Тема лекции №5

Планирование имитационных экспериментов.

Цель лекции – изучить основные правила построения и проведения имитационных экспериментов.

План лекции.

1. Сущность и цели планирования эксперимента.
2. Элементы стратегического планирования экспериментов.
3. Элементы тактического планирования.
4. Точность и количество реализаций модели при определении средних значений параметров.
5. Точность и количество реализаций модели при определении вероятностей исходов.

1. Сущность и цели планирования эксперимента

Для организации экспериментов наиболее важно:

1. Простота повторений условий эксперимента.
2. Возможность управления экспериментом, включая его прерывание и возобновление.
3. Легкость изменения условий проведения эксперимента (воздействий внешней среды).
4. Исключение корреляции между последовательностями данных, снимаемых в процессе эксперимента с моделью.
5. Определение временного интервала исследования модели $(0, T)$.

Имитационный (Компьютерный) эксперимент

представляет собой процесс использования модели с целью получения и анализа интересующей исследователя информации о свойствах моделируемой системы.

***План эксперимента* определяет:**

- объем вычислений на компьютере;
- порядок проведения вычислений на компьютере;
- способы накопления и статистической обработки результатов моделирования.

Планирование экспериментов имеет следующие цели:

- сокращение общего времени моделирования при соблюдении требований к точности и достоверности результатов;
- увеличение информативности каждого наблюдения;
- создание структурной основы процесса исследования.

Средством достижения приемлемого компромисса между максимумом информации и минимумом затрат ресурсов является **план эксперимента**.

Весь комплекс действий *по планированию эксперимента* разделяют на две самостоятельные функциональные части:

- **стратегическое планирование;**
- **тактическое планирование.**

Стратегическое планирование - разработка условий проведения эксперимента, определение режимов, обеспечивающих наибольшую информативность эксперимента.

Тактическое планирование обеспечивает достижение заданных точности и достоверности результатов.

2. Элементы стратегического планирования экспериментов.

Формирование стратегического плана выполняется в так называемом *факто*

Факторное простра... *них* и *внутренних па*... *рых* исследователь мож... *овки* и проведения эксперимента.

Экзогенные переменные — это переменные, задающиеся извне, значения которых задаются вне модели.

Объектами стратегического планирования являются:

- выходные переменные (отклики, реакции, *экзогенные переменные*);
- входные переменные (факторы, *эндогенные переменные*);
- уровни факторов.

Эндогенные переменные — это переменные, значение которых формируется внутри модели.

Математические методы планирования экспериментов основаны на так называемом кибернетическом представлении процесса проведения эксперимента

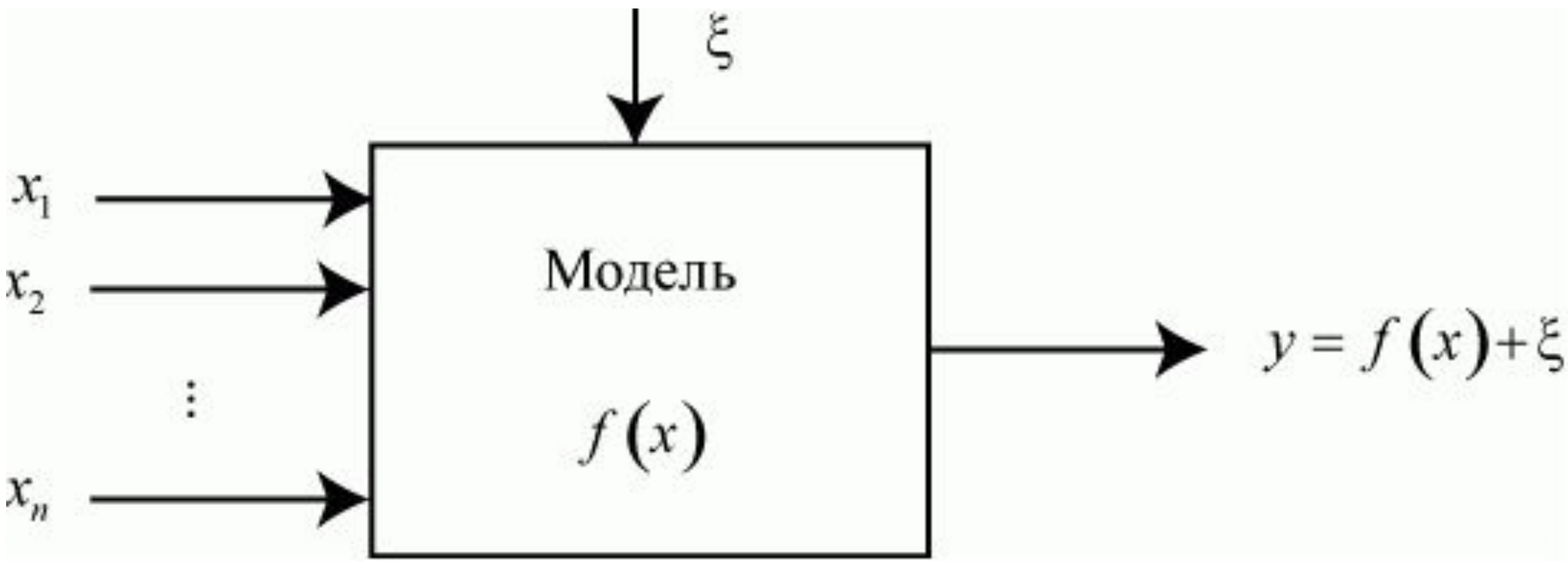

Рисунок 2.1 - Кибернетическое представление эксперимента

На рисунке 2.1:

$x_i, i=\overline{1,n}$ – âõîäíûå ïåðåìåííûå;

$y = f(x) + \xi$ – âûõîäíàÿ ïåðåìåííàÿ;

ξ – îøèáêà, ïîìåõà, âûçûâàåìàÿ

íàëè÷èåì ñëó÷àéíûõ ôàêòîðîâ;

$f(x)$ – îïåðàòîð, ìîäåëèðóþùèé äåéñòâèå

ðåàëüíîé ñèñòåìû, îïðåäåëÿþùèé çàâèñèìîñòü

âûõîäíîé ïåðåìåííîé îò ôàêòîðîâ.

Проблемы, решаемые при стратегическом планировании

1) выбор отклика (реакции), то есть *определение*, какие величины нужно измерять во время эксперимента, чтобы получить искомые ответы.

2) выбор (определение) существенных факторов и их сочетаний, влияющих на работу моделируемого объекта.

В теории систем приводится так называемый принцип Парето:

- 20% факторов определяют 80% свойств системы;
- 80% факторов определяют 20% свойств системы.

Следовательно, надо уметь выделять существенные факторы. А это достигается достаточно глубоким изучением моделируемого объекта и протекающих в нем процессов.

3) выбор значений каждого фактора, называемых *уровнями фактора*.

Анализ данных эксперимента существенно упрощается, если назначить уровни факторов, равноотстоящие друг от друга. Такой план называется *ортогональным*. Ортогональность плана обычно достигают так: две крайние точки области изменения фактора выбирают как два уровня, а остальные уровни располагают так, чтобы они делили полученный *отрезок* на две части.

Эксперимент, в котором реализуются все *сочетания* уровней всех факторов, называется *полным факторным экспериментом* (ПФЭ).

Число измерений откликов (реакций) модели N_c при ПФЭ равно

$$N_c = q_1 \cdot q_2 \cdot \ldots \cdot q_k,$$

где q_i - число уровней i -го фактора, $i = \overline{1,k};$
k - число факторов эксперимента.

Величина N_c определяет структуру стратегического плана, то есть количество наблюдений (информационных точек).

При компьютерной реализации ПФЭ в каждом наблюдении (информационной точке) нужно выполнить определенное число прогонов (реализаций) модели, чтобы обеспечить заданную *точность* и *достоверность* значений откликов. *Определение* числа прогонов модели является предметом тактического планирования.

Обозначим число прогонов в каждом наблюдении p. Тогда для симметричного ПФЭ общее число N необходимых прогонов модели равно:

$$N = p \cdot q^k$$

Пример

Планируется провести компьютерный эксперимент, в котором на отклик модели влияют три фактора. Для каждого фактора установлены три уровня. Требования по точности и достоверности провести 6000 прогонов модели на каждом уровне (для каждого наблюдения). Время одного прогона модели равно 2 с.

Оценить *затраты времени* на проведение компьютерного эксперимента.

Исходные данные: число факторов k=5; число уровней q=2; количество наблюдений p=6000; время одного прогона t_p=1/30мин.

Решение.

Число прогонов: $$N = p \cdot q^k = 6000 \cdot 3^5 =$$

Затраты времени (ч.): $$T = \frac{N \cdot t_p}{60} =$$

3. Элементы тактического планирования.

Основной задачей тактического планирования является обеспечение результатам компьютерного эксперимента заданных точности и достоверности.

Рассмотрим случай, когда имитационная модель строилась для определения характеристик некоторых случайных величин.

Характеристику случайной величины будем обозначать греческой буквой Θ («фита»).

С помощью имитационного моделирования точное *значение* Θ определить нельзя, так как число *N* реализаций модели конечно. При конечном числе реализаций модели определяется приближенное *значение* характеристики. Обозначим это приближение $\overline{\Theta}$

Приближенное *значение* называют *оценка* соответствующей характеристики: оценкой матожидания, оценкой дисперсии, оценкой коэффициента корреляции.

Точностью характеристики $\overline{\Theta}$ называют величину ε в отношении

$$\left|\overline{\Theta} - M(\Theta)\right| < \varepsilon,$$

где $M(\Theta)$ - матожидание случайной величины.

Величина ***ε*** представляет собой абсолютное значение ошибки в определении значения искомой характеристики. Достоверность оценки характеристики $\overline{\Theta}$ называют вероятность ***α*** того, что заданная точность достигается:

$$P(\left|\overline{\Theta} - M(\Theta)\right| < \varepsilon) = \alpha.$$

Достоверность характеризует повторяемость, устойчивость эксперимента и трактуется так:

если для оценки *M(Θ)* использовать величину $\overline{\Theta}$ то в среднем на каждые 1000 применений этого правила в *1000·α* случаев величина $\overline{\Theta}$ будет отличаться от *M(Θ)* на величину меньше ε.

В ряде случаев целесообразно пользоваться понятием относительной точности

$$d = \frac{\varepsilon}{M(\Theta)}.$$

В этом случае *достоверность* оценки имеет вид:

$$P(\frac{\overline{\Theta} - M(\Theta)}{M(\Theta)} < d) = \alpha.$$

4. Точность и количество реализаций модели при определении средних значений параметров.

Найдем функциональную *связь* точности ***e*** и достоверности ***a*** с количеством реализаций модели, когда в качестве показателей эффективности выступают матожидание и *дисперсия* некоторой случайной величины (времени, расстояния и т. п.).

Найдем искомую связь для случая, когда целью эксперимента является определение оценки матожидания некоторой случайной величины.

В N прогонах модели получены независимые значения интересующего нас показателя эффективности:

$$a_1, a_2, ..., a_i, ..., a_N.$$

В качестве оценки матожидания возьмем выборочное среднее (среднее арифметическое):

$$\bar{a} = \frac{\sum_{i=1}^{N} a_i}{N}$$

Согласно центральной предельной теореме, если значения a_i независимы и имеют конечные дисперсии одного порядка, то при большом числе слагаемых N случайная величина $\bar{a}$ имеет практически нормальное распределение с матожиданием и дисперсией соответственно:

$$M(\bar{a}) = M(a), \sigma_{\bar{a}}^2 = \frac{\sigma_a^2}{N}, \sigma_{\bar{a}} = \frac{\sigma_a}{\sqrt{N}}$$

где σ_a - дисперсия искомой случайной величины a.

Следовательно, справедливо

$$P(\left|\bar{a} - M(a)\right| < t_a \sigma_{\bar{a}}) = \Phi^*(t_a)$$

где $\Phi^*(t_a) = \frac{2}{\sqrt{2\pi}} \int_0^{t_a} e^{-\frac{z^2}{2}} dz$ - интеграл вероятности.

Для определения интеграла вероятности вводят понятие интеграл Лапласа $\Phi(t_a)$:

$$\Phi^* = 2\Phi(t_a)$$

Таким образом

$$P(\left|\bar{a} - M(a)\right| < t_a \sigma_{\bar{a}}) = 2\Phi(t_a)$$

Интеграл Лапласа табулирован, следовательно, задаваясь значением достоверности *a*, определяется аргумент t_a.

Итак, искомая связь между точностью ε, достоверностью и числом реализаций модели получена:

$$\varepsilon = t_a \frac{\sigma_a}{\sqrt{N}}, N = t_a^2 \frac{\sigma_a^2}{\varepsilon^2}.$$

Из полученного выражения следует:

- увеличение точности на порядок (уменьшение ошибки на порядок) потребует увеличения числа реализаций на два порядка;
- число необходимых реализаций *N* модели не зависит от величины искомого параметра ***a***, а зависит от дисперсии.

Достоверность результата ***a*** указана значением аргумента функции Лапласа t_a .

Связь значения t_a с ***a*** находится из таблицы значений функции (интеграла) Лапласа. Наиболее употребительные соответствия t_a и ***a*** приведены в таблице.

Таблица - Фрагмент таблицы функции (интеграла) Лапласа

a	0,8	0,85	0,9	0,95	0,99	0,995	0,999
t_a	1,28	1,44	1,65	1,96	2,58	2,81	3,30

Рассмотрим задачу определения оценки *дисперсии* S^2 случайной величины ***a*** также с заданными точностью и достоверностью.

Приведем окончательный вид формул для расчета:

$$N = t_a^2 \frac{(\mu_4 - \sigma^4)}{\varepsilon^2};$$

$$\varepsilon = t_a \sqrt{\frac{(\mu_4 - \sigma^4)}{N}}$$

где μ_4 - эмпирический центральный момент четвертого порядка:

$$\mu_4 = \frac{1}{N} \sum_{i=1}^{N} (a_i - \bar{a})^4$$

Если определяемая случайная величина имеет нормальное распределение, то $\mu_4 = 3a^4 \approx 3S^4$

и выражения для *N* и *ε* принимают вид

$$N = t_a^2 \frac{2S^4}{\varepsilon^2}; \varepsilon = t_a \frac{S^2\sqrt{2}}{\sqrt{N}}.$$

при малых значениях *N* (*N*<*120*) следует использовать параметр распределения Стьюдента t^*_a.

Пример.

В результате предварительных прогонов модели *N=1000* определена оценка дисперсии S^2=10 ед^2.

Определить число реализаций модели N_1(обычное распределение СВ) и N_2(нормальное распределение СВ) для определения оценок матожидания и дисперсии случайной величины ***a*** соответственно с точностью ε=0,1 и достоверностью 0,9.

Решение.

$$N_1 = t_a^2 \frac{S^2}{\varepsilon^2} =$$

$$N_2 = t_a^2 \frac{2S^4}{\varepsilon^2} =$$

5. Точность и количество реализаций модели при определении вероятностей исходов

Рассмотрим случай, когда в качестве показателя эффективности выступает *вероятность* свершения (или не свершения) какого-либо события, например, поражения цели, выхода из строя техники, завершения комплекса работ в заданное время и др.

В качестве оценки вероятности ***P*** события $\overline{a}$ выступает частота его свершения:

$$\overline{P} = \frac{m}{N},$$

где *m* - число реализаций модели;

N - число свершений данного события.

Использование частоты в качестве оценки искомой вероятности основано на теореме Я. Бернулли, которую в данном случае можно в формализованном виде записать так:

$$\lim_{N \to \infty} \frac{m}{N} = P$$

Точность и *достоверность* этой оценки связаны уже с известным определением достоверности:

$$P(\left|\overline{P} - P\right| < \varepsilon) = a$$

Задача сводится к нахождению такого количества реализаций *N,* чтобы оценка $\overline{P}$ отличалась от искомого значения *P* менее, чем на *ε* с заданной достоверностью. Здесь, как и ранее, *ε* - абсолютное *значение*, характеризующее *точность* оценки.

Для нахождения функциональной связи между точностью, достоверностью и числом реализаций модели введем переменную x_i - результат исхода i-й реализации модели:

$$x_i = \begin{cases} 1, & \text{событие свершилось, вероятность } P, \\ 0, & \text{событие не свершилось, вероятность } 1 - P. \end{cases}$$

Тогда частота свершения события (оценка искомой вероятности) будет определяться следующим выражением:

$$\overline{P} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Величина $\sum_{i=1}^{N} x_i$ случайная и дискретная. Она при таком задании x_i имеет *биномиальное распределение* (*распределение Бернулли*) с характеристиками:

матожидание: $$M\left[\sum_{i=1}^{N} x_i\right] = NP$$

дисперсия: $$D\left[\sum_{i=1}^{N} x_i\right] = NP(1-P)$$

Из этого следует

$$M\left[\overline{P}\right] = M\left[\frac{\sum_{i=1}^{N} x_i}{N}\right] = \frac{1}{N} NP = P$$

$$D\left[\overline{P}\right] = D\left[\frac{\sum_{i=1}^{N} x_i}{N}\right] = \frac{1}{N^2} NP(1-P) = \frac{P(1-P)}{N} \qquad \sigma_{\overline{P}} = \sqrt{D\left[\overline{P}\right]} = \sqrt{\frac{P(1-P)}{N}}$$

В теории вероятностей есть теорема Лапласа (частный случай центральной предельной теоремы), сущность которой состоит в том, что при больших значениях числа реализаций *N биномиальное распределение* достаточно хорошо согласуется с нормальным распределением. Следовательно, можно записать:

$$P(\left|\overline{P} - P\right| < t_a \sigma_{\overline{P}}) = 2\Phi(t_a) \Rightarrow \quad P(\left|\overline{P} - P\right|) < t_a \sqrt{\frac{P(1-P)}{N}}) = 2\Phi(t_a).$$

Получим искомые формулы:

$$\varepsilon = t_a \sqrt{\frac{P(1-P)}{N}}, \quad N = t_a^2 \frac{P(1-P)}{\varepsilon^2}.$$

Если априорные сведения хотя бы о порядке искомой вероятности *P* неизвестны, то использование значения абсолютной ошибки *ε* может не иметь смысла. Например, может быть так, что исследователь задал значение абсолютной ошибки *ε*=1, а искомое *значение* вероятности оказалось *P*=0,01. Очевидно, явное несоответствие. Поэтому целесообразно оперировать *относительной погрешностью* $d = \dfrac{\varepsilon}{P}$

В этом случае:

$$d = t_a \sqrt{\frac{1-P}{PN}}, \quad N = t_a^2 \frac{1-P}{Pd^2}.$$

Из формул следует, что при определении оценок малых вероятностей с приемлемо высокой точностью необходимо выполнить очень большое число реализаций модели.

При отсутствии высокопроизводительного компьютера применение статистического моделирования становится проблематичным.

Пример.

Вероятность наступления события P=0,1.Определить число *реализаций модели* и *затраты машинного времени* для оценки данной вероятности с относительной точностью d=0,01 и достоверностью a=0,9. На выполнение одной реализации модели требуется 5 сек.

Решение.

Количество реализаций модели

$$N = t_a^2 \frac{1-P}{Pd^2} = 1,65^2 \frac{1-0,1}{0,1 \cdot 0,01^2} =$$