

6. ЛЕКЦИЯ. Метод главных компонент

6.1. Сущность проблемы снижения размерности и различные методы ее решения

В исследовательской и практической статистической работе приходится сталкиваться с ситуациями, когда общее число p признаков $x^{(1)}, x^{(2)}, \dots, x^{(n)}$, регистрируемых на каждом из множества обследуемых объектов, очень велико – порядка ста и более. Тем не менее имеющиеся многомерные наблюдения следует подвергнуть статистической обработке, осмыслить либо ввести в базу данных для того, чтобы иметь возможность их использовать в нужный момент.

$$X_i = \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \dots \\ x_i^{(p)} \end{pmatrix}, i = 1, 2, \dots, n \quad (1)$$

Желание представить каждое из наблюдений (1) в виде вектора Z некоторых вспомогательных показателей $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ с существенно меньшим (чем p) числом компонент p' бывает обусловлено в первую очередь следующими причинами:

- необходимостью *наглядного представления* (визуализации) исходных данных (1), что достигается их проецированием на специально подобранное трехмерное пространство ($p' = 3$), плоскость ($p' = 2$) или числовую прямую;
- стремлением к *лаконизму исследуемых моделей*, обусловленному необходимостью упрощения счета и интерпретации полученных статистических выводов;
- необходимостью *существенного сжатия объемов хранимой информации* (без видимых потерь в ее информативности), если речь идет о записи и хранении массивов типа (1) в специальной базе данных.

При этом новые (вспомогательные) признаки $z^{(1)}, z^{(2)}, \dots, z^{(p')}$, могут выбираться из числа исходных или определяться по какому-либо правилу по совокупности исходных признаков, например как их линейные комбинации. При формировании новой системы признаков к последним предъявляются разного рода требования, такие, как наибольшая информативность (в определенном смысле), взаимная некоррелированность и т. п. В зависимости от варианта формальной конкретизации этих требований приходим к тому или иному алгоритму снижения размерности. Имеется, по крайней мере, три основных типа принципиальных предпосылок, обуславливающих возможность

перехода от большого числа p исходных показателей состояния (поведения, эффективности функционирования) анализируемой системы к существенно меньшему числу p' наиболее информативных переменных. Это, во-первых, дублирование информации, доставляемой сильно взаимосвязанными признаками; во-вторых, не информативность признаков, мало меняющихся при переходе от одного объекта к другому (малая «вариабельность» признаков); в-третьих, возможность агрегирования, т. е. простого или «взвешенного» суммирования, по некоторым признакам.

Формально задача перехода (с наименьшими потерями в информативности) к новому набору признаков $\widetilde{z}^{(1)}, \widetilde{z}^{(2)}, \dots, \widetilde{z}^{(p')}$ может быть описана следующим образом. Пусть $Z = Z(X)$ – некоторая p -мерная вектор-функция исходных переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ ($p' \ll p$) и пусть $I_{p'}(Z(X))$ – определенным образом заданная мера информативности p' -мерной системы признаков $Z(X) = (z^{(1)}(X), \dots, z^{(p')}(X))$. Конкретный выбор функционала $I_{p'}(Z)$ зависит от специфики решаемой реальной задачи и опирается на один из возможных критериев: критерий автоинформативности, нацеленный на максимальное сохранение информации, содержащейся в исходном массиве $\{X_i\}_{i=1, \dots, n}$ относительно самих исходных признаков; и критерий внешней информативности, нацеленный на максимальное «выжимание» из $\{X_i\}_{i=1, \dots, n}$ информации, содержащейся в этом массиве относительно некоторых других (внешних) показателей.

Задача заключается в определении такого набора признаков \widetilde{Z} , найденного в классе F допустимых преобразований исходных показателей $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, что

$$I_{p'}(\widetilde{Z}(X)) = \max_{Z \in F} \{I_{p'}(Z(X))\} \quad (2)$$

Тот или иной вариант конкретизации этой постановки (определяющий конкретный выбор меры информативности $I_{p'}(Z)$ и класса допустимых преобразований) приводит к конкретному методу снижения размерности: к методу главных компонент, факторному анализу, экстремальной группировке параметров и т. д.

Поясним это на примерах.

1. Метод главных компонент. Именно к p' первым главным компонентам придет исследователь, если в качестве класса допустимых преобразований F определит всевозможные линейные ортогональные нормированные комбинации исходных показателей, т. е.

$$\begin{aligned}
z^{(j)}(X) &= c_{j1}(x^{(1)} - \mu^{(1)}) + \dots + c_{jp}(x^{(p)} - \mu^{(p)}); \\
\sum_{v=1}^p c_{jv}^2 &= 1, j = 1, 2, \dots, p; \\
\sum_{v=1}^p c_{jv}c_{kv} &= 0, j, k = 1, 2, \dots, p; j \neq k
\end{aligned} \tag{3}$$

здесь $\mu^{(v)} = Ex^{(v)}$ – математическое ожидание $x^{(v)}$ а в качестве меры информативности p' -мерной системы показателей $(z^{(1)}(X), \dots, z^{(p')}(X))$ выражение

$$I_{p'}(Z(X)) = \frac{D_{z^{(1)}} + \dots + D_{z^{(p')}}}{D_{x^{(1)}} + \dots + D_{x^{(p)}}} \tag{4}$$

здесь D знак операции вычисления дисперсии соответствующей случайной величины.

2. *Факторный анализ.* Как известно, модель факторного анализа объясняет структуру связей между исходными показателями $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ тем, что поведение каждого из них статистически зависит от одного и того же набора так называемых общих факторов $y^{(1)}, \dots, y^{(p)}$, т. е.

$$x^{(j)} - \mu^{(j)} = \sum_{v=1}^{p'} g_{jv} y^{(v)} + u^{(j)}, j = 1, 2, \dots, p$$

где g_{jv} – «нагрузка» общего фактора $y^{(v)}$ на исходный показатель $x^{(j)}$ а $u^{(j)}$ – остаточная «специфическая» случайная компонента, причем $Ey^{(v)} = 0$, $Eu^{(j)} = 0$, $Dy^{(v)} = 1$ и $y^{(1)}, \dots, y^{(p)}$, $u^{(1)}, \dots, u^{(p)}$ – попарно некоррелированы.

Оказывается, если F определить как класс всевозможных линейных комбинаций $x^{(1)}, \dots, x^{(p)}$ с учетом упомянутых ограничений на $y^{(v)}$, а в качестве меры информативности p -мерной системы показателей выбрать величину $I_{p'}(Z(X)) = 1 - \|R_X - R_{\hat{X}}\|^2$, то решение оптимизационной задачи (2) совпадает с вектором общих факторов $(y^{(1)}, \dots, y^{(p)})$ в модели факторного анализа. Здесь R_X – корреляционная матрица исходных показателей $x^{(1)}, \dots, x^{(p)}$, $R_{\hat{X}}$ – корреляционная матрица показателей $\widehat{x^{(j)}} = \sum_{v=1}^{p'} g_{jv} y^{(v)}$, а $\|A\|$ — евклидова норма матрицы A .

3. *Метод экстремальной группировки признаков.* В данном методе речь идет о таком разбиении совокупности исходных показателей $x^{(1)}, \dots, x^{(p)}$ на заданное число p' групп $S_1, \dots, S_{p'}$, что признаки, принадлежащие одной группе, были бы взаимокоррелированы сравнительно сильно, в то время как признаки, принадлежащие к разным группам, были бы коррелированы слабо. Одновременно решается задача замены каждой (i -й) группы сильно взаимокоррелированных исходных показателей одним вспомогательным «равнодействующим» показателем $z^{(i)}$, который, естественно, должен быть в тесной корреляционной связи с признаками своей группы. Определив в

качестве класса допустимых преобразований F исходных показателей все нормированные ($D_z^i = 1$) линейные комбинации $x^{(1)}, \dots, x^{(p)}$, ищем решение $(S_1^*, \dots, S_{p'}^*; \widetilde{z^{(1)}}, \dots, \widetilde{z^{(p')}})$, максимизируя (по S и $Z(X)$) функционал.

$$I_{p'}(Z(X); S) = \sum_{x^{(k)} \in S_1} r^2(x^{(k)}, z^{(1)}) + \dots + \sum_{x^{(k)} \in S_{p'}} r^2(x^{(k)}, z^{(p')})$$

где $r(x, z)$ — коэффициент корреляции между переменными x и z .

4. *Многомерное шкалирование.* В ряде ситуаций и в первую очередь в ситуациях, когда исходные статистические данные получают с помощью специальных опросов, анкет, экспертных оценок, возможны случаи, когда элементом первичного наблюдения является не состояние i -го объекта, описываемого вектором X_i , а характеристика p_{ij} попарной близости (отдаленности) двух объектов (или признаков) соответственно с номерами i и j . В этом случае исследователь располагает в качестве массива исходных статистических данных матрицей размера $n \times n$ (если рассматриваются характеристики попарной близости объектов) или $p \times p$ (если рассматриваются характеристики попарной близости признаков) вида

$$\rho = (\rho_{ij}), \quad i, j = 1, 2, \dots, t, \quad t = n \text{ или } t = p \quad (5)$$

где величины ρ_{ij} интерпретируются либо как расстояния между объектами (признаками) i и j , либо как ранги, задающие упорядочение этих расстояний. Задача многомерного шкалирования состоит в том, чтобы «погрузить» наши объекты (признаки) в такое p' -мерное пространство ($p' \ll \min(p, n)$), т. е. так выбрать координатные оси $0z^{(1)}, \dots, 0z^{(p')}$, чтобы исходная геометрическая конфигурация совокупности анализируемых точек-объектов (или точек-признаков), заданных с помощью уравнений 1 или 5, оказалась бы наименее искаженной в смысле некоторого критерия средней «степени искажения» $\Delta(Z)$ взаимных попарных расстояний.

Одна из достаточно общих схем многомерного шкалирования определяется критерием

$$\Delta(Z) = \sum_{i,j=1}^n d_{ij}^\alpha |\widehat{d_{ij}}(Z) - d_{ij}|^\beta$$

где d_{ij} — расстояние между объектами O_i и O_j в исходном пространстве, $\widehat{d_{ij}}(Z)$ — расстояние между теми же объектами в искомом пространстве меньшей размерности p' , а α и β — свободные параметры, выбор конкретных значений которых производится по усмотрению исследователя.

Определив меру информативности искомого набора признаков Z , например, как величину, обратную упомянутой выше величине степени искажения геометрической структуры исходной совокупности точек, сведем эту задачу к общей постановке (2), полагая

$$I_{p'}(Z) = \left[1 + \sum_{i,j=1}^n d_{ij}^\alpha |\widehat{d}_{ij}(Z) - d_{ij}|^\beta \right]^{-1}$$

5. *Отбор наиболее информативных показателей в моделях дискриминантного анализа.* Приведенные выше функционалы являются измерителями автоинформативности соответствующей системы признаков. Приведем теперь примеры критериев внешней информативности.

В частности, нас будет интересовать информативность системы показателей $z^{(1)}(X), \dots, z^{(p')}(X)$ с точки зрения правильности классификации объектов по этим показателям в схеме дискриминантного анализа (решения задач распознавания образов, который используется для принятия решения о том, какие переменные разделяют (т.е. «дискриминируют») возникающие наборы данных (так называемые «группы»)). При этом класс допустимых преобразований F определим исходя из требований, что в качестве $z^{(k)}(X)$ могут рассматриваться лишь представители набора исходных показателей, т. е. $Z(X) = (x^{(i_1)}, x^{(i_2)}, \dots, x^{(i_{p'})})$. Распространенным исходным тезисом при решении задачи выявления наиболее информативных p' показателей из исходного набора $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ является утверждение, что вектор показателей $(x^{(i_1)}, x^{(i_2)}, \dots, x^{(i_{p'})})$ заданной размерности p' тем более информативен, чем больше различие в законах его вероятностного распределения, определенных в разных классах в рассматриваемой задаче классификации. Если ввести меру попарного различия $\delta\{P_i(Z), P_j(Z)\}$ законов $P_1(Z), P_k(Z)$, описывающих распределение вероятностей вектора признаков $Z = (x^{(i_1)}, \dots, x^{(i_{p'})})$ в классах с номерами $1, 2, \dots, k$, то можно формализовать вышеприведенный принцип отбора наиболее информативных показателей $x^{(i_1)}, \dots, x^{(i_{p'})}$ определяя их из условия максимизации (по $i_1, i_2, \dots, i_{p'}$) величины

$$I_{p'}(Z(X)) = \sum_{i,j=1}^k \delta\{P_i(Z), P_j(Z)\}$$

Наиболее употребительные меры различия между законами распределения вероятностей $\delta\{P_i, P_j\}$ – это расстояние информационного типа (расстояние Кульбака, расстояние Махаланобиса), а также «расстояние по вариации».

6. *Отбор наиболее информативных переменных в моделях регрессии.* При построении зависимостей регрессионного типа одним из центральных оказывается вопрос выявления сравнительно небольшого числа p' переменных (из априорного набора $x^{(1)}, x^{(2)}, \dots, x^{(p)}$), наиболее существенно влияющих на поведение исследуемого результирующего признака y .

Таким образом, как и в предыдущем пункте, класс F состоит из всевозможных наборов переменных $Z = (x^{(i_1)}, \dots, x^{(i_{p'})})$, отобранных из исходного множества факторов-аргументов $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, и имеем дело с критерием внешней информативности таких наборов. Его вид обычно задается с помощью множественного коэффициента детерминации $R_y^2(x^{(i_1)}, \dots, x^{(i_{p'})})$ – характеристики степени тесноты связи показателя y с набором переменных $x^{(i_1)}, \dots, x^{(i_{p'})}$. При этом для фиксированной размерности $p' < p$ набор переменных $x^{(i_1^*)}, \dots, x^{(i_{p'}^*)}$ будет, очевидно, считаться наиболее информативным (с точки зрения точности описания поведения показателя y), если значение меры информативности на этом наборе достигает максимума.

7. *Сведение нескольких частных критериальных показателей к единому интегральному.* Речь идет о ситуациях, в которых «качество функционирования» исследуемой системы или объекта (предприятия, сложного изделия, отдельного специалиста и т.д.) характеризуется набором поддающихся измерению частных критериальных показателей $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Однако требуется перейти к некоторой не поддающейся непосредственному измерению скалярной интегральной оценке y . При этом постулируется, что латентный показатель y является функцией известного общего вида от $x^{(1)}, \dots, x^{(p)}$, т. е. $y = f(x^{(1)}, \dots, x^{(p)}; \Theta)$, и требуется подобрать лишь неизвестное значение параметра (вообще говоря, векторного) Θ .

Для решения этой задачи к зарегистрированной в результате контрольного обследования исходной статистической информации вида (1) приходится добавлять один из следующих вариантов экспертной информации о показателе y .

В а р и а н т 1: *балльная оценка* «выходного качества» y , т. е. значения $y_{1\Theta}, y_{2\Theta}, \dots, y_{n\Theta}$, экспертно оценивающие в определенной балльной шкале «выходное качество» 1-го, 2-го, ..., n -го объектов.

В а р и а н т 2: *ранжирование* анализируемых объектов, т. е. их упорядочение по степени убывания «выходного качества» y ; таким образом

будем иметь ранги $R_\Theta = \{R_{i\Theta}\}_{i=\overline{1,n}}$, т. е. порядковые номера объектов в этом упорядоченном ряду.

В а р и а н т 3: результаты *парных сравнений* анализируемых объектов по интересующему нас «выходному качеству» или результат разбиения контрольной совокупности объектов на группы, однородные с точки зрения «выходного качества»; и в том и в другом случае экспертные данные могут быть представлены с помощью булевой матрицы $\Gamma = (\gamma_{i,j,\Theta})_{i,j=1}$, где $\gamma_{i,j,\Theta} = 1$ если O_i не хуже O_j , $\gamma_{ij} = 0$ в противном случае.

Алгоритмы определения неизвестного параметра Θ используют в качестве исходной статистическую информацию (1), дополненную одним из вариантов экспертной информации (поэтому метод называется экспертно-статистическим), и построены на следующей идее. Если было бы известно значение параметра $\hat{\Theta}$, можно было бы вычислить значение целевой функции $f(x^{(1)}, \dots, x^{(p)}; \hat{\Theta})$ для каждого из контрольных объектов и определить с помощью этой целевой функции и балльные оценки $f(x^{(1)}, \dots, x^{(p)}; \hat{\Theta})$, и ранги $R(\hat{\Theta}) = \{R_i(\hat{\Theta})\}_{i=\overline{1,n}}$, и матрицу парных сравнений $\Gamma(\hat{\Theta}) = (\gamma_{ij}(\hat{\Theta}))$, $i, j = \overline{1,n}$

Поэтому если хотим формализовать с помощью целевой функции $f(X; \Theta)$ экспертные критерийные установки, в соответствии с которыми формируется единый интегральный показатель «выходного качества» y , естественно подчинить алгоритм поиска параметра Θ оптимизационному критерию вида

$$I_1(Z(X, \Theta)) = \begin{cases} \left[1 + \sum_{i=1}^n (y_{i\Theta} - f(x^{(1)}, \dots, x^{(p)}; \Theta))^2 \right]^{-1} & \text{в варианте 1;} \\ r(R_\Theta, R(\Theta)) & \text{в варианте 2;} \\ \left[1 + \sum_{i=1}^n |\gamma_{i,j,\Theta} - \widehat{\gamma}_{i,j}(\Theta)| \right]^{-1} & \text{в варианте 3;} \end{cases}$$

(здесь под $r(S, Q)$ подразумевается коэффициент ранговой корреляции Спирмэна между ранжировками S и Q). Разработаны алгоритмы и программы, позволяющие вычислять в в задаче максимизации критерия $I_1(Z(X, \Theta))$ для всех трех вариантов.

6.2. Геометрическая интерпретация метода главных компонент

Достаточно часто при проведении исследований число принимаемых во внимание признаков слишком велико. В результате в распоряжении исследователя оказывается матрица очень большой размерности. Во многих случаях кажется естественным предположить, что данные могут быть объединены некоторым небольшим числом новых переменных, которые

непосредственно не измеряются. Такое предположение позволяет понизить размерность пространства наблюдений, и иногда эти новые переменные способствуют выдвижению новых, поддающихся содержательной интерпретации и осмыслению гипотез.

Метод главных компонент (Principal Components Analysis) разработан К. Пирсоном (англ. Karl Pearson) в 1901 г. и затем детально разработан американским экономистом и статистиком Г. Хоттелингом в 1933 г.

В процессе создания метода главных факторов сформировались три различных подхода к выделению главных осей. Пирсон исходил из того, что при отыскании плоскостей или гиперплоскостей, проходящих через центр тяжести облака точек в n -мерном пространстве, обеспечивалась минимальная сумма квадратов расстояний всех точек от этих плоскостей. При втором подходе, использованном Андерсоном, исходят из геометрических представлений. И наконец, третий подход, самый распространенный, исходит из максимизации дисперсии в одном направлении при введении дополнительных условий.

Геометрическая интерпретация

Пусть определены три нормально распределенных параметра у n индивидуумов. Тогда возникает ситуация, изображенная на рис. 1, где n точек сосредоточены в трехмерном пространстве с тремя осями переменными X, Y, Z в облаке вокруг общего центра тяжести.

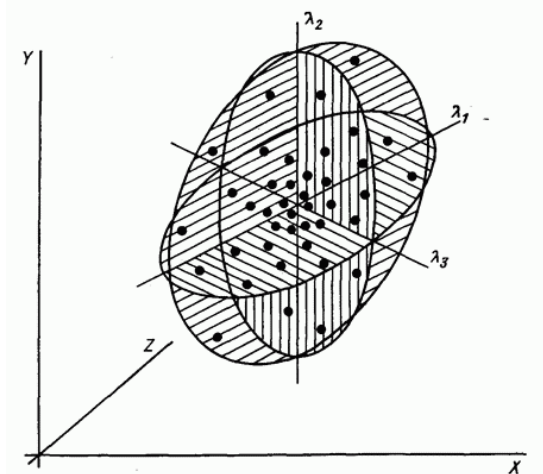


Рис. 1. Трехмерное распределение точек с соответствующими главными осями. X, Y, Z — первоначальная система координат; $\lambda_1, \lambda_2, \lambda_3$ — система главных осей.

Это облако точек наблюдений в общем случае имеет овальную форму и называется эллипсоидом. В частном случае, когда во всех трех направлениях дисперсия одинакова по величине, получают шар. На рис. 1 изображено

овальное тело с тремя секущими плоскостями (различно заштрихованными), проходящими через центр тяжести. Оси координат исходных, переменных X, Y, Z являются более или менее произвольными. Систему координат X, Y, Z можно было бы сместить вдоль одной или нескольких осей, не изменяя облако точек как таковое. Очевидно, систему координат можно вращать как угодно вокруг начала координат в любом из трех измерений и в известной степени удерживать облако точек в неизменном состоянии. Тогда имеется бесконечно много систем координат, в которых можно изобразить наблюдаемые точки.

Но одна из них представляет особый интерес. Это система координат главных осей. Самый длинный диаметр овального тела является первой главной осью λ_1 . Второй главной осью λ_2 является самый длинный диаметр в плоскости, ортогональной к первой главной оси и проходящей через центр тяжести системы (заштрихована вертикально). Третья главная ось λ_3 в трехмерном случае перпендикулярна к первой и второй главным осям и проходит через центр тяжести.

В геометрическом плане метод главных компонент состоит в том, что вначале определяют самую длинную ось эллипсоида. Она является первой главной осью, которая должна пройти через центр тяжести, эта ось обозначена λ_1 . Далее устанавливают подпространство (в данном случае плоскость), которое перпендикулярно к первой главной оси и которое проходит через центр тяжести (заштриховано вертикально). В этом подпространстве находится следующая по величине ось скопления точек и т. д., пока не будут определены последовательно все главные оси.

Длины главных осей пропорциональны величинам дисперсий в направлении соответствующей главной оси. С помощью метода главных компонент устанавливаются направления этих осей относительно первоначальной системы координат. Главные оси соответствуют факторам, которые должны быть лишь надлежащим образом пронормированы, чтобы выполнялось требование единичной дисперсии факторов.

На рис. 2 наглядно изображена ситуация после выделения первой главной оси. Все точки рис. 1 спроецированы на плоскость, проходящую через центр тяжести, перпендикулярно к первой главной оси, причем проецирование производится параллельно λ_1 , как это показано для одной точки. По координатам точек на этой плоскости ищут вторую главную ось. Направление оси λ_2 устанавливается опять таким образом, чтобы максимум дисперсии лежал в направлении этой оси, т. е. определяется самая длинная ось эллипса на этой

плоскости. Третья главная ось проходит через центр тяжести и перпендикулярна к λ_1 и λ_2 .

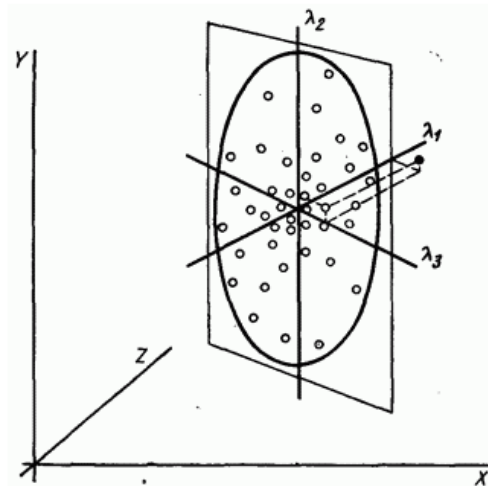


Рис. 2. Ситуация после установления положения первой главной оси.

Определение второй и третьей главных осей. Кружочками обозначены проекции точек на плоскость. Показано проецирование одной из таких точек

Все n точек, изображенные на рис. 1, можно спроецировать параллельно плоскости λ_2, λ_3 на первую главную ось, тогда распределение точек на этой прямой укажет максимум дисперсии в одном измерении. Точно таким же образом можно спроецировать точки на любую другую главную ось или на так называемые гиперплоскости, или подпространства, которые натянуты на них. Плоскость, натянутая на первые две главные оси λ_1 и λ_2 на рис. 1 расположена так, что проекция всех точек на нее дает максимум дисперсии в двух измерениях. Изображение этой плоскости представлено еще раз на рис. 3. Так как полная дисперсия известна, то можно поэтапно установить, какая ее доля (в процентах) содержится в гиперплоскости r главных осей.

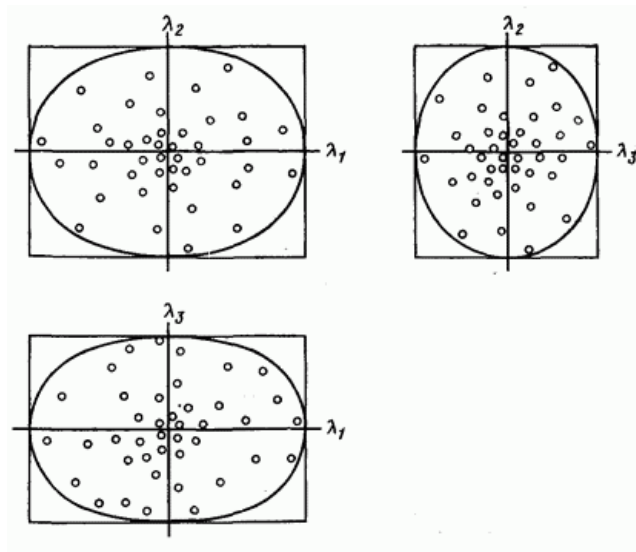


Рис. 3. Проекции точек, изображенных на рис. 1, на трёх плоскостях, проходящих через главные оси $\lambda_1, \lambda_2, \lambda_3$

На рис. 4 демонстрируется случай, когда для описания всех точек достаточно двух главных осей. Хотя точки соответствуют результатам наблюдений по трем переменным, эти три переменные X, Y и Z почти полностью зависят от двух величин. Все точки находятся на одной плоскости или вблизи нее, причем эта плоскость наклонена под некоторым углом к первоначальной системе координат. Если вместо измеренных трех переменных выбрать для изображения точек только две, а именно использовать две главные оси λ_1 и λ_2 , то при небольшом рассеянии точек вне плоскости, натянутой на эти оси, теряется незначительная часть информации.

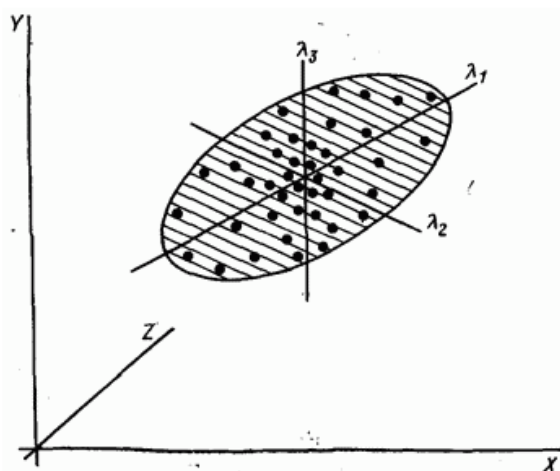


Рис. 4. Распределение результатов наблюдений с небольшим рассеянием вдоль оси λ_3 . Все точки, которые были определены в системе координат XYZ , лежат практически на плоскости λ_1 и λ_2 имеют лишь незначительное рассеяние в направлении λ_3

Изображенная в пространстве на рис. 4 плоскость, определенная главными осями, на рис. 5 представлена в плане. Дисперсия в направлении λ_3 мала. Если бы все точки лежали на одной плоскости, то потребность в третьей главной оси λ_3 отпала бы совсем, и без потери какой-либо информации можно было бы точки изображать не в системе координат XYZ , а в системе λ_1 и λ_2 .

Переход от системы координат XYZ к системе на рис. 1 – 5 соответствует геометрическому решению задачи выделения факторов, осуществляемому с помощью метода главных факторов. Этот переход от одной системы координат к другой без потери информации на практике большей частью возможен лишь тогда, когда определяются все главные компоненты, т. е. для m переменных определяют главных компонент. При благоприятных обстоятельствах, а на практике это встречается довольно часто, довольствуются меньшим числом

главных осей, но достаточным, чтобы воспроизвести большую часть дисперсии, как это схематично изображено на рис. 4 и 5.

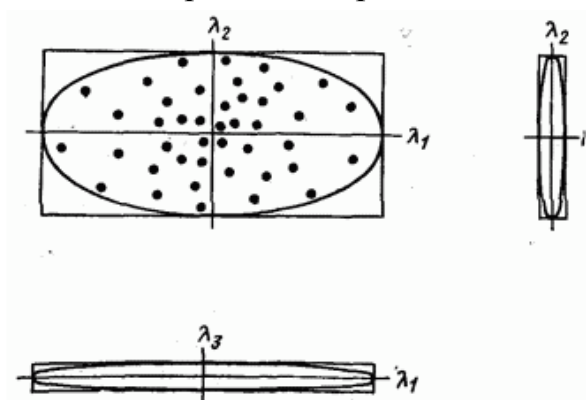


Рис. 5. Три проекции овального тела, изображенного на рис. 4/

При применении метода главных факторов исходят большей частью не из тестового пространства, к которому здесь прибегают ради наглядности, а из корреляционной матрицы. Как будет далее показано, каждой главной оси соответствует собственный вектор α и собственное значение λ корреляционной матрицы.

Собственное значение λ имеет порядок величины дисперсии, корень из него соответствует, поэтому длинам главных осей на рис. 1 – 5. Чтобы добиться наглядности на этих рисунках, вопрос о знаке корня не затрагивался. Для обозначения главных осей использовался лишь символ λ .