

## Тема 2. Методы анализа математических моделей

### Лекция 6. Стохастические методы. Имитационное моделирование

#### 1. Стохастические методы

**Стохастичность** (цель, предположение) означает случайность. **Случайный (стохастический) процесс** — это процесс, поведение которого не является детерминированным, и последующее состояние такой системы описывается как величинами, которые могут быть предсказаны, так и случайными. Однако, по М. Кацу и Э. Нельсону, любое развитие процесса во времени (неважно, детерминированное или вероятностное) при анализе в терминах вероятностей будет случайным процессом (иными словами, все процессы, имеющие развитие во времени, с точки зрения теории вероятностей, стохастические).

Использование термина стохастичность в математике относят к работам Владислава Борткевича, который использовал его в значении выдвигать гипотезы, которое, в свою очередь, отсылает нас к древнегреческим философам, а также к работе Я. Бернулли *Ars Conjectandi* (лат. искусство загадывать).

В математике имеется большой раздел – теория вероятностей и математическая статистика, связанный с наблюдениями и исследованиями случайных величин и процессов. Приложения этой теории используются в других разделах в качестве методического аппарата. Например, методы оптимизации часто используют случайные процессы и статистическую обработку данных.

В области искусственного интеллекта стохастические программы работают с использованием вероятностных методов. Примерами таких алгоритмов могут служить: алгоритм имитации отжига, стохастические нейронные сети, стохастическая оптимизация, генетические алгоритмы. Стохастичность в данном случае может содержаться как в самой проблеме, так и в планировании в условиях неопределённости. Для агента моделирования детерминированное окружение более простое, нежели стохастическое.

В естественных науках также часто рассматриваются случайные процессы. Примером реального случайного процесса в нашем мире может служить моделирование давления газа при помощи Винеровского процесса. Несмотря на то, что каждая молекула газа движется по своему строго определённой пути (в данной модели, а не в реальном газе), движение совокупности таких молекул практически нельзя просчитать и предсказать. Достаточно большой набор молекул будет обладать стохастическими свойствами, такими как наполнение сосуда, выравнивание давления, движение в сторону меньшего градиента концентрации и т. д. Таким образом проявляется эмерджентность системы (то есть появление у системы свойств, не присущих её элементам в отдельности или несводимость свойств системы к сумме свойств её компонентов).

Все более широкое распространение в современных научных исследованиях получают математико-статистические модели: корреляционные и регрессионные, метод главных компонент, кластер-анализ, метод статистических испытаний (метод Монте-Карло), допускающие использование неполной вероятностной (стохастической) связи между исследуемыми показателями.

#### 1.1 Корреляционный анализ

Исследование связей между несколькими показателями какой-либо системы требует статистических наблюдений за всеми изучаемыми показателями. Для установления взаимосвязей между компонентами системы служит **корреляция** (от англ. correlation – соотношение, соответствие, взаимозависимость), которая дает возможность установить, насколько средняя величина одного из показателей меняется в зависимости от другого (других).

Задачи корреляции решают всегда при фиксированном числе учитываемых признаков, т.е. тех, значения которых известны исследователю для каждого элемента системы. Число учитываемых признаков определяется не только средствами, которыми располагает исследователь для решения задачи, но и постановкой задачи, которая видоизменяется и углубляется по мере увеличения этих признаков.

При решении многих исследовательских задач возникает необходимость в установлении связи между  $k$  аргументами  $X_1, X_2, \dots, X_k$  и зависящей от них величиной  $y$ . Возможны следующие основные типы взаимосвязей между переменными величинами.

- **Зависимость между неслучайными величинами.** В этом случае зависимая переменная  $y$  вполне определенно задается независимыми переменными  $X_1, X_2, \dots, X_k$ .

- *Зависимость между случайными величинами.* Между ними может существовать функциональная или стохастическая связь. Последняя проявляется в том, что одна из случайных величин реагирует на изменение другой изменениями своего закона распределения.
- *Зависимость случайной величины  $y$  от неслучайных переменных  $X_1, X_2, \dots, X_k$ .* Природа этой взаимосвязи может быть двоякой:
  - а) определение зависимой переменной  $y$  связано с некоторыми ошибками измерения, а переменные  $X_1, X_2, \dots, X_k$  измеряются без ошибок или эти ошибки пренебрежимо малы по сравнению с ошибкой измерения зависимой переменной;
  - б) значения переменной  $y$  зависят не только от соответствующих значений  $X_1, X_2, \dots, X_k$ , но и от ряда неконтролируемых факторов, поэтому при каждом сочетании значений  $X_1, X_2, \dots, X_k$  зависимая переменная  $y$  подвержена колебаниям случайного характера.

## 1.2 Регрессионный анализ

Часто возникает необходимость в установлении связи между случайной величиной  $y$  и неслучайными переменными  $X_1, X_2, \dots, X_k$ , принимающими в каждой серии опытов определенные значения. Величина  $y$  является случайной, имеющей нормальное распределение, с переменным центром распределения  $M(y)$ , изменяющимся при разных сочетаниях значений факторов  $X_1, X_2, \dots, X_k$ . Случайная величина  $y$  имеет постоянную дисперсию, т.е. не зависящую от  $X_1, X_2, \dots, X_k$ .

Таким образом,  $M(y)$  является функцией  $X_1, X_2, \dots, X_k$ , т.е. на каждое изменение неслучайных величин  $X_1, X_2, \dots, X_k$  случайная величина  $y$  реагирует изменением своего математического ожидания  $M(y)$ . Выражение  $M(y) = F(X_1, X_2, \dots, X_k)$  называют уравнением регрессии математического ожидания случайной величины  $y$  по неслучайным величинам  $X_1, X_2, \dots, X_k$ .

В основе *регрессионного анализа* лежат следующие предположения:

- При каждом значении  $X_1, X_2, \dots, X_k$  величина  $y$  имеет нормальное распределение.
- Дисперсия теоретического распределения случайной величины  $y$  постоянна.
- Тип функции  $M(y) = F(X_1, X_2, \dots, X_k)$  известен.
- Независимые переменные  $X_1, X_2, \dots, X_k$  измеряются с пренебрежимо малой ошибкой по сравнению с ошибкой в определении  $y$ .
- Переменные  $X_1, X_2, \dots, X_k$  линейно независимы.

В случае исследования параметров  $X_1, X_2, \dots, X_k$  различных технологических процессов часто приходится иметь дело со случайной величиной  $y$ , не подчиняющейся нормальному распределению.

При корреляционном или регрессионном анализе на основе полученной корреляционной матрицы стремятся, например, получить уравнения регрессии, связывающие факторные признаки с результативными. Сами уравнения регрессии являются конечной целью исследования. По ним выполняется содержательная интерпретация получаемых результатов и вырабатываются соответствующие решения.

## 1.3 Метод главных компонент

*Метод главных компонент* является разновидностью регрессионного анализа. В *методе главных компонент* корреляционная матрица используется как исходная ступень для дальнейшего анализа наблюдаемых признаков. Появляется возможность извлечения дополнительной информации об изучаемом элементе или процессе, причем весьма ценная новая информация получается на основе статистических данных, ранее собранных для проведения классического регрессионного анализа.

*Методом главных компонент* можно решить задачи четырех основных типов:

- отыскание скрытых, но объективно существующих закономерностей, определяемых воздействием внутренних и внешних причин;
- описание изучаемого процесса числом главных компонент  $m$ , значительно меньшим, чем число первоначально взятых признаков  $k$ ;
- выявление и изучение стохастической связи признаков с главными компонентами;
- прогнозирование хода развития процесса на основе уравнения регрессии, построенного по полученным главным компонентам.

Практические возможности решения вышеуказанных типов задач реализуются в следующих направлениях:

- причинный анализ взаимосвязей показателей и определение их стохастической связи с главными компонентами;

- построение обобщенных технико-экономических показателей;
- ранжирование объектов или наблюдений по главным компонентам;
- классификация объектов наблюдений;
- ортогонализация исходных показателей;
- сжатие исходной информации;
- построение уравнений регрессии по обобщенным технико-экономическим показателям.

#### 1.4 Кластерный анализ

**Кластерный анализ** – способ группировки многомерных объектов, основанный на представлении результатов отдельных наблюдений точками подходящего геометрического пространства с последующим выделением групп (скоплений) этих точек.

Термин «кластерный анализ» (от англ. cluster — *гроздь, скопление, пучок*) предложен в 1939 г. К. Трионом. Синонимами выступают выражения: *автоматическая классификация, таксономия, распознавание без обучения, распознавание образов без учителя, самообучение* и др.

Основная цель кластерного анализа – выделение в исходных многомерных множествах данных таких однородных подмножеств, внутри которых все объекты похожи между собой в определенном смысле, а объекты из разных подмножеств напротив не похожи.

Основными подходами при кластерном анализе являются:

- *Вероятностно-статистический*, предполагающий выделение групп, каждая из которых представляет собой реализацию некоторой случайной величины.
- *Структурный*, который предусматривает выделение компактных групп объектов, удаленных один от другого, отыскивает «естественное» разделение совокупности на области скопления объектов; используется для двух видов исходных данных: для матриц близости, или расстояний между объектами, и для объектов, представленных как точки в многомерном пространстве.
- *Вариативный* (нормативный), заключающийся в разделении совокупности по некоторому признаку на группы в соответствии с определенными интервалами, причем характер распределения объектов практически не влияет на выбор интервалов и на число групп.

#### 1.5 Дисперсионный анализ

**Дисперсионный анализ** предназначен для работы с результатами наблюдений, зависящих от различных одновременно действующих факторов. При таком анализе выделяют доминирующие факторы и количественно оценивают степень их влияния на конечные результаты эксперимента. Дисперсионный анализ основан на предположении, что существенность фактора определяется его вкладом в дисперсию конечных результатов. Различают *однофакторный* и *многофакторный* дисперсионный анализ. Дисперсионный анализ является одним из эффективных средств выделения наиболее существенных факторов при планировании эксперимента.

#### 1.6 Метод статистических испытаний (Монте-Карло)

Нередко исследуемый процесс моделируют многократными повторениями случайных вариантов его реализации. Поведение системы при этом изучают, сопоставляя воздействия на вход в систему с результатами, полученными на выходе из нее. Происходящие внутри системы процессы не рассматривают, поскольку внутренние взаимосвязи, характерные для системы, неизвестны. Единичные реализации называются *статистическими испытаниями*, поэтому рассматриваемый метод называют методом статистических испытаний, или *методом Монте-Карло*.

Метод Монте-Карло (методы Монте-Карло, ММК) – общее название группы методов, основанных на получении большого числа реализаций стохастического (случайного) процесса. Случайный процесс воспроизводится физическими средствами (приборами для генерации случайных величин) либо формируется компьютерной системой по специальному алгоритму таким образом, чтобы его вероятностные характеристики совпадали с аналогичными величинами решаемой задачи. ММК используется для решения задач в различных областях физики, химии, математики, экономики, оптимизации, теории управления и др..

Метод Монте-Карло применяют, когда затруднительно или даже невозможно построить аналитическую модель. Таким образом, его использование целесообразно при решении задач, допускающих теоретико-вероятностное описание, поскольку появляется возможность упрощения промежуточных этапов расчета. Рассматриваемый метод включает три составные части:

- моделирование случайных величин с заданным законом распределения;
- построение вероятностных моделей реальных систем;
- решение задачи статистической теории.

### 1.7 Пример – интегрирование методом Монте-Карло

Предположим, необходимо взять интеграл от некоторой функции. Воспользуемся неформальным геометрическим описанием интеграла и будем понимать его как площадь под графиком этой функции. Для определения этой площади можно воспользоваться одним из обычных численных методов интегрирования: разбить отрезок на подотрезки, подсчитать площадь под графиком функции на каждом из них и сложить. Предположим, что для функции, представленной на рисунке 2, достаточно разбиения на 25 отрезков и, следовательно, вычисления 25 значений функции. Представим теперь, мы имеем дело с  $n$ -мерной функцией. Тогда нам необходимо  $25^n$  отрезков и столько же вычислений значения функции. При размерности функции больше 10 задача становится огромной.

Поскольку пространства большой размерности встречаются, в частности, в задачах теории струн, а также многих других физических задачах, где имеются системы со многими степенями свободы, необходимо иметь метод решения, вычислительная сложность которого бы не столь сильно зависела от размерности. Именно таким свойством обладает метод Монте-Карло.

#### **Обычный алгоритм Монте-Карло интегрирования.**

Предположим, требуется вычислить определённый интеграл

$$I = \int_a^b f(x) dx.$$

Рассмотрим случайную величину  $u$ , равномерно распределённую на отрезке интегрирования  $[a, b]$ . Тогда  $f(u)$  также будет случайной величиной, причём её математическое ожидание выражается как

$$\langle f(u) \rangle = \int_a^b f(x) \varphi(x) dx, \quad \varphi(x) = \frac{1}{b-a},$$

где  $\varphi(x)$  – плотность распределения случайной величины на участке  $[a, b]$ . Таким образом, искомый интеграл выражается как

$$I = \int_a^b f(x) dx = (b-a) \langle f(u) \rangle.$$

Но математическое ожидание случайной величины  $f(u)$  можно легко оценить, смоделировав эту случайную величину и рассчитав выборочное среднее.

Итак, выбираем  $N$  точек, равномерно распределённых на  $[a, b]$ . Для каждой точки  $u_i$  вычисляем  $f(u_i)$ . Затем вычисляем выборочное среднее

$$\frac{1}{N} \sum_{i=1}^N f(u_i),$$

и получаем оценку интеграла:

$$I \approx \frac{(b-a)}{N} \sum_{i=1}^N f(u_i).$$

Точность оценки зависит только от количества точек  $N$ .

Этот метод имеет и геометрическую интерпретацию. Он похож на детерминистический численный метод вычисления интеграла. Разница состоит лишь в том, что вместо равномерного деления области интегрирования на маленькие интервалы и суммирования площадей получившихся "столбиков" здесь отрезок интегрирования покрывается случайными точками, для каждой из которых строится такой же "столбик" и его площадь складывается с остальными.

#### **Геометрический алгоритм Монте-Карло интегрирования**

Набрасываем двумерные точки на прямоугольник  $[a, b] \times [c, d]$  ( $c$  и  $d$  – нижняя и верхняя грани значений функции  $y = f(x)$  на отрезке  $[a, b]$ ) и суммируем только те ( $N_1$ ), которые лежат под кривой

$y = f(x)$ . Интеграл оценим как  $I \approx (b-a)(d-c) \frac{N_1}{N}$ .

## 2. Имитационное моделирование

**Имитационное моделирование** (ситуационное моделирование) – метод, позволяющий строить модели, описывающие процессы так, как они проходили бы в действительности. Такую модель можно "проиграть" во времени как для одного испытания, так и для заданного их множества. При этом результаты будут определяться случайным характером процессов. По этим данным можно получить достаточно устойчивую статистику.

**Имитационное моделирование** – метод исследования, при котором изучаемая система заменяется моделью, с достаточной точностью описывающей реальную систему, с которой проводятся эксперименты с целью получения информации об этой системе. Экспериментирование с моделью называют имитацией (имитация – это постижение сути явления, не прибегая к экспериментам на реальном объекте).

Имитационное моделирование – частный случай математического моделирования. Существует класс объектов, для которых по различным причинам не разработаны аналитические модели, либо не разработаны методы решения полученной модели. В этом случае аналитическая модель заменяется имитатором или имитационной моделью.

Имитационным моделированием иногда называют получение частных численных решений сформулированной задачи на основе аналитических решений или с помощью численных методов.

Имитационная модель – логико-математическое описание объекта, которое может быть использовано для экспериментирования на компьютере в целях проектирования, анализа и оценки функционирования объекта.

К имитационному моделированию прибегают, когда:

- дорого или невозможно экспериментировать на реальном объекте;
- невозможно построить аналитическую модель: в системе есть время, причинные связи, последствие, нелинейности, стохастические (случайные) переменные;
- необходимо симитировать поведение системы во времени.

Цель имитационного моделирования состоит в воспроизведении поведения исследуемой системы на основе результатов анализа наиболее существенных взаимосвязей между её элементами или другими словами – разработке симулятора исследуемой системы для проведения различных экспериментов.

### 2.1 Виды имитационного моделирования

- **Системная динамика** – парадигма моделирования, где для исследуемой системы строятся графические диаграммы причинных связей и глобальных влияний одних параметров на другие во времени, а затем созданная на основе этих диаграмм модель имитируется на компьютере. По сути, такой вид моделирования более всех других парадигм помогает понять суть происходящего выявления причинно-следственных связей между объектами и явлениями. С помощью системной динамики строят модели бизнес-процессов, развития городов, модели производства, динамики популяции, экологии и развития эпидемии. Метод основан [Джеем Форрестером](#) в 1950 годах.
- **Дискретно-событийное моделирование** – подход к моделированию, предлагающий абстрагироваться от непрерывной природы событий и рассматривать только основные события моделируемой системы, такие, как: «ожидание», «обработка заказа», «движение с грузом», «разгрузка» и другие. Дискретно-событийное моделирование наиболее развито и имеет огромную сферу приложений – от логистики и систем массового обслуживания до транспортных и производственных систем. Этот вид моделирования наиболее подходит для моделирования производственных процессов. Основан [Джеффри Гордоном](#) в 1960-х годах.
- **Агентное и мультиагентное моделирование** – относительно новое (1990-е гг.) направление в имитационном моделировании, которое используется для исследования децентрализованных систем, динамика функционирования которых определяется не глобальными правилами и законами (как в других парадигмах моделирования), а наоборот, когда эти глобальные правила и законы являются результатом индивидуальной активности членов группы. Цель агентных моделей – получить представление о глобальных правилах общего поведения системы, исходя из предположений об индивидуальном, частном поведении её отдельных активных объектов и взаимодействии этих объектов в системе. Агент – некая сущность, обладающая активностью, автономным поведением, может принимать решения в соответствии с некоторым набором правил, взаимодействовать с окружением, а также самостоятельно изменяться.