



МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Авторы: С. А. Айвазян

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ, раздел [математической статистики](#), посвящённый математич. методам построения оптимальных планов сбора, систематизации, обработки и интерпретации многомерных статистич. данных. Эти методы предназначены прежде всего для выявления характера и структуры взаимосвязей между компонентами исследуемого многомерного признака и используются для получения теоретич. и практич. выводов. Под многомерным признаком понимается p -мерный вектор $x=(x_1, x_2, \dots, x_p)$ показателей (переменных) x_1, x_2, \dots, x_p , среди которых могут быть количественные, т. е. измеряющие в определённой шкале степень проявления изучаемого свойства объекта; порядковые (или ординальные), т. е. позволяющие упорядочивать анализируемые объекты по степени проявления в них изучаемого свойства; классификационные (или номинальные), т. е. позволяющие разбивать исследуемую совокупность объектов на однородные (по анализируемому свойству) классы. Результаты измерения этих показателей на каждом из анализируемых объектов исследуемой совокупности образуют последовательность многомерных наблюдений, или исходный массив многомерных данных, который используется для проведения М. с. а. В значит. части М. с. а. рассматриваются ситуации, в которых исследуемый многомерный признак интерпретируется как многомерная случайная величина и анализируемая последовательность многомерных наблюдений – как выборка из [совокупности генеральной](#). В этом случае выбор методов обработки исходных статистич. данных и анализ их свойств производятся на основе тех или иных допущений относительно природы многомерного (совместного) закона распределения вероятностей.

По содержанию в М. с. а. выделяют 3 осн. раздела: М. с. а. многомерных распределений, М. с. а. структуры и характера взаимосвязей между компонентами исследуемого многомерного признака, М. с. а. геометрич. структуры исследуемой совокупности многомерных наблюдений.

Многомерный статистический анализ многомерных распределений

охватывает лишь ситуации, в которых обрабатываемые наблюдения имеют вероятностную природу, т. е. интерпретируются как выборка из некоторой генеральной совокупности. К осн. задачам этого раздела относятся: статистич. оценивание исследуемых многомерных распределений и их числовых характеристик; исследование свойств используемых [статистических оценок](#); исследование распределений вероятностей для статистик, с помощью которых строятся статистич. критерии проверки гипотез о вероятностной природе анализируемых многомерных данных.

Многомерный статистический анализ структуры и характера взаимосвязей компонент исследуемого многомерного признака

включает в себя понятия и результаты таких методов и моделей М. с. а., как [регрессионный анализ](#), [дисперсионный анализ](#), [факторный анализ](#), анализ многомерных временных рядов [под многомерным

временным рядом понимается последовательность наблюдений многомерных признаков (x_1, x_2, \dots, x_p) , произведённых во времени].

Многомерный статистический анализ геометрической структуры исследуемой совокупности многомерных наблюдений

объединяет в себе понятия и результаты таких моделей и схем, как дискриминантный анализ, анализ смесей вероятностных распределений, кластерный анализ, многомерное шкалирование. Осн. понятием во всех этих моделях и схемах является понятие расстояния (меры близости, меры сходства) между анализируемыми элементами.

Методы и результаты дискриминантного анализа направлены на решение следующей задачи. Известно о существовании определённого числа $k \geq 2$ генеральных совокупностей и имеется по одной выборке из каждой совокупности (обучающие выборки). Требуется построить основанное на имеющихся обучающих выборках наилучшее (в определённом смысле) классифицирующее правило, позволяющее приписать новый элемент (наблюдение) к своей генеральной совокупности, когда заранее неизвестно, к какой из совокупностей этот элемент принадлежит.

Задача анализа смесей вероятностных распределений чаще всего возникает в связи с исследованием «геометрической структуры» рассматриваемой совокупности. При этом предполагается, что распределение общей генеральной совокупности, из которой извлечена анализируемая выборка, описывается смесью распределений вида $P(x) = \sum_{r=1}^k \pi_r P_r(x|\theta_r)$, где r – номер некоторой однородной совокупности, характеризующейся вероятностным распределением $P_r(x|\theta_r)$, зависящим от параметра θ_r , π_r – априорная вероятность (удельный вес элементов) r -го класса в общей генеральной совокупности. Задача состоит в статистич. оценивании неизвестных параметров $\theta_r, \pi_r, r=1, \dots, k$. Это, в частности, позволяет свести задачу классификации элементов к схеме дискриминантного анализа, хотя в данном случае отсутствуют обучающие выборки.

Методы и результаты кластерного анализа (классификация, распознавание образов «без учителя») направлены на решение следующей задачи. Геометрич. структура анализируемой совокупности элементов задана либо координатами соответствующих точек, либо набором геометрич. характеристик их взаимного расположения, напр. матрицей попарных расстояний. Требуется разбить исследуемую совокупность элементов на сравнительно небольшое (заранее известное или нет) число классов так, чтобы элементы одного класса находились на небольшом расстоянии друг от друга, в то время как разные классы были бы по возможности достаточно удалены один от другого.

Задача многомерного шкалирования относится к ситуации, когда исследуемая совокупность элементов задана с помощью матрицы попарных расстояний, и заключается в приписывании каждому из элементов заданного числа координат таким образом, чтобы структура попарных расстояний между элементами, измеренных с помощью этих вспомогательных координат, в среднем наименее отличалась бы от заданной.

Прикладное назначение М. с. а. состоит в осн. в изучении следующих проблем.

Проблема статистического исследования зависимостей

между анализируемыми показателями. Предполагается, что исследуемый набор статистически регистрируемых показателей x разбит, исходя из содержательного смысла этих показателей и окончательных целей исследования, на q -мерный подвектор $x^{\{1\}}$ предсказываемых (зависимых) переменных и $(p-q)$ -мерный подвектор $x^{\{2\}}$ предсказывающих (независимых) переменных. Проблема состоит в определении на основании имеющихся наблюдений такой q -мерной векторной функции $f(x^{\{2\}})$ из класса допустимых решений F , которая давала бы наилучшую (в определённом смысле) аппроксимацию поведения подвектора показателей $x^{\{1\}}$.

Проблема классификации элементов

(объектов или показателей) в общей постановке заключается в том, чтобы всю анализируемую совокупность элементов разбить на сравнительно небольшое число однородных (в определённом смысле) групп. В зависимости от природы априорной информации и конкретного вида функционала, задающего критерий качества классификации, приходят к тем или иным схемам дискриминантного анализа, кластерного анализа, анализа смесей распределений.

Проблема снижения размерности

исследуемого факторного пространства и отбора наиболее информативных показателей заключается в определении такого набора показателей $Z=(z_1, z_2, \dots, z_m)$ из класса допустимых преобразований $Z(x)$ исходных показателей $x=(x_1, x_2, \dots, x_p)$, где m значительно меньше p , на котором достигается максимум некоторой заданной меры информативности m -мерной системы признаков. Конкретизация функционала, задающего эту меру информативности, приводит, в частности, к разл. схемам факторного анализа или методам отбора наиболее информативных показателей в схемах статистич. исследования зависимостей и дискриминантного анализа.

Осн. математич. методы М. с. а. включают методы теории вероятностей, теории систем линейных уравнений и теории матриц, а также некоторые оптимизационные алгоритмы. М. с. а. применяется в разл. статистич. и эконометрич. исследованиях.

Литература

Лит.: Андерсон Т. В. Введение в многомерный статистический анализ. М., 1963; Кендалл М. Дж., Стьюарт Ф. Многомерный статистический анализ и временные ряды. М., 1976; Прикладная статистика: классификация и снижение размерности. М., 1989; Айвазян С. А. Методы эконометрики. М., 2010.

Processing math: 0%