваться, поэтому оптимальная стратегия игрока A определится из задачи линейного программирования следующего вида: найти минимальное значение целевой функции $f(x) = x_1 + x_2 + \ldots + x_m$ при ограничениях (3.11) и (3.12).

Оптимальная смешанная стратегия игрока B определится решением задачи следующего вида: найти максимальное значение целевой функции $\widetilde{f}(y) = y_1 + y_2 + ... + y_n$ при ограничениях (3.13), (3.14).

Решив пару двойственных задач графическим (для случая двух переменных) или симплексным методом, далее определим:

$$\nu = \frac{1}{\sum_{i=1}^{m} x_i^*}; \quad p_i = \frac{x_i^*}{\sum_{i=1}^{m} x_i^*}; \quad q_j = \frac{y_j^*}{\sum_{i=1}^{m} x_i^*} \quad (i = 1 \ldots m, \quad j = 1 \ldots n).$$

Проиллюстрируем решение матричной игры сведением ее к ЗЛП.

Пример 3.2. Два сельскохозяйственных предприятий *A* и *B* выделяют денежные средства на строительство трех объектов. С учетом особенностей вкладов в строительство и местных условий прибыль предприятия *A* в зависимости в зависимости от объемов финансирования выражается элементами матрицы $\begin{bmatrix} 50 & 15 & 20 \\ 25 & 40 & 30 \\ 10 & 30 & 60 \end{bmatrix}$.

Будем предполагать, что убыток предприятия *B* при этом равен прибыли предприятия *A*. Требуется найти оптимальные стратегии предприятий *A* и *B*.

<u>Решение</u>. Обозначим чистые стратегии предприятий *A* и *B* через A_1, A_2, A_3 и B_1, B_2, B_3 соответственно. Предположим, что предприятие *A* располагает общей суммой *a* тыс. ден. ед., отпускаемой на строительство трех объектов. Аналогично и предприятие *B* имеет сумму в *b* тыс. ден. ед., отпускаемых на строительство тех же трех объектов. Тогда чистая стратегия A_1 – это выделение a_1 тыс. ден. ед. предприятия *A* на строительство первого объекта; A_2 – это выделение a_2 тыс. ден. ед. предприятия *A* на строительство второго объекта; A_3 – это выделение a_3 тыс. ден. ед. предприятия *A* на строительство третьего объекта. Общая сумма средств, выделяемых на строительство трех объектов, $a = a_1 + a_2 + a_3$. Аналогично определяются чистые стратегии и для предприятия *B*.

Проверим игру на наличие седловой точки: $\alpha = \max_i \min_j a_{i,j} = 25$, $\beta = \min_j \max_i a_{i,j} = 40$, $\alpha \neq \beta$, поэтому решение игры определим в смешанных стратегиях. Цена игры ν заключена между нижней α и верхней β ценами, т.е. $25 \leq \nu \leq 40$.

Составим ЗЛП для каждого игрока.

<u>*Для игрока А*</u>: найти минимальное значение функции $f = x_1 + x_2 + x_3$ при ограничениях:

$$\begin{cases} 50x_1 + 25x_2 + 10x_3 \ge 1 \\ 15x_1 + 40x_2 + 30x_3 \ge 1 \\ 20x_1 + 30x_2 + 60x_3 \ge 1 \end{cases}$$

$$x_i \ge 0, \quad (i = 1, 2, 3).$$

<u>*Для игрока В:*</u> найти максимальное значение функции $\widetilde{f} = y_1 + y_2 + y_3$ при ограничениях:

$$\begin{cases} 50y_1 + 15y_2 + 20y_3 \le 1 \\ 25y_1 + 40y_2 + 30y_3 \le 1 \\ 10y_1 + 30y_2 + 60y_3 \le 1 \end{cases}$$

$$y_j \ge 0, \quad (j = 1, 2, 3).$$

Вводя вспомогательные переменные $x_4 \ge 0, x_5 \ge 0, x_6 \ge 0$ для исходной задачи и $y_4 \ge 0, y_5 \ge 0, y_6 \ge 0$ для двойственной задачи, модели задач преобразуем к канонической форме. Соответствие между переменными пары двойственных задач будут следующими

$$\begin{array}{cccccccc} \multicolumn{3}{c}{\textit{Свободные}} & & \multicolumn{3}{c}{\textit{Базисные}} \\ x_1 & x_2 & x_3 & & x_4 & x_5 & x_6 \\ \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \updownarrow & \updownarrow \\ y_4 & y_5 & y_6 & & y_1 & y_2 & y_3 \\ \multicolumn{3}{c}{\textit{Базисные}} & & \multicolumn{3}{c}{\textit{Свободные}} \end{array}$$

Решим, например, двойственную ЗЛП, построенную для определения проигрыша предприятия *В*. Каноническая форма задачи имеет вид:

$$\max : \widetilde{f} = y_1 + y_2 + y_3$$

$$\begin{cases} 50y_1 + 15y_2 + 20y_3 + y_4 = 1 \\ 25y_1 + 40y_2 + 30y_3 + y_5 = 1 \\ 10y_1 + 30y_2 + 60y_3 + y_6 = 1 \end{cases}$$

$$y_j \ge 0, \quad (j = 1 \ldots 6).$$

Решив ее симплекс-методом, приходим к таблице 3.4, в которой содержится оптимальный план $y^* = \left(y_1^*, y_2^*, y_3^*, y_4^*, y_5^*, y_6^*\right)$ =(0,0133; 0,0094; 0,0098;0;0;0). При этом $\widetilde{f}(y^*) = 0{,}0325$.

С учетом основной теоремы двойственности и соответствия между переменными оптимальный план исходной задачи запишется в виде $x^* = \left(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*\right)$=(0,0102;0,0180;0,0043;0;0;0) и $f(x^*) = 0{,}0325$.

По формулам $\nu = \dfrac{1}{f_{\max}}$, $\dfrac{p_i}{\nu} = x_i$, $\dfrac{q_j}{\nu} = y_j$ $(i = 1...m, \quad j = 1...n)$ получим цену игры $\nu = \dfrac{1}{0{,}0325} \approx 30{,}77$ и вероятности p_i^*, q_j^* для оптимальных смешанных стратегий соответственно предприятий A и B:

$$p_1^* = 0{,}0102 \cdot 30{,}77 = 0{,}314 \quad q_1^* = 0{,}0133 \cdot 30{,}77 = 0{,}409$$
$$p_2^* = 0{,}0180 \cdot 30{,}77 = 0{,}554 \quad q_2^* = 0{,}0094 \cdot 30{,}77 = 0{,}289.$$
$$p_3^* = 0{,}0043 \cdot 30{,}77 = 0{,}132 \quad q_3^* = 0{,}0098 \cdot 30{,}77 = 0{,}302$$

Таблица 3.4 – Последняя симплекс-таблица

БП	**СЧ**	$-y_1$	$-y_2$	$-y_3$	$-y_4$	$-y_5$	$-y_6$
y_1	0,0133	1	0	0	0,0234	-0,0047	-0,0055
y_2	0,0094	0	1	0	-0,0188	0,0438	-0,0156
y_3	0,0098	0	0	1	0,0056	-0,0211	0,0254
$\tilde{f}$	0,0325	0	0	0	0,0102	0,0180	0,0043

Таким образом, оптимальными смешанными стратегиями сельскохозяйственных предприятий A и B являются стратегии p^*=(0,314; 0,554; 0,132) и q^*=(0,409; 0,289; 0,302).

Итак, из общей суммы средств a тыс. ден. ед., выделенных предприятием A на строительство трех объектов, на долю первого объекта должно выделяться 31,4%, второго – 55,4% и третьего -13,2% этой суммы. Аналогично распределяются b тыс. ден. ед. предприятием B. Так на долю первого объекта должно выделяться 40,9%, второго – 28,9% и третьего – 30,2% этой суммы. Такое распределение денежных средств предприятиями A и B по трем строящимся объектам позволит им получить максимальную прибыль 30,77 тыс. ден. ед.

3.2 Игры с природой. Критерии для принятия решений

3.2.1 Игры с природой

Управление производственными процессами осуществляется путем реализации последовательности принимаемых решений. Для этого необходима информация о состоянии объекта управления в условиях его работы. В случае отсутствия достаточно полной информации возникает неопределенность в принятии решения.

С целью уменьшения неблагоприятных последствий в каждом конкретном случае следует учитывать степень риска и имеющуюся информацию. И здесь *лицо, принимающее решение* (ЛПР), вступает в игровые отношения с некоторым абстрактным лицом, которое условно можно назвать *«природой».* Иными словами, ЛПР должно уметь находить управленческое решение, когда природа не выбирает сознательно свои оптимальные стратегии. Вместе с тем мы иногда располагаем некоторыми вероятностными характеристиками состояния природы. Такого рода ситуации принято называть *играми с природой* [7, 8].

Любую хозяйственную деятельность человека можно рассматривать как игру с природой. В широком смысле под «природой» будем понимать совокупность неопределенных факторов, влияющих на эффективность принимаемых решений.

Задачей ЛПР является принятие наилучшего управленческого решения в каждой конкретной ситуации. Качество принимаемого решения зависит от информированности ЛПР о ситуации, в которой принимается решение для обоснования принимаемых решений – это задача экономиста.

Безразличие природы к результату игры (выигрышу) и возможность получения экономистом (статистиком) или ЛПР дополнительной информации о ее состоянии отличают игру с природой от обычной матричной игры, в которой принимают участие два сознательных игрока.

Игры с природой представляют собой основную модель теории принятия решений в условиях частичной неопределенности.

Множество состояний природы обозначим через *П*, отдельное состояние – $П_j$, $j = 1...n$. Множество решений (стратегий) статистика обозначим через *А*, отдельное решение – A_i, $i = 1...m$.

Для i-го решения A_i статистика A и j-го состояния природы $П_{\mathrm{j}}$ имеем некоторое число, обозначающее функцию потерь $L(A_i, П_{\mathrm{j}})$, которая, как правило, является случайной переменной.

Во взаимоотношениях с природой статистик может использовать любые из стратегий $A_1, A_2, ..., A_m$ в зависимости от состояний $П_{\mathrm{j}}$ природы. Имея ряд стратегий $A_1, A_2, ..., A_m$, статистик должен руководствоваться некоторым правилом поведения, с помощью которого он определяет, какую стратегию A_i ему выбрать. Иными словами, статистик отыскивает оптимальное поведение, которое и будет его оптимальной стратегией. При этом он может пользоваться как чистыми, так и смешанными стратегиями.

Чтобы выразить в количественной форме упомянутое выше некоторое правило поведения статистика, которым он должен руководствоваться, предположим, что есть возможность численно оценить величиной $a_{i,j}$ эффективность каждой комбинации $(A_i, П_{\mathrm{j}})$, иначе говоря, качество решения A_i. Тем самым будет определена так называемая платежная матрица игры с природой $\left\| a_{i,j} \right\|_{m \times n}$, на основе которой в дальнейшем и будут сформулированы «правила поведения» – критерии выбора оптимальной стратегии статистика.

Элемент $a_{i,j}$ назовем выигрышем статистика, если он использует стратегию A_i при состоянии природы $П_{\mathrm{j}}$.

Решение игры с природой несколько отличается от решения обычной матричной игры, где оба игрока ведут игру сознательно. Отличие состоит прежде всего в упрощении игры. Выявление дублирующих и доминируемых стратегий производится только для стратегий статистика. Стратегии природы нельзя опускать, поскольку она не имеет «умысла» навредить статистику, более того, она может реализовать состояния, заведомо выгодные статистику. Иногда при решении игры с природой используется *матрица рисков* [7]. Элементы $r_{i,j}$ матрицы рисков равны разности между максимально возможным выигрышем и тем выигрышем, который статистик получит в тех же условиях $П_j$, применяя стратегию A_i т. е. $r_{i,j} = \beta_j - a_{i,j}$, где $\beta_j = \max\limits_i a_{i,j}$.

3.2.2 Критерии для принятия решений

Оптимальную стратегию статистика можно определить, используя ряд критериев. Так, при известном распределении вероятностей различных состояний $П_j$ природы пользуются *критерием Байеса* [8]. Показателем в этом критерии служит либо величина среднего выигрыша, либо величина среднего риска.

Платежную матрицу $\|a_{i,j}\|_{m \times n}$ представим в виде табл. 3.5.

По критерию Байеса за оптимальную принимается та чистая стратегия A_i, при которой максимизируется средний выигрыш $\bar{a}_i$ статистика, т. е. обеспечивается $\bar{a} = \max\limits_i \bar{a}_i$, где $\bar{a}_i = \sum\limits_{j=1}^{n} a_{i,j} q_j$, $i = 1...m$.

Таблица 3.5 – Платежная матрица

Стратегия статистика A_i	**Состояние природы** $П_j$				**Средний выигрыш**
A_1	$a_{1,1}$	$a_{1,2}$	…	$a_{1,n}$	$\bar{a}_1$
A_2	$a_{2,1}$	$a_{2,2}$	…	$a_{2,n}$	$\bar{a}_2$
…	…	…	…	…	…
A_m	$a_{m,1}$	$a_{m,2}$	…	$a_{m,n}$	$\bar{a}_m$
q_i	q_1	q_2	…	q_n	

Матрицу рисков представим в виде табл. 3.6. За оптимальную стратегию статистика принимается чистая стратегия A_i, при которой минимизируется средний риск, т.е. обеспечивается $\bar{r} = \min\limits_i \bar{r}_i$, где $\bar{r}_i = \sum\limits_{j=1}^{n} r_{i,j} q_j$, $i = 1...m$.

В случае, когда вероятности состояний природы правдоподобны, для их оценки используют *принцип недостаточного основания*

Лапласа [8], согласно которому все состояния природы полагаются равновероятными, т. е. оптимальной считается стратегия, обеспечивающая максимум среднего выигрыша.

Если вероятности состояний природы неизвестны, то для решения игр с природой – выбора оптимальной стратегии статистика – можно использовать несколько критериев.

Максиминный критерий Вальда совпадает с критерием выбора максиминной стратегии, позволяющей получать нижнюю чистую цену α в парной игре с нулевой суммой. По критерию Вальда за оптимальную принимается чистая стратегия, которая в наихудших условиях гарантирует максимальный выигрыш [8], т. е. $\alpha = \max\limits_i \min\limits_j a_{i,j}$.

Таблица 3.6 – Матрица рисков

Стратегия статистика A_i	**Состояние природы** Π_j				**Средний риск**
A_1	$r_{1,1}$	$r_{1,2}$	…	$r_{1,n}$	$\bar{r}_1$
A_2	$r_{2,1}$	$r_{2,2}$	…	$r_{2,n}$	$\bar{r}_2$
…	…	…	…	…	…
A_m	$r_{m,1}$	$r_{m,2}$	…	$r_{m,n}$	$\bar{r}_m$
q_i	q_1	q_2	…	q_n	

Критерий минимального риска Сэвиджа [8] рекомендует выбирать в качестве оптимальной стратегии ту, при которой величина максимального риска минимизируется в наихудших условиях, т. е. обеспечивается $\min\limits_i \max\limits_j r_{i,j}$.

Критерии Вальда и Сэвиджа ориентируют статистика на самые неблагоприятные состояния природы, т.е. эти критерии выражают пессимистическую оценку ситуации.

Критерий Гурвица является критерием пессимизма-оптимизма [8]. За оптимальную принимается та стратегия, для которой выполняется соотношение $\max\limits_i \left(\lambda \min\limits_j a_{i,j} + (1-\lambda) \max\limits_j a_{i,j} \right)$, где $0 \le \lambda \le 1$.

При $\lambda = 0$ имеем критерий крайнего оптимизма, а при $\lambda = 1$ – критерий пессимизма Вальда. Если $0 < \lambda < 1$, то имеем нечто среднее. При желании подстраховаться в данной ситуации λ принимают близким к единице. В общем случае число λ выбирают исходя из опыта или субъективных соображений.

Решение игры с природой по рассмотренным критериям позволяет более обоснованно принимать ту стратегию, которая гарантирует статистику больший выигрыш по сравнению с выигрышем, принимаемым статистиком интуитивно или исходя из опыта.

Пример 3.3. В соответствии со спросом на продукцию *q-й* номенклатуры в городе планируется построить предприятие по производству этой продукции. Неопределенность спроса в период t приводит к тому, что необходимо рассчитать объем выпускаемой продукции V_q, который должен быть не меньше уровня спроса S_q, чтобы не потерять потенциально возможный доход от реализации продукции, а также не должен превышать уровень спроса, так как предприятие будет нести убытки, связанные в основном с уценкой. Предположим, что в течение года (по кварталам) спрос на продукцию *q-й* номенклатуры выражается величинами 10, 20, 30, 40 тыс. шт. В таком случае и планирующий орган предприятия может принять одно из следующих решений: построить предприятие, которое могло бы удовлетворить спрос потребителей в 10, 20, 30, 40 тыс. шт. *q-й* продукции. Работа подобных предприятий показывает, что предприятие терпит издержки от нереализованной единицы *q-й* продукции 5 ден. ед., а доход от реализации единицы продукции составляет 15 ден. ед. Функцию платежей можно записать в виде кусочно-линейной функции потерь:

$$L(S_q, V_q) = \begin{cases} K'_q(V_q - S_q), & V_q \ge S_q \\ K''_q(S_q - V_q), & V_q < S_q \end{cases}.$$

Требуется: 1) придать описанной ситуации игровую схему, установить характер игры и выявить ее участников; 2) вычислить элементы платежной матрицы и составить ее; 3) дать обоснованные рекомендации планирующему органу на строительство предприятия, которое могло бы обеспечить спрос потребителей на q-ю продукцию.

При изучении работы аналогичных предприятий планирующий орган располагает некоторой дополнительной информацией, снижающей неопределенность ситуации: известны вероятности спроса на данную продукцию по кварталам года: 0,3; 0,2; 0,4; 0,1; спрос на продукцию в каждом квартале равновероятен; о вероятностях спроса на указанную продукцию по кварталам ничего определенного сказать нельзя.

Решение. В качестве статистика выступает планирующий орган, который может принять одно из следующих решений: построить предприятие, способное удовлетворить спрос потребителей в 10 тыс. ед. продукции (стратегия A_1)*;* построить предприятие мощностью в 20 тыс. ед. продукции (стратегия A_2); построить предприятие мощностью в 30 тыс. ед. продукции (стратегия A_3); построить предприятие мощностью в 40 тыс. ед. продукции (стратегия A_4). Второй играющей стороной – природой – будем считать совокупность объективных внешних условий, в которых формируется спрос потребителей. Спрос по кварталам года различен: в первом квартале – 10 тыс. ед. будет означать состояние $П_1$; спрос на продукцию во втором квартале в объеме 20 тыс. ед. – состояние $П_2$; спрос на продукцию в третьем квартале в объеме 30 тыс. ед. – состояние $П_3$; спрос на продукцию в четвертом квартале в объеме 40 тыс. ед. – состояние $П_4$. Итак, описанная ситуация представляет собой игру с природой.

Рассчитаем элементы платежной матрицы (табл. 3.7).

Так, в ситуации $(A_1, П_1)$ элемент $a_{1,1}$ вычисляется следующим образом. Плановый орган принимает решение построить предприятие мощностью в 10 тыс. ед., что и соответствует состоянию спроса в 10 тыс. ед. Доход от производства 10 тыс. ед. продукции $a_{1,1} = 10 \cdot 15 = 150$ тыс. ден. ед. Элемент $a_{1,2}$ в ситуации $(A_1, П_2)$ рассчитываем так. Предприятие строится в расчете на выпуск 10 тыс. ед. продукции, а спрос на нее составляет 20 тыс. ед. Если бы предприятие могло обеспечить этот спрос, то доход составил бы $20 \cdot 15 = 300$ тыс. ден. ед. Однако спрос удовлетворяется лишь на 10 тыс. ед. продукции, следовательно, предприятие не дополучит доход $10 \cdot 15 = 150$ тыс. ден. ед. Элемент

$a_{1,2} = 20 \cdot 15 - 10 \cdot 15 = 150$. Аналогично определяются и другие элементы табл. 3.7, например, элемент $a_{3,1}$ для ситуации (A_3, Π_1). Предприятие строится в расчете на выпуск 30 тыс. ед. продукции, а спрос на нее составляет 10 тыс. ед., тогда доход предприятия от реализации 10 тыс. ед. составит $10 \cdot 15 = 150$ тыс. ден. ед., а от нереализованной продукции (20 тыс. ед.) предприятие терпит издержки $20 \cdot 5 = 100$ тыс. ден. ед., доход предприятия в ситуации (A_3, Π_1) составит $a_{3,1} = 10 \cdot 15 - 20 \cdot 5 = 50$ тыс. ден. ед.

Таблица 3.7 – Расчет платежной матрицы

Мощность предприятия, тыс. ед.	**Стратегия статистка** A_i	**Спрос потребителей по кварталам года**, тыс. ед.			
		10	20	30	40
		Состояние спроса Π_j			
		$\Pi_1(10)$	$\Pi_2(20)$	$\Pi_3(30)$	$\Pi_4(40)$
10	A_1	150	150	150	150
20	A_2	100	300	300	300
30	A_3	50	250	450	450
40	A_4	0	200	400	600
	q_i	0,3	0,2	0,4	0,1

Вычисляем средние выигрыши:

$\overline{a}_1 = 150 \cdot 0{,}3 + 150 \cdot 0{,}2 + 150 \cdot 0{,}4 + 150 \cdot 0{,}1 = 150$.

Аналогично, $\overline{a}_2 = 240, \overline{a}_3 = 290, \overline{a}_4 = 260$.

Оптимальной стратегией по Байесу является A_3 поскольку ей соответствует максимальная средняя прибыль $\overline{a}_3 = \max(150, 240, 290, 260)$ тыс. ден. ед.

По критерию Лапласа, когда средние выигрыши равны: $q_1 = q_2 = q_3 = q_4 = \dfrac{1}{4}$,

$$\overline{\alpha}_1 = \frac{1}{4}(150 + 150 + 150 + 150) = 150, \quad \overline{\alpha}_2 = 250, \quad \overline{\alpha}_3 = 300, \quad \overline{\alpha}_4 = 300.$$

Оптимальными стратегиями по Лапласу являются A_3 и A_4, так как им соответствует максимальная прибыль, равная 300 тыс. ден. ед.

По критерию Вальда оптимальной является стратегия A_1, для которой прибыль достигает наибольшей величины, равной 150 тыс. ден. ед. ($\alpha = \max_i \min_j a_{i,j} = \max(150, 100, 50, 0) = 150$).

Построим матрицу рисков (табл. 3.8). *По критерию Сэвиджа оптимальными являются* стратегии A_3 и A_4, для которых в наихудших условиях величина r риска принимает наименьшее значение, равное 150 тыс. ден. ед. В самом деле, $r = \min_i \max_j r_{i,j} = \min(450, 300, 150, 150) = 150$.

В результате решения игры с природой чаще других рекомендовалась стратегия A_3. Следовательно, нужно строить предприятие мощностью в 30 тыс. ед. продукции. Прибыль при этом, если вероятности спроса известны, составит 290 тыс. ден. ед., при равновероятных условиях спроса – 300 тыс. ден. ед.

Таблица 3.8 – Матрица рисков

Мощность предприятия, тыс. ед.	**Стратегия статистика** A_i	Спрос потребителей по кварталам года, тыс. ед.				$\max_j r_{i,j}$
		10	20	30	40	
		Состояние спроса $П_j$				
		$П_1(10)$	$П_2(20)$	$П_3(30)$	$П_4(40)$	
10	A_1	0	150	300	450	450
20	A_2	50	0	150	300	300
30	A_3	100	50	0	150	150
40	A_4	150	100	50	0	150

Вопросы для самопроверки

1. По каким признакам можно классифицировать игры?
2. Матричные игры.
3. Когда игра имеет седловую точку?
4. Что такое смешанные стратегии для игрока А и игрока В?
5. Применение матричной игры к задаче линейного программирования.
6. Какого рода ситуации принято называть играми с природой?
7. Перечислите критерии принятия решений для игр с природой.

4 ЗАДАЧИ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Задачи массового обслуживания условно делят на задачи анализа и задачи синтеза – оптимизации систем массового обслуживания. Задачи анализа предполагают оценку эффективности функционирования системы массового обслуживания при неизменных наперед заданных исходных характеристик системы: структуре системы, дисциплине обслуживания, потоках требований и законах распределения времени их обслуживания [1, 2]. Задачи синтеза направлены на поиск оптимальных параметров систем массового обслуживания.

Систему массового обслуживания в общем виде можно представить как совокупность последовательно связанных между собой входящих потоков требований на обслуживание, очередей, каналов обслуживания и выходящих потоков требований (рис. 4.1).

Случайный характер входящего потока требований (машин, станков, самолетов, пользователей и т.д.), а также длительности обслуживания каналом (станция техобслуживания, аэродром, ЭВМ и т.д.) приводит к образованию случайного процесса в системе, который необходимо исследовать.

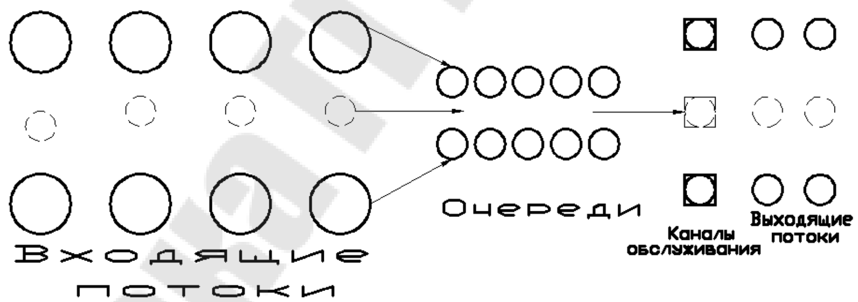

Рисунок 4.1 – Схема системы массового обслуживания (СМО)

4.1 Классификация систем массового обслуживания и их основные характеристики

Системы массового обслуживания по наличию того или иного признака можно разделить следующим образом [1, 2]:

1. *По характеру поступления требований* на системы с регулярным и случайным потоками поступления требований в систему.

Случайный поток требований в систему подразделяется на стационарный и нестационарный:

- если количество поступающих требований в систему в единицу времени (интенсивность потока) постоянно или является заданной функцией времени, то имеем систему с регулярным потоком поступления требований в систему, в противном случае – со случайным;
- если параметры потока требований не зависят от расположения рассматриваемого интервала времени на оси времени, то имеем стационарный поток требований, в противном случае – нестационарный. Например, если число машин, приходящих на склад, не зависит от времени суток, то поток требований – стационарный.

2. *По количеству поступающих требований в один момент времени* на системы с ординарным и неординарным потоками требований. Если вероятность поступления двух или более требований в один момент равна нулю или имеет столь малую величину, что ею можно пренебречь, то имеем систему с ординарным потоком требований.

3. *По связи между требованиями* на системы без последействия от поступивших требований и с последействием. Если вероятность поступления требований в систему в некоторый момент не зависит от того, сколько уже требований поступило в систему, т. е. не зависит от предыстории изучаемого процесса, то мы имеем задачу без последействия, в противном случае – с последействием.

4. *По характеру поведения требования* в системе с отказами, с ограниченным ожиданием и с ожиданием без ограничения:

- если вновь поступившее требование на обслуживание застает все каналы обслуживания уже занятыми и оно покидает систему, то имеем систему с отказами. Требование может покинуть систему и в том случае, когда очередь достигла определенных размеров;
- если поступившее требование застает все каналы обслуживания занятыми и становится в очередь, но находится в ней ограниченное время, после чего, не дождавшись обслуживания, покидает систему, то имеем систему с ограниченным ожиданием;
- если поступившее требование, застав все каналы обслуживания занятыми, вынуждено ожидать своей очереди до тех пор, пока

оно не будет обслужено, то имеем систему с ожиданием без ограничения.

5. *По способу выбора требований на обслуживание* с приоритетом, по мере поступления, случайно, последний обслуживается первым. Иногда, в этом случае говорят о дисциплине обслуживания:

- если система массового обслуживания охватывает несколько категорий требований и по каким-либо соображениям необходимо соблюдать различный подход к их отбору, то имеем систему с приоритетом;
- если освободившийся канал обслуживает требование, ранее других поступившее в систему, то имеем систему с обслуживанием требований по мере их поступления;
- если требования из очереди в канал обслуживания поступают в случайном порядке, то имеем систему со случайным выбором требований на обслуживание;
- последний обслуживается первым. Этот способ выбора требований на обслуживание используется в тех случаях, когда удобнее или экономнее брать но обслуживание требование, позже всех поступившее в систему.

6. *По характеру обслуживания требований* на системы с детерминированным и случайным временем обслуживания. Если интервал времени между моментом поступления требования в канал обслуживания и моментом выхода требования из этого канала постоянно, то имеем систему с детерминированным временем обслуживания, в противном случае – со случайным.

7. *По числу каналов обслуживания* на одноканальные и многоканальные системы.

8. По количеству этапов обслуживания на однофазные и многофазные системы. Если каналы обслуживания расположены последовательно и они неоднородны, так как выполняют различные операции обслуживания, то имеем многофазную систему массового обслуживания.

9. *По однородности требований,* поступающих на обслуживание, на системы с однородными и неоднородными потоками требований.

10. *По ограниченности потока требований* на замкнутые и разомкнутые системы. Если поток требований ограничен и требования, покинувшие систему, могут в нее возвращаться, то имеем замкнутую систему, в противном случае – разомкнутую.

Если изучены или заданы входящие потоки требований, механизм (число каналов обслуживания, продолжительность обслуживания и т. д.) и дисциплина обслуживания, то это дает основание для построения математической модели системы.

4.2 Задачи анализа одноканальных систем массового обслуживания

4.2.1 Задача анализа детерминированной системы

Постановка задачи. Пусть исследуется производственный процесс, в котором поступление требований происходит через равные промежутки времени $\Delta t_{\text{п}} = const$ (т. е. интенсивность потока поступлений требований $\lambda = 1/\Delta t_{\text{п}} = const$) и обслуживание производится через равные промежутки времени $\Delta t_{об} = const$ (т.е. интенсивность обслуживания $\mu = 1/\Delta t_{об} = const$). Имеется один канал обслуживания. Предполагается, что $\Delta_{об} / \Delta t_{\text{п}} = \lambda / \mu < 1$ (в противном случае очередь будет бесконечно возрастать) и что к началу обслуживания в системе имеется уже n требований. Определить через какое время очередь исчезнет [6].

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Величину $\psi = \lambda / \mu$ называют коэффициентом использования. Очередь будет бесконечно возрастать, если $\psi > 1$, если же $\psi = 1$, то очередь будет иметь постоянную длину. Схематически работа рассматриваемой системы массового обслуживания представлена на рис. 4.2.

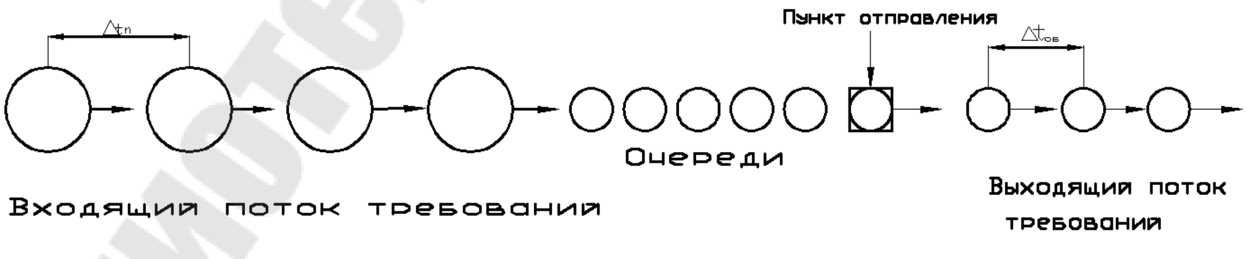

Рисунок 4.2 – Схема работы системы

Пока обслуживается очередь из n требований в течение времени $t = n \cdot \Delta t_{об}$ вновь поступит на обслуживание n_1 требований:

$$n_1 = \frac{t}{\Delta t_{п}} = \frac{n\Delta t_{об}}{\Delta \mathrm{t}_{п}} = n\frac{\lambda}{\mu} = n\psi .$$

Аналогично пока будут обслуживаться n_1 требований в течение времени $t_1 = n_1 \Delta t_{об}$, дополнительно поступит на обслуживание n_2 требований:

$$n_2 = \frac{t_1}{\Delta t_{п}} = \frac{n_1\Delta t_{об}}{\Delta \mathrm{t}_{п}} = n_1\frac{\lambda}{\mu} = n\psi^2 .$$

Это происходит до тех пор, пока $t_к > \Delta t_{п}$ после чего очередь исчезнет.

Построение математической модели. Время, через которое очередь исчезнет, можно представить в таком виде:

$$T = t + t_1 + ... + t_k .$$

Исследование математической модели. Для определения времени, через которое очередь исчезнет, необходимо раскрыть математическую модель.

$$T = n\Delta t_{об} + n_1 \Delta t_{об} + ... + n_k \Delta t_{об} = \frac{n(1-\psi^{k+1})}{\mu(1-\psi)} .$$

В модели использована формула суммы геометрической прогрессии. Чем ближе интенсивность потока λ к интенсивности обслуживания μ, тем через больший промежуток времени исчезнет очередь (при $\frac{\lambda}{\mu} < 1$). Членом ψ^{k+1} можно для упрощения расчетов пренебречь, тогда

$$T \approx \frac{n}{\mu - \lambda} .$$

4.2.2 Задача анализа разомкнутой системы с ожиданием (потоки требований пуассоновские)

Постановка задачи. Пусть задана некоторая система массового обслуживания, для которой справедливы следующие гипотезы:

- вероятность поступления требований не зависит от принятого начала отсчета времени, а зависит только от продолжительности периода наблюдений (стационарность потока);

- не поступают в систему и не покидают ее одновременно два или более требований (поток ординарный);
- поступление одного требования не зависит от поступления другого (отсутствие последствия).

Известны также интенсивность λ поступления потока требований (среднее число поступлений требований в в единицу времени $\lambda = 1/\Delta t_{п}$ и интенсивность μ обслуживания требований (среднее число обслуживаний в единицу времени $\mu = 1/\Delta t_{об}$. Требуется определить основные характеристики системы [6]:

- вероятность простоя канала обслуживания P_0;
- вероятность того, что в системе находится *n* требований, P_n;
- среднее число требований, находящихся в системе, $N_{сист}$(в очереди и обслуживании);
- среднее число требований, находящихся в очереди, $N_{оч}$;
- среднее время ожидания требования в системе $T_{сист}$.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Поток требований, обладающий свойством стационарности, ординарности и отсутствием последействия, называется *простейшим*. В нашей задаче поток требований простейший. Основным понятием при анализе процесса систем массового обслуживания является состояние системы. Зная состояние системы, можно предсказать в вероятностном смысле ее поведение.

Простейший поток – это стационарный пуассоновский поток. Если все потоки событий, переводящие систему из одного состояния в другое, являются пуассоновскими, то для этих систем вероятности состояний описываются с помощью системы обыкновенных дифференциальных уравнений [1].

В большинстве задач прикладного характера замена непуассоновских потоков событий пуассоновскими с теми же интенсивностями приводит к получению решения, которое мало отличается от истинного, а иногда и совсем не отличается. В качестве критерия небольшого отличия реального стационарного потока от пуассоновского можно рассматривать близость математического ожидания и дисперсии числа событий, поступающих на определенном участке времени в реальном потоке.

Существует определенный методический прием, намного облегчающий вывод дифференциальных уравнений для вероятностей состояний. Первоначально строится размеченный граф состояний (рис. 4.3) с указанием возможных переходов – это облегчает исследование и делает его более наглядным. Граф состояний, на котором проставлены не только стрелки переходов, но и интенсивность соответствующих потоков событий, называют *размеченным*.

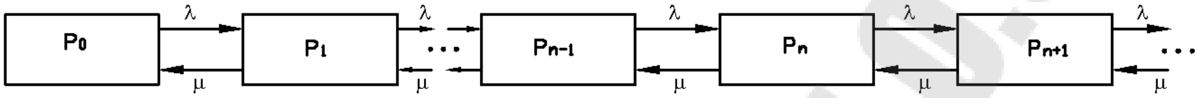

Рисунок 4.3 – Размеченный граф состояний одноканальной разомкнутой СМО с ожиданием

Построение математической модели. Если составлен размеченный граф состояний, то для построения математической модели, т.е. для составления системы обыкновенных дифференциальных уравнений вероятностей состояний, рекомендуется использовать, следующее мнемоническое правило.

Производная $\frac{dP_n(t)}{dt}$ вероятности пребывания системы в состоянии n равна алгебраической сумме нескольких членов:

- число членов этой суммы равно числу стрелок на графе состояний системы, соединяющих состояние n с другими состояниями;
- если стрелка направлена в состояние n, то член берется со знаком плюс;
- если стрелка направлена из состояния n, то со знаком минус;
- каждый член суммы равен произведению вероятностей того состояния, из которого направлена стрелка, на интенсивность потока событий, переводящего систему по данной стрелке.

В соответствии с размеченным графом состояний, используя мнемоническое правило, систему обыкновенных дифференциальных уравнений вероятностей состояний запишем так:

$$\frac{dP_0(t)}{dt} = -P_0(t)\lambda + P_1(t)\mu;$$

....

$$\frac{dP_n(t)}{dt} = P_{n-1}(t)\lambda - P_n(t)(\lambda + \mu) + P_{n+1}(t)\mu.$$

Исследование математической модели. Ограничимся установившегося режима работы разомкнутой одноканальной системы. Тогда

$$\frac{dP_n(t)}{dt} = 0 \quad (n = 0, 1, ...).$$

Вместо системы обыкновенных дифференциальных уравнений получим систему алгебраических уравнений:

$$-P_0\lambda + P_1\mu = 0;$$

$$P_0\lambda - P_1(\lambda + \mu) + P_2\mu = 0;$$

....

$$P_{n-1}\lambda - P_n(\lambda + \mu) + P_{n+1}\mu = 0.$$

Используя полученную систему алгебраических уравнений, легко выразить вероятности состояний системы в виде некоторой рекуррентной формулы.

Из первого уравнения определяется вероятность наличия одного требования в системе

$$P_1 = P_0 \frac{\lambda}{\mu} = \psi P_0;$$

из второго уравнения – вероятность наличия двух требований в системе

$$P_2 = \psi^2 P_0;$$

аналогично из третьего уравнения – вероятность наличия трех требований в системе

$P_3 = \psi^3 P_0$ и т.д.

Суммируя полученные значения для $P_0, P_1, ..., P_n$ и используя формулу суммы геометрической прогрессии, получаем

$$\sum_{i=0}^{n} P_i = P_0 + \psi P_0 + ... + \psi^n P_0 = P_0(1 + \psi + ... + \psi^n) = P_0 \frac{1 - \psi^{n+1}}{1 - \psi}.$$

При $n \to \infty$ $(\psi < 1)$ $\sum_{i=0}^{\infty} P_i = P_0 \frac{1}{1-\psi} = 1$, откуда имеем:

- вероятность простоя канала обслуживания $P_0 = 1 - \psi$;
- вероятность того, что в системе находится n требований,

$$P_n = \psi^n P_0 = \psi^n (1 - \psi);$$

- среднее число требований, находящихся в системе (или математическое ожидание)

$$N_{сист} = \sum_{n=1}^{\infty} n P_n = \frac{\psi}{1 - \psi};$$

- среднее число требований, находящихся в очереди,

$$N_{оч} = \frac{\lambda}{\mu} N_{сист} = \frac{\psi^2}{1 - \psi};$$

- среднее время ожидания требования в системе, которое можно определить, зная среднее число требований, находящихся в системе,

$$T_{сист} = \frac{N_{сист}}{\lambda} = \frac{1}{\mu(1 - \psi)}.$$

4.2.3 Задача анализа замкнутой системы с ожиданием (потоки требований пуассоновские)

Постановка задачи. Пусть исследуется некоторая замкнутая система массового обслуживания с ограниченным количеством требований в системе, т.е. обслуженные требования вновь возвращаются в систему обслуживания (например, экскаватор или автосамосвал). Интенсивность поступления одного требования в систему известна и равна λ. Интенсивность обслуживания требований известна и равна μ. Число требований, нуждающихся в обслуживании равно m. Требуется определить основные характеристики системы [6]:

- вероятность того, что в системе находится n требований, P_n;
- среднее число требований, находящихся в очереди, $N_{оч}$;
- среднее число требований, находящихся в системе, $N_{сист}$;
- среднее время ожидания требования в системе $T_{сист}$;
- среднее время ожидания требования в очереди $T_{оч}$.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Состояние системы будем связывать с числом требований, находящихся в системе. При этом возможны два состояния системы:

- число требований, поступивших в систему, $n = 0$, т.е. канал обслуживания простаивает;
- число требований, поступивших в систему, $0 < n \le m$.

Размеченный граф состояний системы представлен на рис. 4.4.

Построение математической модели. В соответствии с размеченным графом состояний и используя мнемоническое правило, запишем систему дифференциальных уравнений для вероятностей состояний

$$\frac{dP_0(t)}{dt} = -m\lambda P_0(t) + \mu P_1(t);$$

....

$$\frac{dP_n(t)}{dt} = -[(m-n)\lambda + \mu]P_n(t) + (m-n+1)\lambda P_{n-1}(t) + \mu P_{n+1}(t);$$

....

$$\frac{dP_m(t)}{dt} = -\mu P_m(t) + \lambda P_{m-1}(t).$$

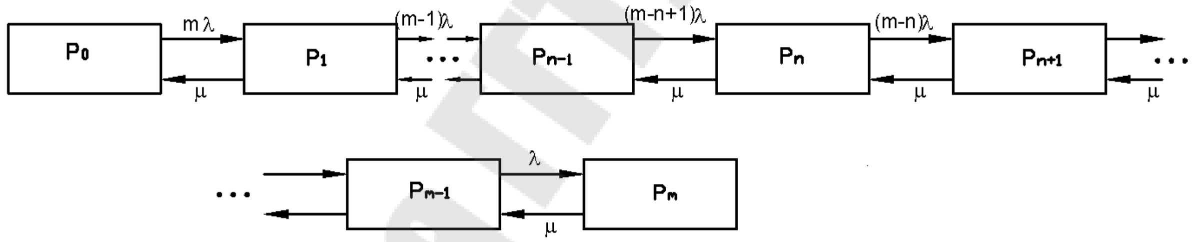

Рисунок 4.4 – Размеченный граф состояний одноканальной замкнутой системы массового обслуживания с ожиданием

Исследование и решение математической модели. Ограничимся установившегося режима работы разомкнутой одноканальной системы. Тогда

$$\frac{dP_n(t)}{dt} = 0 \quad (n = 0, 1, ..., m).$$

Вместо системы обыкновенных дифференциальных уравнений получим систему алгебраических уравнений:

$$m\lambda P_0 - \mu P_1 = 0;$$

....

$$[(m-n)\lambda + \mu]P_n - (m-n+1)\lambda P_{n-1} - \mu P_{n+1} = 0;$$

....

$$\mu P_m - \lambda P_{m-1} = 0.$$

Для $0 < n \le m$ нетрудно получить рекуррентную формулу:

- при n=0 $P_1 = m\psi P_0$;
- при n=1 $P_2 = (m-1)\psi P_1$;
- при n=2 $P_3 = (m-2)\psi P_2$;

…………………………………

- при n-1 $P_n = (m-n+1)\psi P_{n-1}$.

Вероятность того, что в системе находится n требований составит,

$$P_n = (m-n+1)\psi P_{n-1} = \frac{m!\psi^n}{(m-n)!}P_0.$$

Используя равенство $\sum_{n=0}^{m} P_n = \sum_{n=0}^{m} \frac{m!\psi^n}{(m-n)!}P_0 = 1$, можно получить выражение для P_0. Вероятность простоя канала обслуживания

$$P_0 = \left[1 + \sum_{n=1}^{m} \frac{m!\psi^n}{(m-n)!}\right]^{-1}.$$

Среднее число требований, находящихся в очереди,

$$N_{оч} = \sum_{n=2}^{m}(n-1)P_n = m - \frac{1+\psi}{\psi}(1-P_0).$$

Среднее число требований, находящихся в системе,

$$N_{сист} = \sum_{n=1}^{m} nP_n = m - \frac{1}{\psi}(1-P_0).$$

Среднее время ожидания требования в очереди

$$T_{оч} = \frac{N_{оч}}{\lambda(m - N_{оч})} = \frac{1}{\mu}\left[\frac{m}{1-P_0} - \frac{1+\psi}{\psi}\right].$$

Среднее время ожидания требования в системе

$$T_{сист} = \frac{N_{сист}}{\lambda(m - N_{сист})} = \frac{1}{\mu}\left[\frac{m}{1-P_0} - \frac{1}{\psi}\right].$$

4.3 Задачи анализа многоканальных систем массового обслуживания

4.3.1 Задача анализа разомкнутой системы с ожиданием (потоки требований пуассоновские)

Постановка задачи. Пусть известны интенсивность λ поступления потока требований в систему и интенсивность μ обслуживания этих требований. Число каналов обслуживания N_k. Требуется определить [6]:

- вероятность того, что в системе имеется *n* требований, $P_n(t)$;
- вероятность простоя каналов обслуживания $P_0(t)$;
- среднее число требований, находящихся в очереди, $N_{оч}$;
- среднее время ожидания требования в очереди $T_{оч}$.
- среднее число свободных каналов обслуживания $N_{ск}$.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Возможны два случая:

- число требований в системе *n* меньше числа каналов: $0 \le n < N_k$;
- число требований больше или равно числу каналов $n \ge N_k$.

В первом случае все требования, находящиеся в системе, одновременно обслуживаются и не все каналы заняты. Общая интенсивность обслуживания будет равна $n\mu$ (индекс *k* опустим).

Размеченный граф состояний системы представлен на рис. 4.5.

Рисунок 4.5 – Размеченный граф состояний многоканальной разомкнутой системы массового обслуживания с ожиданием

Построение математической модели. В соответствии с размеченным графом состояний и используя мнемоническое правило, запишем систему дифференциальных уравнений для вероятностей состояний системы:

для $n = 0 \quad \frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t);$

……………………

для $1 < n < N \quad \frac{dP_n(t)}{dt} = P_{n-1}(t)\lambda - (\lambda + n\mu)P_n(t) + (n+1)\mu P_{n+1}(t);$

…………………………………

для $n \ge N \quad \frac{dP_n(t)}{dt} = P_{n-1}(t)\lambda - (\lambda + N\mu)P_n(t) + N\mu P_{n+1}(t).$

Исследование и решение математической модели. Ограничимся установившегося режима работы системы $\lambda = const, \quad \mu = const$ при $t \to \infty$, $P_n = const$ и $\frac{dP_n(t)}{dt} = 0 \quad (n = 0, 1, ...)$ и вместо системы обыкновенных дифференциальных уравнений получим систему алгебраических уравнений:

для $n = 0 \quad -\lambda P_0 + \mu P_1 = 0;$

……………………

для $1 < n < N \quad P_{n-1}\lambda - (\lambda + n\mu)P_n + (n+1)\mu P_{n+1} = 0;$

…………………………………

для $n \ge N \quad P_{n-1}\lambda - (\lambda + N\mu)P_n + N\mu P_{n+1} = 0.$

Используя полученные алгебраические выражения, определяем рекуррентные выражения для определения вероятности нахождения системы в состоянии n, P_n:

для n=0 $P_1 = \frac{\lambda}{\mu} P_0 = \psi P_0$;

для n=1 $P_2 = \frac{\psi^2}{2} P_0$;

……………

для n=N-1 $P_N = \frac{\psi^N}{N!} P_0$.

Из этих выражений видно, что для $n < N$ вероятность нахождение в системе n требований определяется по следующей формуле:

$$P'_n = \frac{\psi^n}{n!} P_0.$$

Для состояний системы $n \ge N$ имеем следующие выражения:

для n=N $P_{N+1} = \dfrac{\psi}{N}\dfrac{\psi^N}{N!}P_0$;

для n=N+1 $P_{N+2} = \dfrac{\psi^2}{N^2}\dfrac{\psi^N}{N!}P_0$.

Из полученных выражений видно, что для состояний системы $n \ge N$ вероятность нахождение в системе n требований определяется по следующей формуле:

$$P''_n = \frac{\psi^n}{N^{n-N}N!}P_0.$$

Имея аналитические выражения для всех состояний системы, а также используя очевидное равенство

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{N-1} P'_n + \sum_{n=N}^{\infty} P''_n = 1,$$

определяем вероятность простоя канала обслуживания P_0:

$$P_0\left(\sum_{n=0}^{N-1}\frac{\psi^n}{n!} + \frac{\psi^N}{N!}\sum_{n=N}^{\infty}\frac{\psi^{n-N}}{N^{n-N}}\right) = 1.$$

Если $\dfrac{\psi}{N} < 1$, то $\displaystyle\sum_{n=N}^{\infty}\frac{\psi^{n-N}}{N^{n-N}} = \frac{1}{1-\psi/N}$ и вероятность простоя каналов обслуживания

$$P_0 = \left(\sum_{n=0}^{N-1}\frac{\psi^n}{n!} + \frac{\psi^N}{N!(1-\psi/N)}\right)^{-1}.$$

Среднее число требований, находящихся в очереди,

$$N_{оч} = \sum_{n=N}^{\infty}(n-N)P''_n = \frac{\psi^{N+1}}{N!N(1-\psi/N)^2}P_0;$$

$$T_{оч} = \frac{N_{оч}}{\lambda};\ N_{ск} = \sum_{n=0}^{N-1}(N-n)P_n = \sum_{n=0}^{N-1}(N-n)\frac{\psi^n}{n!}P_0.$$

4.3.2 Задача анализа разомкнутой системы с отказом требований пуассоновские)

Постановка задачи. Пусть имеется некоторая разомкнутая система массового обслуживания. Пусть известны интенсивность поступления потока требований в систему и интенсивность μ

облуживания каждым каналом. Если требование застало все N каналов занятыми, то оно получает отказ и покидает систему.

Требуется определить [6]:

- вероятность P_0 того, что все каналы свободны;
- вероятность P_n того, что занято n каналов обслуживания;
- среднее число занятых каналов обслуживания $N_{зк}$.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Состояние системы будем связывать с числом занятых каналов обслуживания. Перечислим возможные N состояний системы:

- все каналы свободны, ни одно требование не обслуживается;
- один канал занят, обслуживается одна заявка;
- n каналов занято, обслуживаются n требований;
- ……………………;
- все N каналов заняты, обслуживается N требований.

Размеченный граф состояний системы приведен на рис. 4.6.

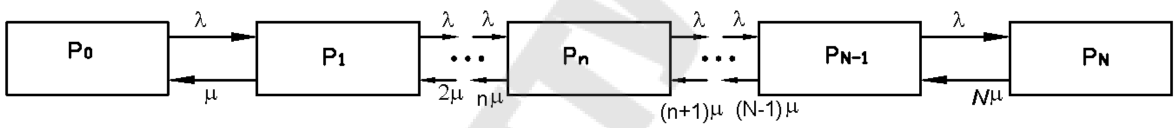

Рисунок 4.6 – Размеченный граф состояний многоканальной разомкнутой СМО

Построение математической модели. В соответствии с размеченным графом состояний и используя мнемоническое правило, запишем систему дифференциальных уравнений для вероятностей состояний:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t);$$

………………………..;

$$\frac{dP_n(t)}{dt} = -(\lambda + n\mu)P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t);$$

……………………….;

$$\frac{dP_N(t)}{dt} = -N\mu P_N(t) + \lambda P_{N-1}(t).$$

Они называются уравнения Эрланга.

Исследование и решение математической модели. Исследуем стационарный режим работы системы при $t \to \infty$, т.е. установившийся режим работы.

Система рекуррентных алгебраических уравнений будет выглядеть:

$$-\lambda P_0 + \mu P_1 = 0;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots;$$
$$-(\lambda + n\mu)P_n + \lambda P_{n-1} + (n+1)\mu P_{n+1} = 0;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots;$$
$$-N\mu P_N + \lambda P_{N-1} = 0.$$

Из первого уравнения

$$P_1 = \frac{\lambda}{\mu} P_0 = \psi P_0;$$

аналогично из второго с учетом первого равенства находим:

$$P_2 = \frac{\psi^2}{2} P_0;\ \ldots;\ P_n = \frac{\psi^n}{n!} P_0 .$$

Используя полученные соотношения, можно определить вероятность P_0 того, что все каналы обслуживания свободны:

$$\sum_{n=0}^{N} P_n = \sum_{n=0}^{N} \frac{\psi^n}{n!} P_0 = P_0 \sum_{n=0}^{N} \frac{\psi^n}{n!} = 1, \text{ откуда } P_0 = \left(\sum_{n=0}^{N} \frac{\psi^n}{n!} \right)^{-1} .$$

Вероятность того, что занято ровно n каналов обслуживания,

$$P_n = \frac{\psi^n}{n! \sum_{n=0}^{N} \frac{\psi^n}{n!}} .$$

Среднее число занятых каналов обслуживания

$$N_{зк} = \sum_{n=1}^{N} n P_n = \sum_{i=1}^{N} \frac{\psi^n}{(n-1)! \sum_{n=0}^{N} \frac{\psi^n}{n!}} .$$

4.3.3 Задача анализа замкнутой системы с ожиданием (потоки требований пуассоновские)

Постановка задачи. Пусть имеется некоторая замкнутая система массового обслуживания, в которой обслуженные требования вновь возвращаются в систему. Интенсивность поступления одного требования в систему известна и равна λ, интенсивность обслуживания каждого канала известна и равна μ. Число каналов обслуживания *N*, число требований, нуждающихся в обслуживании *m*. Будем считать, что $N \le m$. Требуется определить [6]:

- вероятность того, что в системе находятся *n* требований, $P_n(t)$;
- вероятность простоя каналов обслуживания $P_0(t)$;
- среднее число требований, ожидающих начала обслуживания $N_{оч}$;
- среднее время ожидания требования в очереди $T_{оч}$.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Состояние системы будем связывать с числом требований, находящихся в системе. При этом возможны два случая:

- число требований *n*, поступивших в систему, меньше числа каналов обслуживания, т.е. все они находятся на обслуживании $0 \le n < N$;
- число требований *n*, поступивших в систему, больше или равно числу каналов обслуживания $n \ge N$, из них *N* обслуживаются, а *r* требований ожидает в очереди $(r = 1, 2, ..., m - N)$.

Граф состояний системы представлен на рис. 4.7.

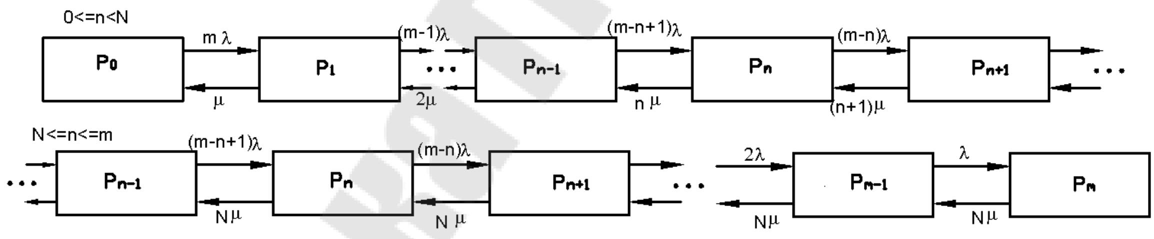

Рисунок 4.7 – Размеченный граф состояний многоканальной замкнутой СМО

Построение математической модели. В соответствии с размеченным графом состояний и используя мнемоническое правило, запишем систему дифференциальных уравнений для вероятностей состояний:

$$\text{для } 0 \le n < N \qquad \frac{dP_n(t)}{dt} = -m\lambda P_0(t) + \mu P_1(t);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots;$$

$$\frac{dP_n(t)}{dt} = -[(m-n)\lambda + n\mu]P_n(t) + (m-n+1)\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t);$$

……………………………………………;

для $N \le n \le m$

$$\frac{dP_n(t)}{dt} = -[(m-n)\lambda + N\mu]P_n(t) + (m-n+1)\lambda P_{n-1}(t) + N\mu P_{n+1}(t);$$

……………………………………………..;

$$\frac{dP_m(t)}{dt} = -N\mu P_m(t) + \lambda P_{m-1}(t).$$

Исследование математической модели. Для установившегося режима работы системы $\lambda = const$, $\mu = const$ при $t \to \infty$, $P_n = const$ и $\frac{dP_n(t)}{dt} = 0$ $(n = 0, 1, ..., m)$ и вместо системы обыкновенных дифференциальных уравнений получаем систему рекуррентных алгебраических уравнений, из которых находим вероятность того, что в системе находится *n* требований:

- для $0 \le n < N$ $\quad P'_n = \frac{m!\psi^n}{n!(m-n)!}P_0;$
- для $N \le n \le m$ $\quad P''_n = \frac{m!\psi^n}{N^{n-N}N!(m-n)!}P_0.$

Для определения вероятности простоя каналов обслуживании P_0 используется равенство

$$\sum_{n=0}^{m} P_n = \sum_{n=0}^{N-1} P'_n + \sum_{n=N}^{m} P''_n = \sum_{n=0}^{N-1}\frac{m!\psi^n}{n!(m-n)!}P_0 + \sum_{n=N}^{m}\frac{m!\psi^n}{N^{n-N}N!(m-n)!}P_0 = 1,$$

откуда

$$P_0 = \left(\sum_{n=0}^{N-1}\frac{m!\psi^n}{n!(m-n)!} + \sum_{n=N}^{m}\frac{m!\psi^n}{N^{n-N}N!(m-n)!}\right)^{-1}.$$

Среднее число требований, ожидающих начала обслуживания (длина очереди):

$$N_{оч} = \sum_{n=N}^{m}(n-N)P''_n = \sum_{n=N}^{m}\frac{(n-N)m!\psi^n}{N^{n-N}N!(m-n)!}P_0.$$

Среднее время ожидания требований в очереди

$$T_{оч} = \frac{N_{оч}}{\mu(N - N_{ск}}.$$

Среднее число свободных каналов обслуживания

$$N_{ск} = \sum_{n=0}^{N-1}(N-n)P''_n = \sum_{n=0}^{N-1}\frac{(N-n)m!\psi^n}{n!(m-n)!}P_0.$$

Можно заметить, что для определения основных характеристик многоканальных систем массового обслуживания, так же как и для одноканальных требуется произвести большой объем вычислений. Для эффективного выполнения необходимых расчетов целесообразно использование современных ЭВМ.

4.4 Задача синтеза (оптимизации) одноканальной замкнутой системы массового обслуживания с ожиданием

Постановка задачи и выбор критерия оптимизации. Пусть исследуется одноканальная замкнутая система массового обслуживания, для которой известны характеристики канала обслуживания и характеристики требований, поступающих на обслуживание. Требуется определить оптимальную структуру системы, т.е. оптимальное число требований $m_{опт}$, необходимых для обслуживания канала, чтобы эффективность системы была максимальна [6]. В качестве критерия оптимизации примем удельные приведенные затраты, характеризующие затраты всей системы на одно обслуживание.

Аналитически выражение критерия оптимизации для определения оптимальной структуры одноканальной замкнутой системы массового обслуживания с ожиданием запишется так:

$$Y = \frac{P_0 C_{пк} + (1-P_0)C_{рк} + mC_{пт}}{\mu(1-P_0)} + \frac{E_н(S_к + mS_т)}{T_г\mu(1-P_0)} \text{ (руб./ед. продукции)},$$

где P_0 – вероятность простоя канала обслуживания; m – число требований, нуждающихся в обслуживании; $C_{пк}$ – средние затраты при простое канала обслуживания в течение часа из-за несвоевременного поступления требований на обслуживание, руб.; $C_{рк}$ – средние затраты при работе канала обслуживания в течение часа, руб; μ – интенсивность обслуживания канала, 1/ч; $\mu = 1/t_{об}$; $t_{об}$ – среднее время обслуживания требования каналом, ч; $C_{пт}$ – средние затраты содержания требования в течение часа, руб; $S_к, S_т$ – капитальные вложения соответственно на канал обслуживания и требование, руб; $T_г$ – годовой режим работы – число часов работы системы в году; $E_н$ – нормативный коэффициент эффективности.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. При определении оптимальной структуры системы массового обслуживания можно предположить:

- вероятность поступления определенного количества требований в систему зависит только от продолжительности периода поступления, а не от расположения этого периода (часа работы) в смене, т.е. поток поступления требований на обслуживание стационарный;
- число требований, поступивших в систему в некоторый момент времени, не зависит от числа поступивших до этого времени;
- вероятность поступления двух или более требований в один момент времени столь малая величина, что ею можно пренебречь, а значит поток требований можно считать ординарным.

Используя ранее полученные зависимости (см. п.4.2.3), получаем:

- вероятность того, что в системе на обслуживании находится n требований,

$$P_n = \frac{m!\psi^n}{(m-n)!} P_0,$$

где $\psi = \dfrac{\lambda}{\mu}$ называют коэффициентом использования; λ – интенсивность поступления требования 1/ч;

- вероятность простоя канала

$$P_0 = \left(1 + \sum_{n=1}^{m} \frac{m!\psi^n}{(m-n)!}\right)^{-1}.$$

Построение математической модели. Аналитическое выражение целевой функции с учетом результатов, полученных выше, запишется в виде

$$Y = \frac{C_{рк} - C_{пк}}{\mu} + \left[C_{пк} + mC_{пт} + \frac{E_н}{T_г}(S_к + S_m m)\right] \cdot \frac{1}{\mu\left[1 - \left(1 + \sum_{n=1}^{m} \frac{m!\psi^n}{(m-n)!}\right)^{-1}\right]}$$

Таким образом, при известных условиях работы системы необходимо определить оптимальную структуру системы, т.е. такое количество требований *m*, которое минимизирует величину критерия оптимизации.

Исследование математической модели. Запишем критерий оптимизации в ином виде, обозначив первое слагаемое через Y_1 и преобразовав второе слагаемое:

$$Y(m) = Y_1 + \frac{C_{пк} + \dfrac{S_к E_н}{T_г} + m(C_{пт} + \dfrac{S_т E_н}{T_г})}{\mu[1 - P_0(m)]}.$$

Как можно заметить, первое слагаемое не зависит от *m*.

Для определения минимального значения целевой функции положим, что он достигается для значения $m = m_{опт}$, тогда $Y(m_{опт} - 1) > Y(m_{опт})$; $Y(m_{опт} + 1) > Y(m_{опт})$.

Найдем соответствующие значения $Y(m-1)$ и $Y(m+1)$:

$$Y(m-1) = Y_1 + \frac{C_{пк} + \dfrac{S_к E_н}{T_г} + (m-1)(C_{пт} + \dfrac{S_т E_н}{T_г})}{\mu[1 - P_0(m-1)]};$$

$$Y(m+1) = Y_1 + \frac{C_{пк} + \dfrac{S_к E_н}{T_г} + (m+1)(C_{пт} + \dfrac{S_т E_н}{T_г})}{\mu[1 - P_0(m+1)]};$$

$$Y(m-1) - Y(m) = -\frac{C_{пт} + \dfrac{S_т E_н}{T_г}}{\mu[1 - P_0(m-1)]} + \frac{C_{пк} + \dfrac{S_к E_н}{T_г} + m\left[C_{пт} + \dfrac{S_т E_н}{T_г}\right]}{\mu} \times$$

$$\times \left[\frac{1}{1 - P_0(m-1)} - \frac{1}{1 - P_0(m)}\right] \geq 0.$$

Аналогично определяется

$$Y(m+1) - Y(m) = \frac{C_{пт} + \dfrac{S_т E_н}{T_г}}{\mu[1 - P_0(m+1)]} + \frac{C_{пк} + \dfrac{S_к E_н}{T_г} + m\left[C_{пт} + \dfrac{S_т E_н}{T_г}\right]}{\mu} \times$$

$$\times \left[\frac{1}{1 - P_0(m+1)} - \frac{1}{1 - P_0(m)}\right] \geq 0.$$

Окончательно получаем следующее неравенство, используя которое можно легко определить оптимальную структуру одноканальной замкнутой системы, зная соответствующую входную информацию. Значение m, которое удовлетворяет неравенству, обеспечивает минимальное значение критерия оптимальности – целевой функции.

Чтобы определить оптимальное значение m, достаточно протабулировать значения неравенства

$$\frac{1}{1-P_0(m-1)}\left(1-\frac{1}{C+m}\right) > \frac{1}{1-P_0(m)} < \frac{1}{1-P_0(m+1)}\left(1+\frac{1}{C+m}\right),$$

где C – коэффициент затрат:

$$C=\frac{C_{пк}+\dfrac{E_н}{T_г}S_к}{C_{пт}+\dfrac{E_н}{T_г}S_т}.$$

4.5 Задача синтеза (оптимизации) многоканальной замкнутой системы массового обслуживания с ожиданием

Постановка задачи и выбор критерия оптимизации. Пусть исследуется некоторая многоканальная замкнутая система массового обслуживания. Известны характеристики каналов обслуживания и характеристики требований, поступающих на обслуживание. Требуется определить такую структуру многоканальной системы, чтобы эффективность системы была максимальна. В качестве критерия оптимизации принимаем целевую функцию – удельные приведенные затраты, т.е. затраты, приходящиеся на одно обслуживание [6]. Используем обозначения и допущения те же, что и для одноканальной замкнутой системы.

Аналитическое выражение критерия оптимизации для определения оптимальной структуры многоканальной замкнутой системы массового обслуживания с ожиданием запишется так:

$$Y=\frac{P_0C_{пк}N+(1-P_0)C_{рк}N+NC_{пт}}{\mu(1-P_0)}+\frac{E_н(S_кN+mS_т)}{T_г\mu(1-P_0)},$$

где N – число каналов обслуживания в системе.

Выявление основных особенностей, взаимосвязей и количественных закономерностей. Используя ранее полученные зависимости (см. п.4.3.3), представляем искомые аналитические выражения в таком виде:

- вероятность того, что в системе на обслуживании находится n требований:

для $0 \le n < N \quad P_n = \dfrac{m!\psi^n}{n!(m-n)!}P_0$;

для $N \le n \le m \quad P_n = \dfrac{m!\psi^n}{N^{n-N}N!(m-n)!}P_0$.

- вероятность простоя канала обслуживания из-за отсутствия требований в системе

$$P_0 = \left(1 + \sum_{n=1}^{N-1}\frac{m!\psi^n}{n!(m-n)!} + \sum_{n=N}^{m}\frac{m!\psi^n}{N^{n-N}N!(m-n)!}\right)^{-1}.$$

Построение математической модели. Аналитическое выражение целевой функции с учетом полученных результатов

$$Y = \frac{C_{рк}N - C_{пк}N}{\mu} + \frac{C_{пк}N + mC_{пт} + \dfrac{E_н}{T_г}S_к N + \dfrac{E_н}{T_г}S_т m}{\mu\left[1 - \left(1 + \sum\limits_{n=1}^{N-1}\dfrac{m!\psi^n}{n!(m-n)!} + \sum\limits_{n=N}^{m}\dfrac{m!\psi^n}{N^{n-N}N!(m-n)!}\right)^{-1}\right]}.$$

Таким образом, имея развернутое выражение целевой функции при известных условиях работы системы, можно определить оптимальную структуру многоканальной системы, т.е. оптимальное число требований, функционирующих в системе.

Исследование математической модели. Для определения минимального значения целевой функции положим, что он достигается для значения $m = m_{опт}$, тогда $Y(m_{опт} - 1) > Y(m_{опт})$; $Y(m_{опт} + 1) > Y(m_{опт})$.

Первое слагаемое целевой функции Y_1 не зависит от *m*. Запишем целевую функцию в таком виде:

$$Y = Y_1 + \frac{(C_{пк} + \dfrac{E_н}{T_г}S_к)N + m(C_{пт} + \dfrac{E_н}{T_г}S_т)}{\mu[1 - P_0(m)]}.$$

Для определения оптимального числа требований найдем соответствующие значения:

$$Y(m-1) = Y_1 + \frac{(C_{пк} + \dfrac{E_н}{T_г}S_к)N + (m-1)(C_{пт} + \dfrac{E_н}{T_г}S_т)}{\mu[1 - P_0(m-1)]};$$

$$Y(m+1) = Y_1 + \frac{(C_{пк} + \frac{E_н}{T_г} S_к)N + (m+1)(C_{пт} + \frac{E_н}{T_г} S_m))}{\mu\left[1 - P_0(m+1)\right]}.$$

Обозначим отношение затрат работы канала обслуживания к требованию через $C = \dfrac{C_{пк} + \dfrac{E_н S_к}{T_г}}{C_{пт} + \dfrac{E_н S_m}{T_г}}$, и проведя некоторые преобразования, окончательно получим

$$\frac{1}{1 - P_0(m-1)}\left(1 - \frac{1}{CN+m}\right) > \frac{1}{1 - P_0(m)} < \frac{1}{1 - P_0(m+1)}\left(1 + \frac{1}{CN+m}\right).$$

Вопросы для самопроверки

1. Приведите примеры систем массового обслуживания (СМО).
2. Что может быть каналами обслуживания?
3. Для чего предназначена всякая СМО?
4. Что изучает предмет теория массового обслуживания?
5. Дайте классификацию систем массового обслуживания.
6. Задача анализа детерминированной системы. Постановка задачи.
7. Задача анализа разомкнутой системы с отказом требований пуассоновские). Постановка задачи.
8. Задача анализа разомкнутой многоканальной системы с ожиданием (потоки требований пуассоновские). Постановка задачи.
9. Задача анализа разомкнутой многоканальной системы с отказом требований пуассоновские). Постановка задачи.
10. Задача анализа замкнутой многоканальной системы с ожиданием (потоки требований пуассоновские). Постановка задачи.
11. Задача синтеза (оптимизации) одноканальной замкнутой системы массового обслуживания с ожиданием.
12. Задача синтеза (оптимизации) многоканальной замкнутой системы массового обслуживания с ожиданием.

5 ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ И ТЕХНИЧЕСКИХ СИСТЕМ

Основой современного технического прогресса является опережающее развитие машиностроения, осуществляемое по следующим главным направлениям:

- существенное улучшение и разработка новых технологических процессов производства деталей и сборки машин;
- создание систем высокопроизводительного технологического оборудования, оснастки и инструмента;
- повышение автоматизации на всех этапах производственного процесса.

Принятие проектных решений в машиностроении и оценка их качества в основном осуществляются на основании данных эксперимента.

Задача извлечения наибольшего объема информации об изучаемых процессах или устройствах при ограничениях по затратам является в настоящее время достаточно актуальной. Решению указанной проблемы способствует широкое внедрение в практику прикладных исследований статистических методов планирования экспериментов, которые дают не только способ обработки экспериментальных данных, но позволяют также оптимально организовывать эксперимент.

5.1 Эксперимент как предмет исследования

В технической литературе термину эксперимент устанавливается следующее определение – система операций, воздействий и (или) наблюдений, направленных на получение информации об объекте исследования.

Являясь источником познания и критерием истинности теорий и гипотез, эксперимент играет очень важную роль как в науке, так и в инженерной практике. Эксперименты ставятся в исследовательских лабораториях и на действующем производстве, в медицинских клиниках и на опытных сельскохозяйственных полях, в космосе и в глубинах океана.

Прежде всего отметим, что любой эксперимент предполагает проведение тех или иных опытов.

Опыт – воспроизведение исследуемого явления в определенных условиях проведения эксперимента при возможности регистрации его результатов.

По цели проведения и форме представления полученных результатов эксперимент делят на качественный и количественный.

Качественный эксперимент устанавливает только сам факт существования какого-либо явления, но при этом не дает никаких количественных характеристик объекта исследования. Любой эксперимент, каким бы сложным он ни был, всегда заканчивается представлением его результатов, формулировкой выводов, выдачей рекомендаций. Эта информация может быть выражена в виде графиков, чертежей, таблиц, формул, статистических данных или словесных описаний. Качественный эксперимент как раз и предусматривает именно словесное описание его результатов.

Однако словесное описание – не самый эффективный и информативный способ представления результатов эксперимента, поскольку он не позволяет дать количественных рекомендаций, проанализировать свойства объекта в иных условиях.

Количественный эксперимент не только фиксирует факт существования того или иного явления, но, кроме того, позволяет установить соотношения между количественными характеристиками явления и количественными характеристиками способов внешнего воздействия на объект исследования.

Итак, количественный эксперимент прежде всего предполагает количественное определение всех тех способов внешнего воздействия на объект исследования, от которых зависит его поведение – количественное описание всех факторов.

Фактор – переменная величина, по предположению влияющая на результаты эксперимента.

В отдельном конкретном опыте каждый фактор может принимать одно из возможных своих значений – уровень фактора.

Уровень фактора – фиксированное значение фактора относительно начала отсчета.

Фиксированный набор уровней всех факторов в каждом конкретном опыте как раз и определяет одно из возможных состояний объекта исследования.

При проведении опытов очень многое зависит от того, насколько активно экспериментатор может «вмешиваться» в исследуемое явление, имеет он или нет возможность устанавливать те уровни факторов, которые представляют для него интерес.

С этой точки зрения все факторы можно разбить на три группы:

- *контролируемые и управляемые* – это факторы, для которых можно не только зарегистрировать их уровень, но еще и задать в каждом конкретном опыте любое его возможное значение;
- *контролируемые, но неуправляемые факторы* – это факторы, уровни которых можно только регистрировать, а вот задать в каждом опыте их определенное значение практически невозможно;
- *неконтролируемые* – это факторы, уровни которых не регистрируются экспериментатором и о существовании которых он даже может и не подозревать.

В количественном эксперименте необходимо не только регистрировать уровни всех контролируемых факторов, но и иметь возможность устанавливать количественное описание того свойства (отклика) исследуемого явления, которое изучает (наблюдает) экспериментатор. Причем поскольку на объект исследования в процессе эксперимента всегда влияет огромное количество неконтролируемых факторов, что вносит в получаемые результаты некоторый элемент неопределенности, значение отклика, в каждом конкретном опыте, невозможно предсказать заранее. Поэтому воспроизведение исследуемого явления при одном и том же фиксированном наборе уровней всех контролируемых факторов всегда будет приводить к различным значениям отклика, т.е. отклик – это всегда случайная величина.

Отклик – наблюдаемая случайная переменная, по предположению зависящая от факторов.

И наконец, в результате количественного эксперимента необходимо найти зависимость между откликом и факторами – функцию отклика. Причем поскольку отклик – это случайная величина, то, с точки зрения теории вероятностей, его можно задать одним из параметров своего распределения, например математическим ожиданием.

Функция отклика – зависимость математического ожидания отклика от факторов.

С учетом приведенного выше деления факторов на три группы, функцию отклика в самом общем случае можно записать в виде:

$$M_y = f(x_i, h_j) + \varepsilon_\delta \qquad (5.1)$$

где M_y – математическое ожидание отклика; x_i – контролируемые и управляемые факторы; h_j – контролируемые, но неуправляемые факторы; ε_δ – ошибка эксперимента, учитывающая влияние неконтролируемых факторов.

По тому, какой группой факторов располагает исследователь, количественный эксперимент в свою очередь можно разделить еще на два вида. Если в распоряжении экспериментатора нет управляемых факторов, то такой эксперимент носит название пассивного.

Пассивный эксперимент – эксперимент, при котором уровни факторов в каждом опыте регистрируются исследователем, но не задаются.

Поскольку при пассивном эксперименте исследователь не имеет возможность задать уровень ни одного из факторов, то при проведении опытов ему остается лишь «пассивно» наблюдать за явлением и регистрировать результаты. Планирование пассивного эксперимента сводится к определению числа опытов, которые необходимо провести исследователю для решения поставленной перед ним задачи, а конечной целью пассивного эксперимента в большинстве случаев является получение функции отклика в виде

$$M_y = f(h_j) + \varepsilon_\delta \qquad (5.2)$$

Если же экспериментатор имеет возможность не только контролировать факторы, но и управлять ими, то такой эксперимент носит название активного.

Активный эксперимент – эксперимент, в котором уровни факторов в каждом опыте задаются исследователем.

Поскольку в этом случае экспериментатор имеет возможность «активно» вмешиваться в исследуемое явление, то естественно, что активный эксперимент всегда предполагает какой-либо план его проведения.

План эксперимента – совокупность данных, определяющих число, условия и порядок реализации опытов.

Поэтому активный эксперимент всегда должен начинаться с планирования.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего поставленным требованиям.

Целью активного эксперимента может быть либо определение функции отклика в виде

$$M_y = f(x_i) + \varepsilon_\delta \tag{5.3}$$

либо поиск такого сочетания уровней управляемых факторов x_i, при котором достигается оптимальное (экстремальное – минимальное или максимальное) значение функции отклика. В этом последнем случае эксперимент носит еще название поискового (экстремального) эксперимента.

И наконец, по условиям проведения различают лабораторный и промышленный эксперименты.

Лабораторный эксперимент. В лаборатории меньше влияние случайных погрешностей, обеспечивается большая «стерильность» условий проведения опытов, в большинстве случаев осуществляется и более тщательная подготовка, одним словом, выше «культура эксперимента». Как правило, в лабораторных условиях экспериментатор может воспроизвести опыт «одинаково» значительно лучше, чем в промышленности. Другое важное отличие – это большая возможность варьировать (изменять) уровни факторов.

Промышленный эксперимент. В промышленных условиях обеспечить условия лабораторного эксперимента значительно труднее. Усложняются измерения и сбор информации, значительно большее влияние на объект исследования и измерительные приборы оказывают различного рода помехи (, поэтому в промышленном эксперименте особенно необходимо использовать специальные статистические методы обработки результатов. Кроме того, на реальном действующем производстве всегда желательно по возможно меньшему числу измерений получить наиболее достоверные результаты.

5.2 Планирование эксперимента

5.2.1 Основные понятия и определения

Теория планирования эксперимента началась с работ знаменитого английского ученого Р.Фишера в 30-х годах XX столетия, использовавшего ее для решения агробиологических задач. В дальнейшем это направление было развито в пятидесятых годах в США Дж.Боксом и его сотрудниками.

Под математической теорией планирования эксперимента будем понимать науку о способах составления экономичных экспериментальных планов, которые позволяют извлекать наибольшее количество информации об объекте исследования, о способах проведения эксперимента, о способах обработки экспериментальных данных и их использования для оптимизации производственных процессов, а также инженерных расчетов.

Истинный вид функции отклика $y=f(x_1, ..., x_i, ..., x_k)$ до эксперимента чаще всего неизвестен, в связи с чем для математического описания поверхности отклика используют уравнение

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{\substack{i,u=1 \\ i \neq u}}^{k} \beta_{iu} x_i x_u + \ldots, \qquad (5.4)$$

где x_i, x_u – переменные факторы при i=1, ..., k; u=1, ..., k; $i \neq u$;

$$\beta_i = \left(\frac{\partial f}{\partial x_i}\right)_0; \quad \beta_{iu} = \left(\frac{\partial^2 f}{\partial x_i \partial x_u}\right)_0; \quad \beta_{ii} = \left(\frac{\partial^2 f}{2\partial x_i^2}\right)_0$$ – коэффициенты.

Это уравнение является разложением в ряд Тейлора неизвестной функции отклика в окрестности точки с $x_i = x_{i0}$.

На практике по результатам эксперимента производится обработка данных по методу наименьших квадратов. Этот метод позволяет найти оценку b коэффициентов β, и данный полином заменяется уравнением вида

$$\tilde{y} = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{\substack{i,u=1 \\ i \neq u}}^{k} b_{iu} x_i x_u + \ldots, \qquad (5.5)$$

которое является регрессионной моделью (моделью регрессионного анализа). В этом выражении $\tilde{y}$ означает модельное, т.е. рассчитываемое по уравнению модели, значение выхода. Коэффициенты регрессии определяются экспериментально и служат для статистической оценки теоретических коэффициентов, т.е.
$b_0 \to \beta_0, b_i \to \beta_i, \; b_{iu} \to \beta_{iu}, b_{ii} \to \beta_{ii}$.

В регрессионной модели члены второй степени $x_i x_u$, x_i^2 характеризуют кривизну поверхности отклика. Чем больше кривизна этой поверхности, тем больше в модели регрессии членов высшей степени. На практике чаще всего стремятся ограничиться линейной моделью.

Последовательность активного эксперимента заключается в следующем:

1. разрабатывается схема проведения исследований, т.е. выполняется планирование эксперимента. При планировании экспериментов обычно требуется с наименьшими затратами и с необходимой точностью либо построить регрессионную модель процесса, либо определить его оптимальные условия;
2. осуществляется реализация опыта по заранее составленному исследователем плану, т.е. осуществляется сам активный эксперимент;
3. выполняется обработка результатов измерений, их анализ и принятие решений.

Таким образом, планирование эксперимента – это процедура выбора условий проведения опытов, их количества, необходимых и достаточных для решения задач с поставленной точностью.

Использование теории планирования эксперимента обеспечивает:

- минимизацию, т.е. предельное сокращение необходимого числа опытов;
- одновременное варьирование всех факторов;
- выбор четкой стратегии, что позволяет принимать обоснованные решения после каждой серии опытов;
- минимизацию ошибок эксперимента за счет использования специальных проверок.

Использование теории планирования эксперимента может явиться одним из путей существенного повышения эффективности многофакторных экспериментальных исследований.

В планировании экспериментов применяются в основном планы первого и второго порядков. Планы более высоких порядков используются в инженерной практике редко. В связи с этим далее приводится краткое изложение методики составления планов эксперимента для моделей первого и второго порядков.

Под планами первого порядка понимают такие планы, которые позволяют провести эксперимент для отыскания уравнения регрессии, содержащего только первые степени факторов и их произведения:

$$\tilde{y} = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{\substack{i,u=1 \\ i \neq u}}^{k} b_{iu} x_i x_u + \sum_{\substack{i,j,u=1 \\ i \neq j \neq u}}^{k} b_{iju} x_i x_j x_u + \ldots \quad (5.6)$$

Планы второго порядка позволяют провести эксперимент для отыскания уравнения регрессии, содержащего и вторые степени факторов:

$$\tilde{y} = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{\substack{i,u=1 \\ i \neq u}}^{k} b_{iu} x_i x_u + \ldots \qquad (5.7)$$

Нахождение уравнения регрессии методом планирования экспериментов состоит из следующих этапов:

- выбор основных факторов и их уровней;
- планирование и проведение собственно эксперимента;
- определение коэффициентов уравнения регрессии;
- статистический анализ результатов эксперимента.

5.2.2 Планы второго порядка

Описание поверхности отклика полиномами первого порядка часто оказывается недостаточным. Во многих случаях удовлетворительная аппроксимация может быть достигнута, если воспользоваться полиномом второго порядка (5.7).

В этом случае требуется, чтобы каждый фактор варьировался не менее чем на трех уровнях. В этом случае ПФЭ содержит слишком большое количество опытов, равное 3^k. Так, при k=3 их 27, а число коэффициентов b=10, при k=5 число опытов 243, а коэффициентов 21. В связи с этим осуществление ПФЭ для планов второго порядка не только сложно, но и нецелесообразно.

Сократить число опытов можно, воспользовавшись так называемым композиционным или последовательным планом, разработанным Боксом и Уилсоном. Так, при двух факторах модель функции отклика $y = f(x_1, x_2)$ второго порядка представляет собой поверхность в виде цилиндра, конуса, эллипса и т.д., описываемую в общем виде уравнением

$$\tilde{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 .$$

Для определения такой поверхности необходимо располагать координатами не менее трех ее точек, т.е. факторы x_1 и x_2 должны варьироваться не менее чем на трех уровнях. Поэтому план эксперимента в плоскости факторов x_1 и x_2 на рис. 5.1а не может состоять лишь из опытов 1, 2, 3, 4 ПФЭ 2^2, располагающихся в вершинах квадрата, как для модели первого порядка. К ним должны быть добавлены опыты (звездные точки) 5, 6, 7, 8, расположенные на осях x_1 и x_2 с координатами (±α;0), (0;±α) и обязательно опыт 9 в центре квадрата, чтобы по любому направлению (5-9-6), (1-9-4) и т.д. располагалось три точки, определяющие кривизну поверхности в этом направлении.

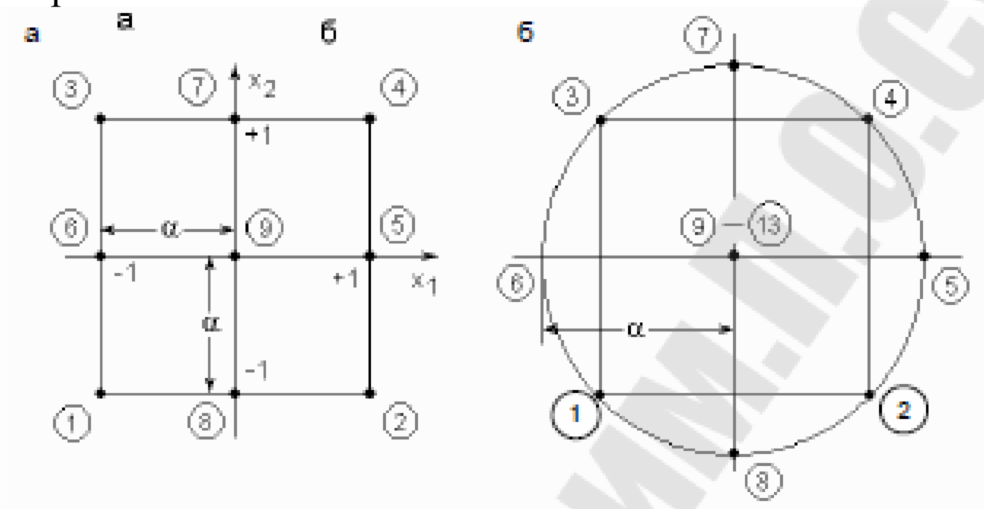

Рисунок 5.1 – Планы второго порядка при k=2:
а – ортогональный; б – ротатабельный

Таким образом, в общем случае ядро композиционного плана составляет при k<5 ПФЭ 2^k, а при k≥5 – дробную реплику от него.

Если линейное уравнение регрессии оказалось неадекватным, необходимо:

1) добавить 2 k звездных точек, расположенных на координатных осях факторного пространства (±α,0,0,...,0), (0,±α,0,...,0), ..., (0,0,...,±α), где α – звездное плечо, или расстояние до звездной точки;

2) провести n_0 опытов при значениях факторов в центре плана.

При k факторах общее число опытов в матрице композиционного плана составит

$n=2^k + 2\cdot k + n_0$ при $k<5$,

$n=2^{k-1} + 2\cdot k + n_0$ при $k\geq 5$.

При этом величина звездного плеча α и число опытов в центре плана n_0 зависит от выбранного вида композиционного плана.

5.2.3 Ротатабельные планы второго порядка

Ротатабельным называют планирование, для которого дисперсия отклика (выходного параметра) $\tilde{y}$, предсказанного уравнением регрессии, постоянна для всех точек, находящихся на равном расстоянии от центра эксперимента.

Экспериментатору заранее не известно, где находится та часть поверхности отклика, которая представляет для него особый интерес, поэтому следует стремиться к тому, чтобы количество информации, содержащееся в уравнении регрессии, было одинаково для всех равноотстоящих от центра эксперимента точек. Бокс и Хантер предложили ротатабельные планы 2-го порядка. Для того чтобы композиционный план был ротатабельным, величину звездного плеча α выбирают из условия:

$\alpha = 2^{\frac{k}{4}}$ *при* $k < 5$ и $\alpha = 2^{\frac{k-1}{4}}$ *при* $k \geq 5$

или в общем случае

$$\alpha = 2^{\frac{k-p}{4}}, \qquad (5.8)$$

где k – число факторов; p – дробность реплики (для ПФЭ p=0, для полуреплики p=1, для четверть реплики p=2 и т.д.).

В таблице 5.1 указаны возможные для расчета сочетания *K* и *KPR* и соответствующие им значения *Z*, *NC*, *N0* и *NI*. В таблице 2.1 в колонке «K» указано количество факторов (до 10 факторов); в колонке «KPR» – под «0» понимается полный факторный эксперимент, под «1» – полуреплика, «2» – 1/4-реплика, под «3» – 1/4-реплика и под «4» – 1/16-реплика; в колонке «NC» – число точек ядра; в колонке «N0» – число центральных точек; ; в колонке «NI» – общее число опытов; в колонке «Z» – величина плеча для звездных точек α.

Для конкретного эксперимента выбирается количество факторов *K* и требуемое значение *KPR* исходя из таблицы 5.1, а затем по значению KP=K-KPR определяется план эксперимента с использованием таблиц 5.2, …, 5.6. Например, требуется провести восьмифакторный эксперимент (K=8) с использованием 1/4-реплики от ПФЭ, т.е. KPR=2. Определяем значения KP=K-KPR=8-2=6. Следовательно, план будем брать из таблицы 5.6 следующим образом. Вычеркиваются (не используются) колонки 9 и10, а также строки 73,74, 83 и 84, т.е. в плане остаются 8 колонок и 80 строк, к которым добавляются 13 строк для проведения эксперимента в центральных точках (N0=13). В результате получается в плане 8 колонок и 93 строки, что соответствует строке12 таблицы 5.1 (NI=93).

Таблица 5.1 – Возможные для расчета сочетания *K* и *KPR* и соответствующие им значения *Z*, *NC*, *N0* и *NI*

№	K	KPR	NC	N0	NI	Z
1	2	0	4	5	13	1,41421
2	3	0	8	6	20	1,68179
3	4	0	16	7	31	2,0
4	4	1	8	4	20	1,68179
5	5	0	32	10	52	2,37841
6	5	1	16	6	32	2,0
7	6	0	64	15	91	2,82843
8	6	1	32	9	53	2,37841
9	6	2	16	5	33	2,0
10	7	1	64	14	92	2,82843
11	7	2	32	8	54	2,37841
12	8	2	64	13	93	2,82843
13	8	3	32	6	54	2,37841
14	9	3	64	11	93	2,82843
15	10	4	64	10	94	2,82843

Таблица 5.2 – Значения кодов уровней при KP=K-KPR=2

№ экспериментального результата	К=2	
	1	2
1	+1	+1
2	-1	+1
3	+1	-1
4	-1	-1
5	$+\alpha$	0
6	0	$+\alpha$
7	$-\alpha$	0
8	0	$-\alpha$
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0

Таблица 5.3 – Значения кодов уровней при KP=K-KPR=3

№ экспериментального результата	К=4			
	К=3			
	1	2	3	4
1	+1	+1	+1	+1
2	-1	+1	+1	-1
3	+1	-1	+1	-1
4	-1	-1	+1	+1
5	+1	+1	-1	-1
6	-1	+1	-1	+1
7	+1	-1	-1	+1
8	-1	-1	-1	-1
9	$+\alpha$	0	0	0
10	0	$+\alpha$	0	0
11	0	0	$+\alpha$	0
12	0	0	0	$+\alpha$
(*NC*+*K*+1)	$-\alpha$	0	0	0
(*NC*+*K*+2)	0	$-\alpha$	0	0
(*NC*+*K*+3)	0	0	$-\alpha$	0
(NC+*K*+4)	0	0	0	$-\alpha$
(*NI*-*N*0+1)	0	0	0	0
…	0	0	0	0
NI	0	0	0	0

Таблица 5.4 – Значения кодов уровней при KP=K-KPR=4

№ экспериментального результата	К=6					
	К=5					
	К=4					
	1	2	3	4	5	6
1	+1	+1	+1	+1	+1	+1
2	-1	+1	+1	+1	-1	+1
3	+1	-1	+1	+1	-1	-1
4	-1	-1	+1	+1	+1	-1
5	+1	+1	-1	+1	-1	-1
6	-1	+1	-1	+1	+1	-1
7	+1	-1	-1	+1	+1	+1
8	-1	-1	-1	+1	-1	+1
9	+1	+1	+1	-1	-1	-1

10	-1	+1	+1	-1	+1	-1
11	+1	-1	+1	-1	+1	+1
12	-1	-1	+1	-1	-1	+1
13	+1	+1	-1	-1	+1	+1
14	-1	+1	-1	-1	-1	+1
15	+1	-1	-1	-1	-1	-1
16	-1	-1	-1	-1	+1	-1
17	$+\alpha$	0	0	0	0	0
18	0	$+\alpha$	0	0	0	0
19	0	0	$+\alpha$	0	0	0
20	0	0	0	$+\alpha$	0	0
21	0	0	0	0	$+\alpha$	0
22	0	0	0	0	0	$+\alpha$
(NC+*K*+1)	$-\alpha$	0	0	0	0	0
(NC+*K*+2)	0	$-\alpha$	0	0	0	0
(NC+*K*+3)	0	0	$-\alpha$	0	0	0
(NC+*K*+4)	0	0	0	$-\alpha$	0	0
(NC+*K*+5)	0	0	0	0	$-\alpha$	0
(NC+*K*+6)	0	0	0	0	0	$-\alpha$
(*NI*-*N*0+1)	0	0	0	0	0	0
…	0	0	0	0	0	0
NI	0	0	0	0	0	0

Таблица 5.5 – Значения кодов уровней при KP=K-KPR=5

№ экспериментального результата	К=8							
	К=7							
	К=6							
	К=5							
	1	2	3	4	5	6	7	8
1	+1	+1	+1	+1	+1	+1	+1	+1
2	-1	+1	+1	+1	+1	-1	+1	-1
3	+1	-1	+1	+1	+1	-1	-1	+1
4	-1	-1	+1	+1	+1	+1	-1	-1
5	+1	+1	-1	+1	+1	-1	-1	-1
6	-1	+1	-1	+1	+1	+1	-1	+1
7	+1	-1	-1	+1	+1	+1	+1	-1
8	-1	-1	-1	+1	+1	-1	+1	+1

9	+1	+1	+1	-1	+1	-1	-1	-1
10	-1	+1	+1	-1	+1	+1	-1	+1
11	+1	-1	+1	-1	+1	+1	+1	-1
12	-1	-1	+1	-1	+1	-1	+1	+1
13	+1	+1	-1	-1	+1	+1	+1	+1
14	-1	+1	-1	-1	+1	-1	+1	-1
15	+1	-1	-1	-1	+1	-1	-1	+1
16	-1	-1	-1	-1	+1	+1	-1	-1
17	+1	+1	+1	+1	-1	-1	-1	-1
18	-1	+1	+1	+1	-1	+1	-1	+1
19	+1	-1	+1	+1	-1	+1	+1	-1
20	-1	-1	+1	+1	-1	-1	+1	+1
21	+1	+1	-1	+1	-1	+1	+1	+1
22	-1	+1	-1	+1	-1	-1	+1	-1
23	+1	-1	-1	+1	-1	-1	-1	+1
24	-1	-1	-1	+1	-1	+1	-1	-1
25	+1	+1	+1	-1	-1	+1	+1	+1
26	-1	+1	+1	-1	-1	-1	+1	-1
27	+1	-1	+1	-1	-1	-1	-1	+1
28	-1	-1	+1	-1	-1	+1	-1	-1
29	+1	+1	-1	-1	-1	-1	-1	-1
30	-1	+1	-1	-1	-1	+1	-1	+1
31	+1	-1	-1	-1	-1	+1	+1	-1
32	-1	-1	-1	-1	-1	-1	+1	+1
33	$+\alpha$	0	0	0	0	0	0	0
34	0	$+\alpha$	0	0	0	0	0	0
35	0	0	$+\alpha$	0	0	0	0	0
36	0	0	0	$+\alpha$	0	0	0	0
37	0	0	0	0	$+\alpha$	0	0	0
38	0	0	0	0	0	$+\alpha$	0	0
39	0	0	0	0	0	0	$+\alpha$	0
40	0	0	0	0	0	0	0	$+\alpha$
(NC+K+1)	$-\alpha$	0	0	0	0	0	0	0
(NC+K+2)	0	$-\alpha$	0	0	0	0	0	0
(NC+K+3)	0	0	$-\alpha$	0	0	0	0	0
(NC+K+4)	0	0	0	$-\alpha$	0	0	0	0
(NC+K+5)	0	0	0	0	$-\alpha$	0	0	0

(NC+*K*+6)	0	0	0	0	0	-α	0	0
(NC+*K*+7)	0	0	0	0	0	0	-α	0
(NC+*K*+8)	0	0	0	0	0	0	0	-α
(*NI*-*N*0+1)	0	0	0	0	0	0	0	0
…	0	0	0	0	0	0	0	0
NI	0	0	0	0	0	0	0	0

Таблица 5.6 – Значения кодов уровней при KP=K-KPR=6

№ Экспериментального результата	К=10									
	К=9									
	К=8									
	К=7									
	К=6									
	1	2	3	4	5	6	7	8	9	10
1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	-1	+1	+1	+1	+1	+1	-1	+1	-1	-1
3	+1	-1	+1	+1	+1	+1	-1	-1	+1	-1
4	-1	-1	+1	+1	+1	+1	+1	-1	-1	+1
5	+1	+1	-1	+1	+1	+1	-1	-1	-1	+1
6	-1	+1	-1	+1	+1	+1	+1	-1	+1	-1
7	+1	-1	-1	+1	+1	+1	+1	+1	-1	-1
8	-1	-1	-1	+1	+1	+1	+1	+1	+1	+1
9	+1	+1	+1	-1	+1	+1	-1	-1	-1	-1
10	-1	+1	+1	-1	+1	+1	+1	-1	+1	+1
11	+1	-1	+1	-1	+1	+1	+1	+1	-1	+1
12	-1	-1	+1	-1	+1	+1	-1	+1	+1	-1
13	+1	+1	-1	-1	+1	+1	+1	+1	+1	-1
14	-1	+1	-1	-1	+1	+1	-1	+1	-1	+1
15	+1	-1	-1	-1	+1	+1	-1	-1	+1	+1
16	-1	-1	-1	-1	+1	+1	+1	-1	-1	-1
17	+1	+1	+1	+1	-1	+1	-1	-1	-1	-1
18	-1	+1	+1	+1	-1	+1	+1	-1	+1	+1
19	+1	-1	+1	+1	-1	+1	+1	+1	-1	+1
20	-1	-1	+1	+1	-1	+1	-1	+1	+1	-1
21	+1	+1	-1	+1	-1	+1	+1	+1	+1	-1
22	-1	+1	-1	+1	-1	+1	-1	+1	-1	+1
23	+1	-1	-1	+1	-1	+1	-1	-1	+1	+1
24	-1	-1	-1	+1	-1	+1	+1	-1	-1	-1

25	+1	+1	+1	-1	-1	+1	+1	+1	+1	+1
26	-1	+1	+1	-1	-1	+1	-1	+1	-1	-1
27	+1	-1	+1	-1	-1	+1	-1	-1	+1	-1
28	-1	-1	+1	-1	-1	+1	+1	-1	-1	+1
29	+1	+1	-1	-1	-1	+1	-1	-1	-1	+1
30	-1	+1	-1	-1	-1	+1	+1	-1	+1	-1
31	+1	-1	-1	-1	-1	+1	+1	+1	-1	-1
32	-1	-1	-1	-1	-1	+1	-1	+1	+1	+1
33	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1
34	-1	+1	+1	+1	+1	-1	+1	-1	+1	+1
35	+1	-1	+1	+1	+1	-1	+1	+1	-1	+1
36	-1	-1	+1	+1	+1	-1	-1	+1	+1	-1
37	+1	+1	-1	+1	+1	-1	+1	+1	+1	-1
38	-1	+1	-1	+1	+1	-1	-1	+1	-1	+1
39	+1	-1	-1	+1	+1	-1	-1	-1	+1	+1
40	-1	-1	-1	+1	+1	-1	+1	-1	-1	-1
41	+1	+1	+1	-1	+1	-1	+1	+1	+1	+1
42	-1	+1	+1	-1	+1	-1	-1	+1	-1	-1
43	+1	-1	+1	-1	+1	-1	-1	-1	+1	-1
44	-1	-1	+1	-1	+1	-1	+1	-1	+1	-1
45	+1	+1	-1	-1	+1	-1	-1	-1	-1	+1
46	-1	+1	-1	-1	+1	-1	+1	-1	+1	-1
47	+1	-1	-1	-1	+1	-1	+1	+1	-1	-1
48	-1	-1	-1	-1	+1	-1	-1	+1	+1	+1
48	+1	+1	+1	+1	-1	-1	+1	+1	+1	+1
50	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1
51	+1	-1	+1	+1	-1	-1	-1	-1	+1	-1
52	-1	-1	+1	+1	-1	-1	+1	-1	-1	+1
53	+1	+1	-1	+1	-1	-1	-1	-1	-1	+1
54	-1	+1	-1	+1	-1	-1	+1	-1	+1	-1
55	+1	-1	-1	+1	-1	-1	+1	+1	-1	-1
56	-1	-1	-1	+1	-1	-1	-1	+1	+1	+1
57	+1	+1	+1	-1	-1	-1	-1	-1	-1	-1
58	-1	+1	+1	-1	-1	-1	+1	-1	+1	+1
59	+1	-1	+1	-1	-1	-1	+1	+1	-1	+1
60	-1	-1	+1	-1	-1	-1	-1	+1	+1	-1
61	+1	+1	-1	-1	-1	-1	+1	+1	+1	-1
62	-1	+1	-1	-1	-1	-1	-1	+1	-1	+1

63	+1	-1	-1	-1	-1	-1	-1	-1	+1	+1
64	-1	-1	-1	-1	-1	-1	+1	-1	-1	-1
65	+α	0	0	0	0	0	0	0	0	0
66	0	+α	0	0	0	0	0	0	0	0
67	0	0	+α	0	0	0	0	0	0	0
68	0	0	0	+α	0	0	0	0	0	0
69	0	0	0	0	+α	0	0	0	0	0
70	0	0	0	0	0	+α	0	0	0	0
71	0	0	0	0	0	0	+α	0	0	0
72	0	0	0	0	0	0	0	+α	0	0
73	0	0	0	0	0	0	0	0	+α	0
74	0	0	0	0	0	0	0	0	0	+α
(NC+*K*+1)	-α	0	0	0	0	0	0	0	0	0
(NC+*K*+2)	0	-α	0	0	0	0	0	0	0	0
(NC+*K*+3)	0	0	-α	0	0	0	0	0	0	0
(NC+*K*+4)	0	0	0	-α	0	0	0	0	0	0
(NC+*K*+5)	0	0	0	0	-α	0	0	0	0	0
(NC+*K*+6)	0	0	0	0	0	-α	0	0	0	0
(NC+*K*+7)	0	0	0	0	0	0	-α	0	0	0
(NC+*K*+8)	0	0	0	0	0	0	0	-α	0	0
(NC+*K*+9)	0	0	0	0	0	0	0	0	-α	0
(NC+*K*+10)	0	0	0	0	0	0	0	0	0	-α
(*NI*-*N*0+1)	0	0	0	0	0	0	0	0	0	0
…	0	0	0	0	0	0	0	0	0	0
NI	0	0	0	0	0	0	0	0	0	0

Для проведения эксперимента переход от кодированных значений переменных к натуральным осуществляется с использованием следующих зависимостей:

$$\frac{x_i - x_{0i}}{\varepsilon_i} = \pm 1 \quad u \quad \frac{x_i - x_{0i}}{\varepsilon_i} = \pm \alpha_i, \qquad (5.9)$$

где ε_i – значения интервалов варьирования факторов.

Учитывая специфический характер ротатабельного плана в общем виде, можно также получить формулы для расчета коэффициентов уравнения регрессии и их дисперсий:

$$b_0 = \frac{A}{n}\left[2\lambda^2(k+2)(0y) - 2\lambda c\sum_{i=1}^{k}(iiy)\right] \qquad (5.10)$$

$$b_i = (c/n)(iy); \qquad (5.11)$$

$$b_{ii} = \frac{A}{n}\left[c^2\left[(k+2)\lambda - k\right](iiy) + c^2(1-\lambda)\sum_{i=1}^{k}(iiy) - 2\lambda c(0y)\right]; \qquad (5.12)$$

$$b_{iu} = \frac{c^2}{n\lambda}(iuy); \qquad (5.13)$$

$$S_{b_0}^2 = \frac{2A\lambda^2(k+2)}{n}S_{восп}^2; \qquad (5.14)$$

$$S_{b_{ii}}^2 = \frac{A\left[(k+1)\lambda - (k-1)c^2\right]}{n}S_{восп}^2; \qquad (5.15)$$

$$S_{b_{iu}}^2 = \frac{c^2}{\lambda n}S_{восп}^2, \qquad (5.16)$$

где $(0y) = \sum_{j=1}^{n} x_{oj} y_j$; (5.17)

$$(iiy) = \sum_{j=1}^{n} x_{ij}^2 y_j; \qquad (5.18)$$

$$(iy) = \sum_{j=1}^{n} x_{ij} y_j; \qquad (5.19)$$

$$(iuy) = \sum_{j=1}^{n} x_{ij} x_{uj} y_j; \qquad (5.20)$$

$$C = \frac{n}{\sum_{j=1}^{n} x_{ij}^2 y_j}; \qquad (5.21)$$

$$A = \frac{1}{2\lambda\left[(k+2)\lambda - k\right]}; \qquad (5.22)$$

$$\lambda = \frac{nk}{(k+2)n_1} = \frac{k(n_1+n_0)}{(k+2)n_1}. \qquad (5.23)$$

Значимость коэффициентов проверяем по критерию Стьюдента

$$t_i = \frac{|b_i|}{S_{b_i}} \qquad (5.24)$$

Если $t_i > t_{критическое(0{,}05;n_0-1)}$, то коэффициент считается значимым.

Здесь n_0 – число опытов в центре плана; n_1 – число остальных опытов.

Матрица ротатабельного планирования, оказывается неортогональной, так как

$$\sum_{j=1}^{n} x_{0j} x_{uj}{}^2 \neq 0; \quad \sum_{j=1}^{n} x_{ij}{}^2 x_{uj}{}^2 \neq 0; \quad i \neq u\,.$$

Следовательно, если какой-либо из квадратичных эффектов оказался незначимым, то после его исключения коэффициенты уравнения регрессии необходимо пересчитать заново.

При использовании ротатабельных планов второго порядка дисперсию воспроизводимости можно определить по опытам в центре плана. В связи с этим при проверке адекватности уравнения регрессии, полученного по ротатабельному плану второго порядка, поступают следующим образом.

Находят остаточную сумму квадратов

$$S_1^2 = \sum_{j=1}^{n} (y_j - \tilde{y}_j)^2 \tag{5.25}$$

с числом степеней свободы

$$m_1 = n - l = n - \frac{(k+2)(k+1)}{2}\,.$$

По опытам в центре плана определяют сумму квадратов воспроизводимости

$$S_2^2 = \sum_{j=1}^{n_0} (y_{0j} - \bar{y}_{0j})^2 \tag{5.26}$$

с числом степеней свободы $m_2 = n_o - 1$.

Далее находят сумму квадратов, характеризующих неадекватность

$S_3{}^2 = S_1{}^2 - S_2{}^2$, число степеней свободы которой

$$m_3 = m_1 - m_2 = n - \frac{(k+2)(k+1)}{2} - (n_0 - 1)\,.$$

Проверяют адекватность по F-критерию Фишера:

$$F = \frac{S_3^2 / m_3}{S_2^2 / m_2}\,. \tag{5.27}$$

Уравнение адекватно, если $F < F_{\alpha;m3;m2}$.

Если модель второго порядка оказалась неадекватной, следует повторить эксперименты на меньшем интервале варьирования факторов или перенести центр плана в другую точку факторного пространства. В тех случаях, когда адекватность модели по-прежнему не достигается, рекомендуется перейти к планам третьего порядка.

5.3 Планирования эксперимента и латинские квадраты

Дисперсионный анализ тесно связан с планированием эксперимента. Удачно спланированный эксперимент, выявляя все необходимые факторы (эффекты), оказывается всегда либо более точным, либо менее трудоемким по сравнению с непродуманным экспериментом. В случае, если на результат эксперимента действуют одновременно несколько факторов, то наилучший эффект дает одновременный дисперсионный анализ всех этих факторов, а не по парное сравнения этих факторов.

Эксперимент, спланированный так, что в нем встречаются все возможные сочетания уровней изучаемых факторов называется полным факторным экспериментом (ПФЭ). Число испытаний при таком эксперименте равно произведению количества уровней изучаемых факторов. Если изучению подлежит один или два фактора, то используется обычно ПФЭ [12].

Если же изучению подлежит три и более факторов, то целесообразно использовать дробный факторный эксперимент (ДФЭ), при котором некоторые сочетания уровней пропущены. ДФЭ дает возможность при неизменном числе испытаний исследовать гораздо большее число факторов, чем ПФЭ. При этом часть информации теряется, но при удачном спланированном ДФЭ теряется именно та часть информации, которая по тем или иным причинам не в данный момент несущественна. Особенно широко используется ДФЭ, в которых теряется лишь информация о взаимодействиях изучаемых факторов и поэтому наблюдения при каждом уровне изучаемых факторов не дублируются. При изучении двух и более факторов целесообразно, чтобы количество уровней всех факторов было одинаковым (K).

Таким образом, если проводится исследование двух факторов A и B с числом уровней обоих факторов по K, то получается ПФЭ с числом испытаний равным K^2. Если попытаться, не меняя числа испытаний, присоединить к этим факторам третий фактор C, имеющий *К* уровней, то в результате получится ДФЭ. Задача спланировать его так, чтобы можно было оценить эффекты всех трех факторов. Для этого необходимо так расположить уровни фактора *С*, чтобы в каждом столбце и в каждой строке встречались все уровни без повторений.

Такие расположения существуют и при том не единственные. Согласно Фишеру их называют латинскими квадратами (Фишер обозначал уровни добавляемого фактора латинскими буквами – отсюда и получилось название квадрата). Эти расположения приводятся в специальных справочниках.

К исследуем трем факторам *A*, *B*, *C*, не меняя общего числа наблюдений, можно попытаться добавить еще и четвертый фактор *D*. Это нетрудно сделать, если удастся найти такое расположение уровней факторов *C* и *D*, при котором в каждой строке и в каждом столбце имеются все *К* уровней фактора *C* и все *К* уровней фактора *D* и в то же время никаких два уровня факторов *C* и *D* не встречаются во всей таблице вместе более одного раза. Расположение такого типа называется квадратом второго порядка. Строить квадраты второго порядка довольно сложно. В табл. 5.7 представлен пример латинского квадрата второго порядка при *К=4*.

Таблица 5.7– Квадрат второго порядка при К=4

	A_1	A_2	A_3	A_4
B_1	C_1 D_1	C_2 D_2	C_3 D_3	C_4 D_4
B_2	C_2 D_3	C_1 D_4	C_4 D_1	C_3 D_2
B_3	C_3 D_4	C_4 D_3	C_1 D_2	C_2 D_1
B_4	C_4 D_2	C_3 D_1	C_2 D_4	C_1 D_3

5.4 **Определение количества параллельных наблюдений**

При увеличении числа параллельных наблюдений *n* уменьшается доверительный интервал. Однако системная ошибка β_0 при этом не уменьшается, так что суммарная ошибка все равно будет больше β_0.

Поэтому целесообразно выбирать *n* так, чтобы ширина доверительного интервала составляла $50-100\%\ \beta_0$. Этому условию соответствует $n=3-4$. Считается нецелесообразным брать $n=5-10$, так как возрастающая трудоемкость эксперимента не оправдывается достигаемой точностью.

Если не учитывается взаимное влияние основных факторов, то проводить параллельные наблюдения не требуется, что обычно бывает при ДФЭ.

5.5 Определение значимости основных факторов

Рассмотрим случай, когда один или несколько основных факторов изменяется заданным образом. Эти изменения могут повлиять на результаты наблюдений. Степень влияния, его качественные и количественные характеристики являются объектами дисперсионного анализа.

Предполагаем, что случайные ошибки наблюдений имеют нормальное распределение. Влияние изучаемых факторов может быть двояким: они могут изменять как истинный результат (среднее) наблюдений α, так и дисперсию этих наблюдений σ^2. Однако, будем предполагать, что дисперсия наблюдений σ^2 остается неизменной (используется одна методика проведения эксперимента и те же приборы). Если же стабильность дисперсии вызывает опасение, то нужно провести специальное исследование с помощью критериев Бартлера и Кохрана.

Для того, чтобы сравнить влияние различных факторов, нужно найти какой-нибудь достаточно надежный и универсальный показатель этого влияния. Допустим, что изучаемый фактор A на различных уровнях привел к серии истинных результатов: $\alpha_1, \alpha_2, ..., \alpha_k$. В качестве показателей влияния фактора A будем брать число $\sigma_A^2 = \frac{1}{k}\sum_{i=1}^{k}(\alpha_i - \overline{\alpha})^2$, где $\overline{\alpha} = \frac{\alpha_1 + \alpha_2 + ... + \alpha_k}{k}$ – среднее арифметическое наблюдений α_i.

Число σ_A^2 называется дисперсией фактора A. Это наименование дано по аналогии с обычной дисперсией, но числа α_i в эксперименте не являются случайными, поэтому σ_A^2 не связана ни с какой случайной величиной. Выбор σ_A^2 удобен по двум причинам. Во-первых, дисперсия является простейшей меркой рассеивания. Во-вторых, показатель влияния фактора A определен теперь аналогично показателю случайного фактора (т.е. обычной дисперсии σ^2). Это позволяет непосредственно сравнивать фактор A с эффектом случайности.

Рассмотрим самый простой случай, когда дисперсия наблюдений σ^2 известна заранее и исследуется один переменный фактор A. Пусть при изменении фактора A получились результаты наблюдений $x_1, x_2, ..., x_k$. Найдем выборочную дисперсию

$$S^2 = \frac{1}{k-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 = \frac{1}{k-1}\left(\sum_{i=1}^{k} x_i^2 - \frac{\left(\sum_{i=1}^{k} x_i\right)^2}{k}\right).$$

Сравним эту дисперсию, имеющую $f = k-1$ степеней свободы, с генеральной дисперсией σ^2 с использованием критерия Фишера. Если S^2 от σ^2 отличается незначительно, то и влияние фактора A нужно признать незначимым, так как он не сумел существенно увеличить случайный разброс наблюдений. Если же S^2 отличается значимо от σ^2, то это может быть вызвано только влиянием фактора A, которое теперь нужно признать значимым. Для тог, чтобы оценить σ_A^2, воспользуемся тем. Что дисперсия суммы двух независимых случайных величин равна их дисперсий. В нашем случае складываются эффект случайности (с дисперсией σ^2) и эффект воздействия фактора A (с дисперсией σ_A^2), которые очевидным образом независимы. Поэтому общая дисперсия наблюдений должна быть равна $\sigma^2 + \sigma_A^2$. А величина S^2 является оценкой общей дисперсии. Следовательно, $\sigma_A^2 \approx S^2 - \sigma^2$.

Рассмотрим сравнение дисперсий S_1^2 и S_2^2, имеющих соответственно f_1 и f_2 степеней свободы. Будем считать, что первая выборка сделана из генеральной совокупности с дисперсией σ_1^2, а вторая – из генеральной совокупности с дисперсией σ_2^2. Выдвигается нулевая гипотеза, то есть гипотеза о равенстве $\sigma_1^2 = \sigma_2^2$. Для того чтобы отвергнуть гипотезу, нужно доказать значимость расхождения между S_1^2 и S_2^2 при выбранном уровне значимости p'. Причем, $p' = 1 - p$, где p – вероятность. В качестве критерия значимости обычно используется так называемое *распределение Фишера* (или F –

распределение), которым называется распределение случайной величины: $F = \frac{S_1^2}{\sigma_1^2} : \frac{S_2^2}{\sigma_2^2}$. Это распределение зависит только от f_1 и f_2, при этом $F(f_1, f_2) = \frac{1}{F(f_2, f_1)}$. Кривые этого распределения имеют ассиметричную форму.

Квантили F_{1-p} для некоторых наиболее употребительных уровней значимости и различных комбинаций f_1 и f_2 приводятся в соответствующих статистических таблицах. При нахождении квантилей F_p для значений вероятностей p, не вошедших в таблицу, используется соотношение: $F_p(f_1, f_2) = \frac{1}{F_{1-p}(f_2, f_1)}$. Например, $F_{0,95}(4{,}3) = 9{,}1; \quad F_{0,05}(3{,}4) = \frac{1}{9{,}1} = 0{,}11$.

С помощью F – распределение находятся доверительные оценки для отношений $\frac{\sigma_1^2}{\sigma_2^2}$, либо $\frac{S_1^2}{S_2^2}$, либо $\frac{S_1^2}{\sigma_2^2}$. С вероятностью $1-p$ должно выполняться двустороннее неравенство

$$\frac{1}{F_{1-p/2}(f_2, f_1)} \le \frac{S_1^2}{S_2^2} \le F_{1-p/2}(f_1, f_2),$$

или одно из односторонних неравенств:

$$\frac{S_1^2}{S_2^2} \le F_{1-p}(f_1, f_2); \quad \frac{S_1^2}{S_2^2} \ge \frac{1}{F_{1-p}(f_2, f_1)}.$$

Причем если отношение $\frac{S_1^2}{S_2^2}$ попадает в критическую область, то различие между дисперсиями следует признать значимым.

Сформулируем получившиеся критерии значимости. Для удобства изложения будем обозначать через S_1^2, большую из сравниваемых дисперсий. Если большей выборочной дисперсии заведомо не может соответствовать меньшая генеральная, то применяется односторонний критерий. Нулевая гипотеза отвергается, если $\frac{S_1^2}{S_2^2} > F_{1-p}(f_1, f_2)$.

При неизвестном соотношении между σ_1^2 и σ_2^2 применяется двусторонний критерий и нулевая гипотеза отвергается, если $\frac{S_1^2}{S_2^2} > F_{1-p/2}(f_1, f_2)$, причем $S_1^2 > S_2^2$.

Если рассматривается отношение $\frac{S_1^2}{\sigma_2^2}$, то для генеральной дисперсии принимается $f_2 = \infty$.

5.6 Анализ при четырех основных факторах

Дисперсионный анализ особенно эффективен при одновременном изучении нескольких факторов, при котором каждое наблюдение служит для одновременной оценки всех факторов и их взаимодействий. При этом можно не делать параллельных наблюдений, ограничиваясь лишь одним наблюдением для каждого сочетания уровней изучаемых факторов.

Обозначим через $x_{i,j}$ наблюдение получаемое при уровнях A_j и B_i (табл. 5.8.).

Таблица 5.8. -Исходные данные

	A_1	A_2	A_3	A_4
B_1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$
B_2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$
B_3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$
B_4	$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	$x_{4,4}$

Если изучаются одновременно четыре фактора A, B, C и D на уровнях $A_1, A_2, ..., A_к$, $B_1, B_2, ..., B_к$, $C_1, C_2, ..., C_к$, $D_1, D_2, ..., D_к$, то есть проводится ДФЭ, то рекомендуется следующая схема вычислений [12].

1. Определяются значения сумм наблюдений

- по столбцам $XJ_j = \sum_{i=1}^{k} x_{i,j}, \quad j = 1, 2, ..., 4$;

- по строкам $XI_i = \sum_{j=1}^{k} x_{i,j}, \quad i = 1, 2, ..., 4$.

2. Находится сумма квадратов всех наблюдений $Q1 = \sum_{i=1}^{k} \sum_{j=1}^{k} x_{i,j}^2$.
3. Находится сумма квадратов итогов по столбцам (XJ_j), деленная на k: $Q2 = \frac{1}{k} \sum_{j=1}^{k} XJ_j^2$.
4. Находится сумма квадратов итогов по строкам, деленная на k: $Q3 = \frac{1}{k} \sum_{i=1}^{k} XI_i^2$.
5. Находится сумма квадрата общего итога, деленного на k^2: $Q4 = \frac{1}{k^2} (\sum_{i=1}^{k} XI_i)^2 = \frac{1}{k^2} (\sum_{j=1}^{k} XJ_j)^2$.
6. Находится сумма квадратов итогов по уровням фактора C, деленная на k:

$$Q5 = \frac{1}{k} \cdot ((x_{1,1} + x_{2,2} + x_{3,3} + x_{4,4})^2 + (x_{1,4} + x_{2,3} + x_{3,2} + x_{4,1})^2 + \\ + (x_{1,2} + x_{2,1} + x_{3,4} + x_{4,3})^2 + (x_{1,3} + x_{2,4} + x_{3,1} + x_{4,2})^2).$$

7. Находится сумма квадратов итогов по уровням фактора D, деленная на k:

$$Q6 = \frac{1}{k} ((x_{1,1} + x_{2,3} + x_{3,4} + x_{4,2})^2 + (x_{1,2} + x_{2,4} + x_{3,3} + x_{4,1})^2 + \\ + (x_{1,3} + x_{2,1} + x_{3,2} + x_{4,4})^2 + (x_{1,4} + x_{2,2} + x_{3,1} + x_{4,3})^2).$$

8. Определяется совместная дисперсия воспроизводимости и взаимодействия $S_0^2 = \frac{Q1 + 3 \cdot Q4 - Q2 - Q3 - Q5 - Q6}{(k-1)(k-4)}$.
9. Определяются выборочные дисперсии факторов A, B, C, D: $S_A^2 = \frac{Q2 - Q4}{k-1}$, $S_B^2 = \frac{Q3 - Q4}{k-1}$, $S_C^2 = \frac{Q5 - Q4}{k-1}$, $S_D^2 = \frac{Q6 - Q4}{k-1}$.
10. Определяются отношения $\frac{S_A^2}{S_0^2}$, $\frac{S_B^2}{S_0^2}$, $\frac{S_C^2}{S_0^2}$, $\frac{S_D^2}{S_0^2}$.

11. Определяется значимость основных факторов по критерию Фишера. Для этого необходимо определить значение квантиля F_{i-p} для уровня значимости 0,05 при $f_1 = k-1 = 3$ *и* $f_2 = (k-1)\cdot(k-3) = 1$ (в данном случае, $F_{i-p}(f_1, f_2) = 9{,}3$). При этом, если $\frac{S_i^2}{S_0^2} > F_{i-p}(f_1, f_2)$, то нулевая гипотеза отвергается и фактор, имеющий дисперсию S_i^2 признается значимым, в противном случае он признается незначимым.

12. Ищется эмпирическая зависимость вида $y = n \cdot A^a \cdot B^b \cdot C^c \cdot D^d$, для которой необходимо определить параметры *n, a, b, c, d.*

5.7 Определение параметров эмпирической зависимости

При установлении связи между значимыми факторами используют в качестве эмпирических зависимостей степенные функции, которые путем логарифмирования легко приводятся к линейным зависимостям. Для определения параметров эмпирических зависимостей используется метод наименьших квадратов (МНК), который включает получение нормальной системы линейных уравнений.

5.7.1 Определения параметров эмпирических зависимостей с использованием МНК

Данный метод основан на том, что при определении параметров эмпирической зависимости (математической модели), их значения рассчитываются таким образом, чтобы расчетные по этой модели значения зависимой переменной $\mathcal{Y}_j$ имели минимум суммы квадратов отклонений от оценок зависимой переменной Y_j, полученных в опытах. Сумма квадратов отклонений, рассчитанных значений от оценок в опытах, рассчитывается по формуле [19]:

$$Sr = \sum_{j=1}^{m} (\mathcal{Y}_j - Y_j)^2, \qquad (5.28)$$

где Y_j – оценка зависимой переменной величины в *j*-ом опыте; *m* – количество опытов; $У_j$ – значение зависимой переменной величины, рассчитанное по математической модели для условий *j*-ого опыта.

Если в формуле (5.28) каждое значение зависимой переменной величины $У_j$, получаемое расчетом по некоторой математической модели, представить как некоторую функцию, зависящую не только от факторов *X*1, *X*2, …, *X*К, но и от параметров, *b*0, *b*1, …, *b*k, которые требуется определить, то есть

$$У_j = f(X1_j, X2_j, ..., XK_j, b0, b1, \ ..., \ bk) \qquad (5.29)$$

где Xi_j – значение *i*-ой независимой переменной в *j*-ом опыте; *b*i – значение *i*-ого параметра в математической модели, то сумму квадратов отклонений (5.28) можно рассматривать как непрерывную сложную функцию от нескольких переменных, которыми будут являться параметры функции (13.2):

$$Sr = \sum_{j=1}^{m}(f(X1_j, \quad X2_j, \quad ..., Xk_j, ..., b0, b1, ..., bk) - Y_j,)^2 \qquad (5.30)$$

Необходимые условия минимума функции *Sr* записываются в виде системы *k*+1 уравнений:

$$\frac{\partial Sr}{\partial b0} = 0, \quad \frac{\partial Sr}{\partial b1} = 0, ..., \quad \frac{\partial Sr}{\partial bk} = 0\,,$$

где $\frac{\partial Sr}{\partial bi}$ - частная производная функции (5.30) по *i*-му параметру.

5.7.2 Определения параметров эмпирической зависимости для четырех основных факторов

Эмпирическую зависимость, например, зависимость качества обработанной поверхности от факторов *A*, *B*, *C*, *D* будем искать в виде [12]

$$Y = n \cdot A^a \cdot B^b \cdot C^c \cdot D^d\,. \qquad (5.31)$$

Чтобы найти неизвестные параметры с использованием МНК, предварительно ее нужно привести к линейному виду логарифмированием. Для полученного уравнения функция суммы квадратов отклонений будет иметь вид:

$$Sr = \sum_{j=1}^{m}(\lg n + a \cdot \lg A_j + b \cdot \lg B_j + c \cdot \lg C_j + d \cdot \lg D_j - \lg Y_j)^2 \to \min$$

Нахождение частных производных от функции суммы квадратов отклонений по неизвестным параметрам приведет к получению системы линейных уравнений.

Обозначим через суммы следующие выражения, получим следующую систему линейных уравнений:

$$S1=\sum_{j=1}^{m}\lg A_j,\ S2=\sum_{j=1}^{m}\lg B_j,\ S3=\sum_{j=1}^{m}\lg C_j,\ S4=\sum_{j=1}^{m}\lg D_j,\ S5=\sum_{j=1}^{m}\lg Y_j,$$

$$S6=\sum_{j=1}^{m}(\lg A_j)^2,\ S7=\sum_{j=1}^{m}\lg B_j\cdot\lg A_j,\ S8=\sum_{j=1}^{m}\lg C_j\cdot\lg A_j,$$

$$S9=\sum_{j=1}^{m}\lg D_j\cdot\lg A_j,\ S10=\sum_{j=1}^{m}\lg Y_j\cdot\lg A_j,\ S11=\sum_{j=1}^{m}(\lg B_j)^2,$$

$$S12=\sum_{j=1}^{m}\lg C_j\cdot\lg B_j,\ S13=\sum_{j=1}^{m}\lg D_j\cdot\lg B_j,\ S14=\sum_{j=1}^{m}\lg Y_j\cdot\lg B_j,$$

$$S15=\sum_{j=1}^{m}(\lg C_j)^2,\ S16=\sum_{j=1}^{m}\lg D_j\cdot\lg C_j,\ S17=\sum_{j=1}^{m}\lg Y_j\cdot\lg C_j,$$

$$S18=\sum_{j=1}^{m}(\lg D_j)^2,\ S19=\sum_{j=1}^{m}\lg Y_j\cdot\lg D_j.$$

Получаем систему линейных уравнений с неизвестными параметрами lg*n*, *a*, *b*, *c*, *d*:

$$\begin{cases} m\cdot\lg n+a\cdot S1+b\cdot S2+c\cdot S3+d\cdot S4=S5\\ \lg n\cdot S1+a\cdot S6+b\cdot S7+c\cdot S8+d\cdot S9=S10\\ \lg n\cdot S2+a\cdot S7+b\cdot S11+c\cdot S12+d\cdot S13=S14,\\ \lg n\cdot S3+a\cdot S8+b\cdot S12+c\cdot S15+d\cdot S16=S1\\ \lg n\cdot S4+a\cdot S9+b\cdot S13+c\cdot S16+d\cdot S18=S19\end{cases}$$

где $m=k^2=16$.

Вопросы для самопроверки

1. Что такое полный факторный эксперимент?
2. Что такое дробный факторный эксперимент?
3. Как построить латинский квадрат?
4. В какой последовательности определяются значимые факторы?
5. Сущность метода наименьших квадратов (МНК).

Литература

1. Вентцель Е.С. Исследование операций: задачи, принципы, методология. – М.: Наука, 1980. – 208 с.
2. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. – М.: Радио и связь, 1983. – 416 с.
3. Волков И. К., Загоруйко Е. А. Исследование операций: Учеб. для вузов / Под ред. В.С. Зарубина, А.П. Крищенко. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2000. – 436 с.
4. Климович Ф.Ф., Присевок А.Ф. Математическое моделирование технологических задач в машиностроении: Учебно-методическое пособие по лабораторным работам для студентов машиностроительных специальностей высших учебных заведений. – Мн.: БГПА, 2000. – 88с.
5. Конвей Р.В., Максвелл В.Л., Миллер Л.В. Теория расписаний. – М.: Наука, 1975. –360 с.
6. Кудрявцев Е.М. Исследование операций в задачах, алгоритмах программах. – М.: Радио и связь, 1984. – 184 с.
7. Кузнецов А.В., Холод Н.И. Математическое программирование: учеб. – 2-ое изд., перераб. и доп. – Минск: Выш. шк., 2001. – 351 с.
8. Кузнецов А.В., Сакович В.А., Холод Н.И. Высшая математика. Математическое программирование: учеб. пособие для эконом. спец. вузов. – Минск: Выш. шк., 1994. – 221 с.
9. Кузнецов А.В., Холод Н.И., Костевич Л.С. Руководство к решению задач по математическому программированию: Учеб. пособие/ Под общ. ред. А.В. Кузнецова. – 2-ое изд., перераб. и доп. – Мн.: Выш. шк., 2001. – 448 с.
10. Кузнецов Ю. И., Кузубов В. И, ВолощенкоА.В. Математическое программирование: Учеб. пособие. – 2-е изд., перераб. и доп. – М.: Высш. школа, 1980. – 300 с.
11. Мурашко В.С. Практическое пособие к выполнению лабораторных работ по курсу «Математическое моделирование технологических задач в машиностроении» для студ. спец. Т03.01.01 – «Технология машиностроения». – Гомель: ГГТУ им. П.О. Сухого, 1999. – 60с.
12. Мурашко В.С. Основы систем автоматизированного проектирования: прак. рук. к контрольным работам по одноим. курсу для студентов заоч. отд-ния специальностей 36 01 01 «Технология машиностроения» и 36 01 03 «Технологическое

оборудование машиностроительного производства». – Гомель: ГГТУ им. П.О. Сухого, 2004. – 36 с.

13. Мурашко, В. С. Математическое моделирование и алгоритмизация инженерных задач: учебное пособие для вузов / В. С. Мурашко. - Гомель : ГГТУ им. П. О. Сухого, 2012 – 225 с.
14. Мурашко, В. С. Математическое моделирование и алгоритмизация инженерных задач: лабораторный практикум для вузов / В. С. Мурашко. – Гомель: ГГТУ им. П. О. Сухого, 2012 – 151 с.
15. Просветов Г. И. Дискретная математика: задачи и решения: Учеб. пособие. – М.: БИНОМ. Лаборатория знаний, 2008. – 222 с.
16. Пляскин И.И. Оптимизация технических решений в машиностроении . – Москва: Машиностроение, 1982. – 176 с.
17. Сборник задач и упражнений по высшей математике: Математическое программирование: Учеб. пособие / Под общ. ред. Кузнецова А.В., Рутковского Р.А. – 2-ое изд., перераб. и доп. – Мн.: Выш. шк., 2002. – 447 с.
18. Сакович В.А. Исследование операций (детерминированные методы): Справочное пособие. – Мн.: Выш. шк.,1985. – 256с.
19. Солонин И. С. Математическая статистика в технологии машиностроения. – М.: Машиностроение, 1972. – 216с.
20. Тарасик В.П. Математическое моделирование технических систем: Учебник для вузов.- Мн.: Дизайн ПРО, 1998. – 640 с.
21. Таха Х. Введение в исследование операций. – 7-е изд. – М.: Издательский дом «Вильямс», 2005. – 912 с.
22. Тимковский В.Г. Дискретная математика в мире станков и деталей – Москва: Наука, 1992 – 145 с.

Содержание

Мурашко Валентина Семеновна

МЕТОДЫ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В МАШИНОСТРОЕНИИ

Пособие
по курсу «Математическое моделирование и методы исследования операций» для студентов специальности 1-53 01 01 «Автоматизация технологических процессов и производств (по направлениям)» дневной формы обучения

Подписано к размещению в электронную библиотеку
ГГТУ им. П. О. Сухого в качестве электронного
учебно-методического документа 07.12.18.

Рег. № 46Е.
http://www.gstu.by