



DOI: 10.5862/JPM.218.13

УДК 519.226.2-519.248

Ю.А. Пичугин

Российский государственный педагогический университет им. А.И. Герцена

ГЕОМЕТРИЧЕСКИЕ АСПЕКТЫ ПРОВЕРКИ СЛОЖНЫХ СТАТИСТИЧЕСКИХ ГИПОТЕЗ В МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ

Рассмотрены геометрические аспекты задачи проверки сложной статистической гипотезы о принадлежности вектора параметров модели некоторой области. Сформулирована и доказана основополагающая теорема для решения этой задачи. Теорема утверждает, что задача решается проверкой простой статистической гипотезы относительно точки максимального правдоподобия. Рассмотрены типовые примеры применения в математическом моделировании.

ГРАНИЧНАЯ ТОЧКА МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ, ЗАДАЧА ЛАГРАНЖА, ЭЛЛИпсоИД НАДЕЖНОСТИ, ОБОБЩЕННАЯ ЕВКЛИДОВА МЕТРИКА.

Введение

Процесс построения каких-либо математических моделей реальных процессов обычно содержит два неотъемлемых этапа: спецификация и идентификация. При учете факторов стохастической природы, которые естественно возникают на этапе идентификации, состоящем, главным образом, в оценке параметров, возникает необходимость проверки статистических гипотез. Обычно эти гипотезы относятся к значениям параметров. Например, в хорошо известных моделях эконометрики проверяются, в основном, гипотезы о равенстве нулю некоторых параметров модели. Это позволяет определять целесообразность как включения анализируемых переменных в модель, так и построения самой модели. Проверка таких статистических гипотез осуществляется посредством хорошо известного аппарата математической статистики. Однако на практике исследователь нередко сталкивается с ситуацией, когда интересен вопрос о принадлежности набора параметров, рассматриваемого как точка пространства или вектор, некоторой (наперед заданной) области. В этом случае возникает задача проверки сложной многомерной статистической гипотезы.

В работе [1], опубликованной в настоящем журнале, решалась задача проверки такой сложной гипотезы, в частности, речь шла о принадлежности вектора пара-

метров модели Вольтерры (системы двух дифференциальных уравнений) некоторой области, что позволяло делать выводы об устойчивости решения. Представленное в упомянутой публикации и в ряде других работ (например, в [2]) решение содержало в качестве обоснования лишь лаконичную ссылку на геометрические свойства нормального распределения.

С учетом важности этой проблемы для математического моделирования в целом, настоящая работа ставит целью полностью обосновать предлагаемый метод проверки сложной статистической гипотезы и сделать все необходимые уточнения.

При этом делается акцент на геометрические аспекты развития общей идеи и обоснования этого метода, а также на указанные аспекты решения различных прикладных задач математического моделирования, с тем, чтобы показать диапазон применимости рассматриваемого математического аппарата. По убеждению автора, такое представление метода должно улучшить понимание его сути и прикладного значения.

Далее везде в статье предполагается, что исходные данные подчиняются многомерному нормальному распределению.

Граничная точка максимального правдоподобия и эллипсоид надежности

Хорошо известно, что когда одномер-

ный параметр θ оценивается статистически, например, как среднее значение по некоторой выборке, и проверяется $H_0 : \theta = \theta_0$, то доверительная область есть симметричный относительно θ_0 интервал, а критическая область есть внешность или дополнение этого интервала. Ширина доверительной области определяется выбором уровня значимости α (вероятность попадания в критическую область). При этом обычно используется распределение Стьюдента, симметрия функции плотности которого и приводит к симметрии доверительной области.

Однако при достаточно большом объеме выборки можно использовать нормальное стандартное распределение $N(0, 1)$. Именно так и поступали в эпоху статистических таблиц. Обобщение этой простой идеи на многомерный случай оказывается вполне конструктивным. Действительно, предположим, что объем выборки настолько велик, что можно использовать нормальное распределение и в многомерном случае. Пусть вектор параметров

$$\theta = (\theta_1, \theta_2, \dots, \theta_k)^T$$

(T – знак транспонирования) размерности k оценен по выборке $\{\theta_i, i = 1, 2, \dots, n\}$.

Как известно, несмещенные оценки математического ожидания θ и матрицы взаимных ковариаций компонент этой оценки θ , соответственно, равны

$$\hat{\theta} = n^{-1} \sum_{i=1}^n \theta_i, \quad \hat{V}_0 = n^{-1} \hat{V}_0, \quad (1)$$

где

$$\hat{V}_0 = (n-1)^{-1} \sum_{i=1}^n (\theta_i - \hat{\theta})(\theta_i - \hat{\theta})^T.$$

При проверке простой гипотезы $H_0 : \theta = \theta_0$ доверительная область ω_0^α есть часть пространства, которая имеет вероятностную меру $(1 - \alpha)$ и ограничена эллипсоидом

$$W^\alpha = \{\theta : f(\theta, \theta_0, V_0) = c_\alpha\},$$

т. е. $P(\text{Int } W^\alpha) = (1 - \alpha)$.

Этот эллипсоид W^α хорошо известен в литературе как эллипсоид рассеяния, а

$$f(\theta, \theta_0, V_0) = (2\pi)^{-k/2} [\det V_0]^{-1/2} \times \exp(-0,5(\theta - \theta_0)^T V_0^{-1} (\theta - \theta_0))$$

– плотность многомерного нормального распределения $N(\theta_0, V_0)$ [3].

Здесь и далее предполагается, что матрица V_0 – невырожденная (определение функции плотности в вырожденном случае см. в монографии [3]).

Предположим, что мы хотим проверить сложную гипотезу вида $H_G : \theta \in G$, где G – заданная область пространства параметров, которую, не уменьшая общности, будем считать замкнутой, так как при непрерывном распределении граница ∂G есть множество вероятностной меры нуль.

Будем рассматривать H_G как объединение простых гипотез вида $H_\tau : \theta = \theta_\tau$, где $\theta_\tau \in G$. Тогда доверительная область гипотезы H_G есть объединение $\bigcup_{\theta_\tau \in G} \omega_\tau^\alpha$ доверительных областей гипотез H_τ , т. е.

$$\omega_\tau^\alpha = \{\theta : f(\theta, \theta_\tau, V_0) < c_\alpha\},$$

а критическая область H_G есть пересечение $\bigcap_{\theta_\tau \in G} c\omega_\tau^\alpha$ дополнений $c\omega_\tau^\alpha$ к ω_τ^α .

Границей раздела доверительной и критической областей будет огибающая (кривая, поверхность или гиперповерхность, соответственно при $k = 2, 3$ и более) семейства эллипсоидов

$$W_\tau^\alpha = \{\theta : f(\theta, \theta_\tau, V_0) = c_\alpha, \theta_\tau \in \partial G\},$$

центры которых лежат на границе ∂G .

Проверку такой сложной гипотезы в первом приближении можно осуществить по следующему алгоритму, основанному на том, что в функции плотности распределения вероятностей аргумент и параметр сдвига занимают симметричное положение относительно друг друга.

Геометрическая интерпретация этого факта состоит в следующем. Пусть два эллипсоида W_1^α и W_2^α конгруэнтны и совпадают с точностью до параллельного переноса, т. е. определяются одним набором параметров $\{V_0, c_\alpha\}$. Тогда, если центр W_1^α лежит на W_2^α , то и центр W_2^α лежит на W_1^α . Рассмотрим множество эллипсоидов, центры которых совпадают с $\hat{\theta}$, а форма определена матрицей V_0 . Один из эллипсоидов этого множества W^α соприкасается с границей ∂G в некоторой точке θ_* . Значение c_γ для этого эллипсоида будет вычис-

ляться непосредственно как правдоподобие распределения $N(\hat{\theta}, \mathbf{V}_\theta)$ в найденной точке θ_* , т. е. $c_\gamma = f(\theta_*, \hat{\theta}, \mathbf{V}_\theta)$. Тогда, если $c_\gamma > c_\alpha$, то следовательно, $P(\text{Int } W^\gamma) < P(\text{Int } W^\alpha)$ и точка θ_* попадает в доверительную область одной из гипотез $H_\tau : \theta = \theta_\tau$. Это означает принятие общей гипотезы $H_G : \theta \in G$. При выполнении условия $c_\gamma < c_\alpha$, наоборот, точка θ_* не попадает в доверительную область ни одной из гипотез $H_\tau : \theta = \theta_\tau$ и общая гипотеза H_G отвергается.

В часто встречающейся ситуации, когда оценка $\hat{\theta} \notin G$ и мы заинтересованы отвергнуть гипотезу H_G , вероятностная мера пространства $P(\text{Int } W^\gamma)$, ограниченного эллипсоидом W^γ , по распределению $N(\hat{\theta}, \mathbf{V}_\theta)$, есть оценка меры надежности [4, 5], поэтому сам эллипсоид W^γ целесообразно именовать эллипсоидом надежности.

Решение задачи нахождения точки θ_* , как хорошо известно, дается методом Лагранжа. Геометрическая интерпретация и сама суть метода состоит в том, что в точке касания нормали к соприкасающимся поверхностям W^γ и ∂G коллинеарны. Решение сводится к дифференцированию по всем аргументам функции Лагранжа вида [1, 2]:

$$L(\theta, \lambda) = \ln(f(\theta, \hat{\theta}, \mathbf{V}_\theta) - \lambda\phi(\theta)). \quad (2)$$

Здесь предполагается, что граница ∂G задается уравнением $\phi(\theta) = 0$ для некоторой гладкой функции ϕ . Если граница ∂G является кусочногладкой, т. е. задается некоторым набором уравнений для системы гладких функций

$$\{\phi_j(\theta) = 0, j = 1, 2, \dots, J\},$$

то вместо одной задачи (2) мы будем иметь систему задач и функций Лагранжа:

$$L_j(\theta, \lambda) = \ln(f(\theta, \hat{\theta}, \mathbf{V}_\theta) - \lambda\phi_j(\theta)), \quad (3)$$

$$j = 1, 2, \dots, J.$$

Из всех решений системы (3) выбирается значение θ_* , которое доставляет максимум правдоподобия $f(\theta_*, \hat{\theta}, \mathbf{V}_\theta)$. Найденная таким образом точка θ_* , очевидно, является граничной точкой максимального правдоподобия и для распределения $N(\hat{\theta}, \mathbf{V}_\theta)$.

Проверка простой гипотезы относительно граничной точки максимального правдоподобия

Рассмотрим простую гипотезу $H_* : \theta = \theta_*$. Ее проверку можно осуществить посредством следующих статистик. Если ортогональная матрица \mathbf{Q} такова, что

$$\mathbf{Q}^T \mathbf{V}_\theta \mathbf{Q} = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k),$$

где $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$, то тогда компоненты вектора

$$\mathbf{s} = (s_1, s_2, \dots, s_k)^T = \Lambda^{-1/2} \mathbf{Q}^T (\hat{\theta} - \theta_*) \quad (4)$$

будут иметь в точности распределение Стьюдента с числом степеней свободы $(n - 1)$, т. е.

$$s_j \sim t_{n-1} \quad (j = 1, 2, \dots, k).$$

Величина

$$r = (\hat{\theta} - \theta_*)^T \mathbf{V}_\theta^{-1} (\hat{\theta} - \theta_*) \quad (5)$$

при верной гипотезе H_* имеет распределение, которое с увеличением объема выборки n асимптотически приближается к χ^2 -квадрат с k степенями свободы, т. е. $r \sim \chi_k^2$.

Предположим, что вектор параметров оценивается по схеме линейной регрессии

$$\mathbf{Y} = \mathbf{X}\theta + \varepsilon,$$

где \mathbf{Y} — вектор-столбец измерений размерности n , \mathbf{X} — $(n \times k)$ -матрица, ε — вектор-столбец погрешностей измерений размерности n .

В предположении независимости и нормального распределения погрешностей регрессии

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

где $\mathbf{0}$ — нулевой вектор-столбец, \mathbf{I} — единичная матрица.

Оценка методом наименьших квадратов (МНК) θ и матрица взаимных ковариаций этой оценки, соответственно, равны [3]:

$$\hat{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad \mathbf{V}_\theta = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (6)$$

Здесь несмещенная оценка σ^2 равна

$$\hat{\sigma}^2 = (n - k)^{-1} \sum_{i=1}^n (y_i - \mathbf{X}_i \hat{\theta})^2,$$

где y_i — i -я компонента, а \mathbf{X}_i — i -я строка, соответственно, вектора-столбца \mathbf{Y} и матрицы \mathbf{X} (см. [3]).

Тогда, при условии, что гипотеза H_* верна, величина

$$q = \frac{k^{-1}(\hat{\theta} - \theta_*)^T (\mathbf{X}^T \mathbf{X})(\hat{\theta} - \theta_*)}{(n - k)^{-1} \sum_{i=1}^n (y_i - \mathbf{X}_i \hat{\theta})^2} = \hat{\sigma}^{-2} k^{-1} (\hat{\theta} - \theta_*)^T (\mathbf{X}^T \mathbf{X})(\hat{\theta} - \theta_*) = k^{-1} (\hat{\theta} - \theta_*)^T \mathbf{V}_0^{-1} (\hat{\theta} - \theta_*) = rk^{-1} \quad (7)$$

будет иметь в точности F -распределение со степенями свободы $(k, n - k)$, т. е. $q \sim F_{k, n-k}$.

Для критерия Стьюдента (см. выше) в этом случае число степеней свободы равно $n - k$. Здесь следует напомнить, что среднее арифметическое (см. раздел «Граничная точка максимального правдоподобия и эллипсоид надежности») также является МНК-оценкой.

При выполнении хотя бы одного из следующих неравенств, для некоторого, заранее заданного, значения α , т. е.

$$r > \chi_k^2(\alpha); \quad (8)$$

$$q > F_{k, n-k}(\alpha); \quad (9)$$

$$|s_j| > t_{n-k}^{\alpha/2}, \quad j = 1, 2, \dots, k, \quad (10)$$

где $\chi_k^2(\alpha)$, $F_{k, n-k}(\alpha)$ и $t_{n-k}^{\alpha/2}$ — суть критические значения соответствующих распределений (здесь используется значение $\alpha/2$, так как критерий Стьюдента относится к двусторонним), гипотеза H_* отвергается.

Для полноты общей картины линейного МНК-оценивания параметров рассмотрим случай, когда многомерный параметр представляет собой не вектор, а $(m \times h)$ -матрицу Θ . При этом его оценка получается из системы линейных уравнений регрессии:

$$\mathbf{Y}_l = \mathbf{X} \theta_l + \varepsilon_l, \quad l = 1, 2, \dots, m \quad (11)$$

(как, например, в упомянутой выше работе [1], где эта матрица параметров обозначалась B , а сами параметры — β), т. е. общее число параметров $k = m \times h$ (m — число уравнений, h — число параметров в каждом уравнении).

В работе [1] рассматривался случай $m = h = 2$, $k = 4$, однако предложенная в данной работе методика легко обобщается на произвольные значения m и h . Единственное препятствие, состоящее в возмож-

ной вырожденности вычисляемых матриц (параметров многомерных распределений), преодолевается хорошо известными приемами. Как показано в статье [1], при оценке матрицы параметров

$$\Theta = (\theta_1, \theta_2, \dots, \theta_m),$$

где θ_l — вектор параметров l -го уравнения системы (11), $l = 1, 2, \dots, m$, следует переходить к центрированным переменным, когда в матрицу параметров не входят свободные члены, а входят только коэффициенты при переменных правой части (как правило, именно они и представляют главный интерес).

Действительно, пусть \mathbf{Y} имеет размерность $(n \times m)$ (матрица, составленная из левых частей системы (11), т. е.

$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m),$$

а $\mathbf{X} - (n \times h)$, и эти матрицы центрированы на средние по выборке значения переменных.

Тогда МНК-оценка матрицы параметров равна

$$\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (12)$$

Если вместо матрицы рассматривать составной вектор параметров $\theta = (\theta_1^T, \theta_2^T, \dots, \theta_m^T)^T$, то матрица взаимных ковариаций \mathbf{V}_θ компонент оценки этого вектора

$$\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T, \dots, \hat{\theta}_m^T)^T$$

строится следующим образом.

Пусть $\mathbf{R} - (m \times m)$ - и $\mathbf{W} - (h \times h)$ -ортогональные матрицы, приводящие к диагональному виду матрицы $\mathbf{Y}^T \mathbf{Y}$ и $\mathbf{X}^T \mathbf{X}$, соответственно.

Тогда, как показано в статье [1],

$$\mathbf{V}_\theta = \mathbf{R} \otimes \mathbf{I}_h \mathbf{V}_\xi (\mathbf{R} \otimes \mathbf{I}_h)^T,$$

где $\mathbf{R} \otimes \mathbf{I}_h$ — кронекеровское произведение \mathbf{R} (см. [6]) на единичную матрицу порядка h , а $\mathbf{V}_\xi - [(m \cdot h) \times (m \cdot h)]$ -блочно-диагональная матрица следующего вида:

$$\mathbf{V}_\xi = \begin{pmatrix} \hat{\delta}_1^2 (\mathbf{X}^T \mathbf{X})^{-1} & \mathbf{0} & \dots \\ \dots & \dots & \dots \\ \dots & \mathbf{0} & \hat{\delta}_m^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{pmatrix}.$$

Здесь, как обычно, $\mathbf{0}$ — нулевой блок,

а коэффициенты диагональных блоков $\{\hat{\delta}_1^2, \hat{\delta}_2^2, \dots, \hat{\delta}_m^2\}$ есть несмещенные оценки дисперсий погрешностей регрессии $\{\delta_1^2, \delta_2^2, \dots, \delta_m^2\}$ на центрированные значения \mathbf{X} также центрированных (см. ниже), но независимых переменных \mathbf{Z}_l ($l = 1, 2, \dots, m$), выборочная $(n \times m)$ -матрица для которых равна

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m) = \mathbf{Y}\mathbf{R},$$

так как $\mathbf{Z}^T \mathbf{Z} = \mathbf{R}^T \mathbf{Y}^T \mathbf{Y} \mathbf{R}$ – диагональная матрица (см. выше).

Из структуры \mathbf{V}_0 следует, что ортогональная матрица \mathbf{Q} , приводящая \mathbf{V}_0 к диагональному виду, равна

$$\mathbf{Q} = (\mathbf{R} \otimes \mathbf{I}_h)(\mathbf{I}_m \otimes \mathbf{W}).$$

При этом имеет место следующее, легко проверяемое, соотношение:

$$\begin{aligned} (\mathbf{R} \otimes \mathbf{I}_h)(\mathbf{I}_m \otimes \mathbf{W}) &= \mathbf{R} \otimes \mathbf{W} = \\ &= (\mathbf{I}_m \otimes \mathbf{W})(\mathbf{R} \otimes \mathbf{I}_h), \end{aligned}$$

означающее, что порядок перехода к независимым регрессорам и независимым компонентам отклика, когда \mathbf{V}_0 принимает диагональный вид, не имеет принципиального значения.

Имея МНК-оценку

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T, \dots, \hat{\boldsymbol{\theta}}_m^T)^T$$

и матрицу взаимных ковариаций \mathbf{V}_0 компонент этого вектора, находим граничную точку максимального правдоподобия

$$\boldsymbol{\theta}_* = (\boldsymbol{\theta}_{*1}^T, \boldsymbol{\theta}_{*2}^T, \dots, \boldsymbol{\theta}_{*m}^T)^T$$

для распределения $N(\hat{\boldsymbol{\theta}}, \mathbf{V}_0)$.

При этом для проверки H_* также получают χ^2 - и t -статистики вида (5) и (4) (т. е. величины r и s). В работе [1] вместо общей F -статистики вида (7) (т. е. q) было предложено использование m F -статистик для каждого l -го уравнения ($l = 1, 2, \dots, m$):

$$\begin{aligned} q_l &= \frac{(h)^{-1}(\hat{\boldsymbol{\theta}}_l - \hat{\boldsymbol{\theta}}_{*l})^T (\mathbf{X}^T \mathbf{X})(\hat{\boldsymbol{\theta}}_l - \hat{\boldsymbol{\theta}}_{*l})}{(n - h - 1)^{-1} \sum_{i=1}^n (y_{l,i} - \mathbf{X}_i \hat{\boldsymbol{\theta}}_l)^2} = \\ &= \hat{\sigma}_l^{-2} h^{-1} (\hat{\boldsymbol{\theta}}_l - \hat{\boldsymbol{\theta}}_{*l})^T (\mathbf{X}^T \mathbf{X})(\hat{\boldsymbol{\theta}}_l - \hat{\boldsymbol{\theta}}_{*l}), \end{aligned} \quad (13)$$

которые, при верной гипотезе H_* , имеют распределение $F_{h, n-h-1}$ (здесь число степеней свободы не $(n - h)$, а $(n - h - 1)$, так как учитывается центрирование, приводящее к «исчезновению» свободного члена).

Однако следует заметить, что только при определенном условии, а именно при верной и легко проверяемой гипотезе

$$H_0 : \delta_1^2 = \delta_2^2 = \dots = \delta_m^2,$$

которая эквивалентна гипотезе для дисперсий погрешностей системы (11):

$$H'_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2,$$

выполняется весьма важное соотношение:

$$r = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*)^T \mathbf{V}_0^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) = h \sum_{l=1}^m q_l. \quad (14)$$

В противном случае в качестве q_l следует брать статистики

$$\begin{aligned} q_l &= \frac{(h)^{-1}(\hat{\boldsymbol{\xi}}_l - \hat{\boldsymbol{\xi}}_{*l})^T (\mathbf{X}^T \mathbf{X})(\hat{\boldsymbol{\xi}}_l - \hat{\boldsymbol{\xi}}_{*l})}{(n - h - 1)^{-1} \sum_{i=1}^n (z_{l,i} - \mathbf{X}_i \hat{\boldsymbol{\xi}}_l)^2} = \\ &= \hat{\delta}_l^{-2} h^{-1} (\hat{\boldsymbol{\xi}}_l - \hat{\boldsymbol{\xi}}_{*l})^T (\mathbf{X}^T \mathbf{X})(\hat{\boldsymbol{\xi}}_l - \hat{\boldsymbol{\xi}}_{*l}), \end{aligned} \quad (15)$$

где

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= (\hat{\boldsymbol{\xi}}_1^T, \hat{\boldsymbol{\xi}}_2^T, \dots, \hat{\boldsymbol{\xi}}_m^T)^T = (\mathbf{R} \otimes \mathbf{I}_h)^T \hat{\boldsymbol{\theta}}, \\ \boldsymbol{\xi}_* &= (\boldsymbol{\xi}_{*1}^T, \boldsymbol{\xi}_{*2}^T, \dots, \boldsymbol{\xi}_{*m}^T)^T = (\mathbf{R} \otimes \mathbf{I}_h)^T \boldsymbol{\theta}_*. \end{aligned}$$

Тогда условие (14) будет выполнено, т. е.

$$\begin{aligned} r &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*)^T \mathbf{V}_0^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) = h \sum_{l=1}^m q_l = \\ &= (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_*)^T \mathbf{V}_{\hat{\boldsymbol{\xi}}}^{-1} (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_*), \end{aligned} \quad (16)$$

так как столбцы $\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m\}$ матрицы \mathbf{Z} не коррелируют между собой, а матрица взаимных ковариаций параметров регрессии для этих переменных $\mathbf{V}_{\hat{\boldsymbol{\xi}}}$ имеет блочно-диагональную структуру (см. выше).

При верной гипотезе H_0 выражения (13) и (15) совпадают. Условие (14) (или (16)) является весьма существенным уточнением результатов работы [1].

При выполнении хотя бы одного из неравенств

$$q_l > F_{k, n-h-1}(\alpha) \quad (17)$$

для каждого l -го уравнения ($l = 1, 2, \dots, m$) гипотеза H_* отвергается.

Обобщенная евклидова метрика и основная теорема

Теперь, когда выяснен алгоритм нахождения граничной точки максимального правдоподобия $\boldsymbol{\theta}_*$ и проверки простой гипотезы $H_* : \boldsymbol{\theta} = \boldsymbol{\theta}_*$, придадим задаче проверки сложной гипотезы $H_G : \boldsymbol{\theta} \in G$ совершенно

строгий доказательный вид, отличный от предложенного в разделе «Граничная точка максимального правдоподобия и эллипсоид надежности». При этом мы не будем делать никаких существенных предположений относительно объема выборки n . Для всех рассмотренных выше оценок имеет место следующая теорема.

Теорема. Пусть $\hat{\theta}$ (линейная МНК-оценка вектора параметров θ) не принадлежит замкнутой области G , и $V_{\hat{\theta}}$ – матрица взаимных ковариаций компонент этой оценки.

Тогда, если для граничной точки максимального правдоподобия θ_* ($\theta_* \in \partial G$) по распределению $N(\hat{\theta}, V_{\hat{\theta}})$ на некотором уровне значимости α гипотеза $H_*: \theta = \theta_*$ отвергается по какому-либо критерию (χ^2 , t или F), то и для любой другой точки $\theta_{\tau} \in G$ на этом же уровне значимости α отвергается гипотеза $H_{\tau}: \theta = \theta_{\tau}$, а, следовательно, отвергается гипотеза $H_G: \theta \in G$.

Доказательство. Введем в вещественном k -мерном пространстве R^k обобщенную евклидову метрику [7, 8]:

$$d_2(\mathbf{u}, \mathbf{v}) = \sqrt{(\mathbf{u} - \mathbf{v})^T V_{\hat{\theta}}^{-1} (\mathbf{u} - \mathbf{v})} \quad (18)$$

($\mathbf{u}, \mathbf{v} \in R^k$), порожденную распределением $N(\hat{\theta}, V_{\hat{\theta}})$.

Напомним, что метризовать R^k мы можем посредством любой невырожденной, положительно определенной, симметричной матрицы A . Если определить скалярное произведение векторов по формуле

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T A \mathbf{v},$$

то норма (длина) вектора, а также расстояние и угол между векторами определяются соответственно следующими формулами:

$$\|\mathbf{u}\|_2 = \sqrt{\mathbf{u}^T A \mathbf{u}};$$

$$d_2(\mathbf{u}, \mathbf{v}) = \sqrt{(\mathbf{u} - \mathbf{v})^T A (\mathbf{u} - \mathbf{v})};$$

$$\varphi = \arccos(\mathbf{u}^T A \mathbf{v} / (\|\mathbf{u}\|_2 \cdot \|\mathbf{v}\|_2)).$$

Если матрица A не является положительно определенной (среди собственных значений есть отрицательные), то введенная посредством A метрика (типа d_2) называется индефинитной [7]. Тот факт, что невырожденная матрица $V_{\hat{\theta}}$, как и обратная к ней, положительно определены, хорошо известен. Здесь как раз и рассматривается

невырожденный случай.

Очевидно, что граничная точка θ_* максимального правдоподобия является ближайшей из всех точек области G , включая границу, в метрике (18) к точке $\hat{\theta}$ (к полученной оценке вектора параметров по (1), (6), (12)). Следовательно, при выполнении неравенства (8) для θ_* , это неравенство будет выполняться и для любой другой точки $\theta_{\tau} \in G$. То же самое утверждение справедливо и для неравенства (9), использующего F -тест, так как в этом случае (оценка по одному уравнению регрессии) мы имеем дело с эквивалентной метрикой

$$d'_2(\mathbf{u}, \mathbf{v}) = k^{-1/2} d_2(\mathbf{u}, \mathbf{v})$$

(см. неравенство (9)).

Из равенства (4) непосредственно вытекает соотношение

$$d_2(\hat{\theta}, \theta_*) = \sqrt{(\hat{\theta} - \theta_*)^T V_{\hat{\theta}}^{-1} (\hat{\theta} - \theta_*)} = \sqrt{\sum_{j=1}^k s_j^2}, \quad (19)$$

которое в свою очередь означает, что использование студентовских статистик, с геометрической точки зрения, есть переход к обобщенной метрике

$$d_{\infty}(\hat{\theta}, \theta_*) = \max_{j \in \{1, 2, \dots, k\}} |s_j|.$$

Эта метрика совпадает с обычной метрикой d_{∞} (максимум модуля компоненты) в пространстве, подвергнутом линейному преобразованию с матрицей оператора $\Lambda^{-1/2} Q^T$, которое, как всякое линейное преобразование, является аффинным, т. е. сохраняет все метрические соотношения. Следовательно, при выполнении какого-либо из неравенств (10) для θ_* это неравенство будет выполняться и для любой другой точки θ_{τ} из области G .

Осталось рассмотреть случай применения статистик (13) и (15), когда параметры оцениваются по системе регрессионных уравнений, так как использование в этом случае статистик (5) и (4) (т. е. величин r и s) ничем принципиально не отличается от случая оценивания параметров как средних по выборке или по схеме одного уравнения линейной регрессии.

Представим пространство R^k в виде декартова произведения m подпространств размерности h , т. е.

$$R^k = R^h \times \underset{m \text{ сомножителей}}{R^h \times \dots \times R^h},$$

и в каждом подпространстве R^h введем следующую метрику.

Пусть

$$\tilde{d}_{2,l}(\mathbf{g}, \mathbf{h}) = \sqrt{(\mathbf{g} - \mathbf{h})^T \mathbf{V}_l^{-1} (\mathbf{g} - \mathbf{h})},$$

$$(\mathbf{g}, \mathbf{h} \in R^h),$$

где

$$\mathbf{V}_l = \widehat{\delta}_l^2 (\mathbf{X}^T \mathbf{X})^{-1},$$

$$\widehat{\delta}_l^2 = (n - h - 1)^{-1} \sum_{i=1}^n (z_{l,i} - \mathbf{X}_i \widehat{\xi}_l)^2,$$

\mathbf{X}_i – i -я строка центрированной матрицы \mathbf{X} .

Тогда

$$d_{2,l}(\widehat{\theta}_l, \theta_{*l}) = \tilde{d}_{2,l}(\widehat{\xi}_l, \xi_{*l}) = \sqrt{h \cdot q_l}.$$

Последнее означает, как это следует из равенств (16), что аналогично равенству (19) выполняется и равенство

$$d_2(\widehat{\theta}, \theta_*) = \sqrt{\sum_{l=1}^m d_{2,l}^2(\widehat{\theta}_l, \theta_{*l})}. \quad (20)$$

Теперь, если в R^k ввести метрику

$$d_{\infty(k/h)}(\mathbf{u}, \mathbf{v}) = \max_{l \in \{1, 2, \dots, m\}} d_{2,l}(\mathbf{u}_l, \mathbf{v}_l),$$

то из равенства (20) следует, что эта метрика так же, как и метрика $d_{\infty}(\mathbf{u}, \mathbf{v})$, будет эквивалентна метрике $d_2(\mathbf{u}, \mathbf{v})$.

Иными словами, метрика $d_{\infty(k/h)}(\mathbf{u}, \mathbf{v})$ играет роль $d_{\infty}(\mathbf{u}, \mathbf{v})$ в метрическом факторпространстве размерности m ($R^k / R^h \cong R^m$), подвергнутом, за исключением случая верной гипотезы H_0 , ортогональному, т. е. изометрическому, преобразованию с матрицей оператора \mathbf{R}^T . Вполне очевидно, что выполнение какого-либо из неравенств (17) для θ_* влечет за собой непрямое выполнение этого неравенства и для любой другой точки θ_{τ} из области G .

Таким образом, теорема доказана.

Примеры применения методики в различных областях математического моделирования

Рассматривая примеры применения разработанной методики проверки сложной многомерной гипотезы, мы для вектора параметров θ иногда будем использовать и другие обозначения, соответствующие принятым в рассматриваемой предметной области. Остальные обозначения, по воз-

можности, будем сохранять неизменными. Начнем с рассмотрения задач математической экономики.

Оценка экономической состоятельности предприятий по модели Альтмана. Оценка осуществляется сравнением значения линейной формы

$$\varphi(\mathbf{x}) = \mathbf{c}^T \mathbf{x}$$

от выбранных Θ . Альтманом экономических показателей [9], образующих вектор $\mathbf{x} = (x_1, x_2, \dots, x_5)^T$ размерности $k = 5$, с двумя значениями φ_C и φ_R ($\varphi_C > \varphi_R$). Коэффициенты $\varphi(\mathbf{x})$, образующие вектор (вектор Альтмана) $\mathbf{c} = (c_1, c_2, \dots, c_5)^T$, как и значения φ_C и φ_R , определяются статистическими методами. Вектор \mathbf{c} есть нормаль к гиперплоскости, которая минимизирует сумму квадратов отклонений по данным экономически стабильных предприятий. При условии $\varphi(\mathbf{x}) > \varphi_C$ предприятие считается успешным (принадлежащим области стабильности G_C), а при $\varphi(\mathbf{x}) < \varphi_R$ – несостоятельным (принадлежащим области банкротства G_R). Двойное неравенство

$$\varphi_C \geq \varphi(\mathbf{x}) \geq \varphi_R$$

трактруется как неопределенность (область неопределенности G_N).

Пусть по выборке данных показателей какого-либо стабильного предприятия $\{\mathbf{x}_i, i = 1, 2, \dots, n\}$ получены оценки

$$\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^n \mathbf{x}_i, \quad \mathbf{V}_{\bar{\mathbf{x}}} = n^{-1} \widehat{\mathbf{V}}_{\bar{\mathbf{x}}},$$

$$\text{где } \widehat{\mathbf{V}}_{\bar{\mathbf{x}}} = (n - 1)^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

Тогда, решив задачу Лагранжа для

$$L(\mathbf{x}, \lambda) = \ln(f(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{V}_{\bar{\mathbf{x}}}) - \lambda(\mathbf{c}^T \mathbf{x} - \varphi_C)),$$

найдем точку максимального правдоподобия на граничной плоскости $\varphi(\mathbf{x}) = \varphi_C$.

Решение для этого случая имеет достаточно простой вид:

$$\mathbf{x}_C = \bar{\mathbf{x}} + \lambda \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{c}, \quad \lambda = (\varphi_C - \mathbf{c}^T \bar{\mathbf{x}}) (\mathbf{c}^T \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{c})^{-1}. \quad (21)$$

Используя в решении (21) значение φ_R , получаем точку \mathbf{x}_R . Как было отмечено в конце раздела «Граничная точка...», полученные значения не изменятся, если заме-

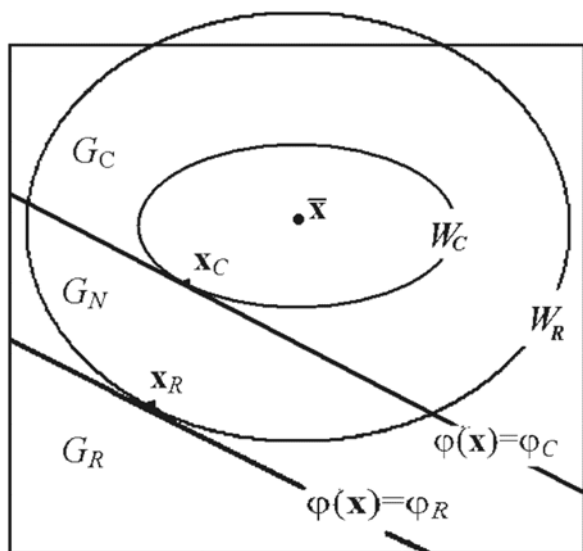


Рис. 1. Схема модели Алтмана для оценки экономической состоятельности предприятий: W_C – эллипсоид надежности; W_R – эллипсоид риска; G_C, G_N, G_R – области успешности, неопределенности и банкротства соответственно (остальные обозначения даны в тексте)

нить в (21) V_x на \hat{V}_x . Более того, для эллипсоида надежности имеют место равенства

$$W_C = \{x : f(x, \bar{x}, V_x) = f(x_C, \bar{x}, V_x)\} = \{x : f(x, \bar{x}, \hat{V}_x) = f(x_C, \bar{x}, \hat{V}_x)\},$$

как и для эллипсоида риска, определяемого точкой x_R (рис. 1).

Очевидно, что экономист предпочтет оценивать вероятностную меру надежности предприятия $P(\text{Int } W_C)$ или меру риска $1 - P(\text{Int } W_R)$ по распределению вектора самих показателей $x \sim N(\bar{x}, V_x)$, а не по распределению оценки вектора его средних значений \bar{x} (с матрицей взаимных ковариаций V_x), так как последнее было бы с его точки зрения легкомысленным благодушием (если учесть, что область допустимых значений ограничена, эти величины можно вычислять в терминах условной вероятности). Аналогично дело обстоит с гипотезами потери стабильности и банкротства. Экономиста вряд ли будет интересовать формулировка этих гипотез в терминах математических ожиданий, т. е. $H_N: Ex \notin G_C$ и $H_R: Ex \in G_R$, соответственно. В последнем случае величины $x_C - \bar{x}$ и $x_R - \bar{x}$, соответ-

ственно, анализировались бы переходом к распределениям t_{n-1} или χ_5^2 (поскольку рассмотрено пять показателей, (см. выше)), от распределения $N(\bar{x}, V_x)$. Это означает, что при вычислении соответствующих статистик использовалась бы матрица V_x (см. формулы (4) и (5)). Однако, если учесть вышеизложенное, то следует использовать матрицу \hat{V}_x . Иными словами, проверяется не возможность попадания средних значений экономических показателей в опасную область, а возможность попадания в опасную область текущих значений этих показателей.

При этом наши статистики уменьшаются, по сравнению с вышеуказанным случаем, на величины \sqrt{n} и n , соответственно. Именно этот момент, собственно, и уменьшает излишнее благодушие. Гипотезы $H_N: x \notin G_C$ и $H_R: x \in G_R$ следует интерпретировать как случаи неизбежной, хоть и случайной потери стабильности и банкротства.

Относительно альтмановской модели можно сделать следующие замечания.

Замечание 1. Если по всей совокупности стабильных предприятий оценены параметры распределения $x \sim N(\theta_x, V_x)$ то $c = P_5$, где P_5 – последний столбец ортогональной матрицы P , такой, что

$$P^T V_x P = \text{diag}(d_1, d_2, \dots, d_5)$$

и $d_1 \geq d_2 \geq \dots \geq d_5$.

Действительно, сумма квадратов отклонений всех значений x от гиперплоскости $c^T x = c^T \theta_x$ минимальна и равна $(N - 1)d_5$, где N – объем всей выборки стабильных предприятий, по которой производилась оценка параметров распределения $N(\theta_x, V_x)$. Кроме того, если определен уровень доверия $(1 - \alpha)$ области стабильности, то

$$\phi_C = c^T \theta_x - (d_5)^{1/2} t_{N-1}^{\alpha/2},$$

а для значения ϕ_R уровень доверия (т. е. $t_{N-1}^{\alpha/2}$) берется предельно допустимым.

Следует отметить, что минимальное собственное значение d_5 , с одной стороны, должно быть отлично от нуля (требование невырожденности или независимости показателей), а с другой стороны, оно должно быть относительно мало (существенно

меньше $\text{tr} \mathbf{V}_x / 5$, где tr – след матрицы). Последнее означает, что эллипсоиды рассеяния распределения $N(\boldsymbol{\theta}_x, \mathbf{V}_x)$ должны быть сжатыми в направлении вектора \mathbf{P}_5 , иначе модель Альтмана не имеет смысла.

Замечание 2. При оцененном распределении $\mathbf{x} \sim N(\boldsymbol{\theta}_x, \mathbf{V}_x)$ можно рассматривать не плоскости, а эллиптические поверхности, задаваемые как

$$W_C^\alpha = \{\mathbf{x} : f(\mathbf{x}, \boldsymbol{\theta}_x, \mathbf{V}_x) = c_\alpha\}$$

или

$$W_C^\alpha = \{\mathbf{x} : (\mathbf{x} - \boldsymbol{\theta}_x)^T \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\theta}_x) = g_\alpha\},$$

где g_α , (так же, как и c_α) зависит от выбора уровня доверия $1 - \alpha$, так что

$$P(\text{Int } W_C^\alpha) = 1 - \alpha$$

(W_R^α определяется аналогично, см. замечание 1).

При оценке отдельного предприятия функция Лагранжа примет вид

$$L(\mathbf{x}, \lambda) = (\mathbf{x} - \bar{\mathbf{x}})^T \hat{\mathbf{V}}_x^{-1} (\mathbf{x} - \bar{\mathbf{x}}) - \lambda((\mathbf{x} - \boldsymbol{\theta}_x)^T \mathbf{V}_x^{-1} (\mathbf{x} - \boldsymbol{\theta}_x) - g_\alpha).$$

Замечание 3. Используя вектор $\mathbf{c} = \mathbf{P}_5$, можно ввести безразмерные показатели [10] относительной вариабельности

$$V_A = \{\mathbf{c}^T \hat{\mathbf{V}}_x \mathbf{c} (\mathbf{c}^T \mathbf{c})^{-1}\} / \text{tr} \hat{\mathbf{V}}_x$$

и относительной стабильности

$$S_A = 1 - \{\mathbf{c}^T \hat{\mathbf{V}}_x \mathbf{c} (\mathbf{c}^T \mathbf{c})^{-1}\} / \lambda_1,$$

где λ_1 – наибольшее собственное значение матрицы $\hat{\mathbf{V}}_x$.

Показатель относительной вариабельности V_A есть отношение дисперсии в направлении вектора \mathbf{c} к общей дисперсии (т. е. V_A определяет долю дисперсии в опасном направлении).

Показатель относительной стабильности S_A приближается к единице, когда дисперсия в направлении вектора \mathbf{c} мала, по сравнению с дисперсией первой главной компоненты (к направлению максимальной дисперсии), и близок к нулю в противоположном случае. Ситуация $S_A \approx 0$ может иметь место, когда направление первой главной компоненты близко к направлению вектора \mathbf{c} (существенные колебания в опасном направлении), или из-за того, что все собственные значения $\hat{\mathbf{V}}_x$ мало различают-

ся между собой (хаотичные колебания).

Здесь мы рассмотрели случай непосредственной оценки параметров распределения. Перейдем к рассмотрению примеров оценки параметров по регрессионной схеме.

Оценка параметров производственной функции. Указанная функция имеет вид

$$Y = aK^{\alpha_1} L^{\alpha_2},$$

где K – капиталовложения (фонды); L – людские ресурсы (труд); Y – выпуск, который осуществляется по данным наблюдений посредством перехода к уравнению линейной регрессии для логарифмов:

$$\ln Y_i = \ln a + \alpha_1 \ln K_i + \alpha_2 \ln L_i + \varepsilon_i \quad (22)$$

($i = 1, 2, \dots, n$).

Экономисты имеют твердое убеждение, что истинные значения параметров α_1 и α_2 удовлетворяют соотношению

$$\alpha_1 + \alpha_2 = 1, \quad (23)$$

т. е. в плоскости с координатами (α_1, α_2) они лежат на прямой (23).

Это убеждение они, как правило, выражают записью производственной функции в виде

$$Y = aK^\alpha L^{1-\alpha}.$$

Однако на практике нередко бывает так, что полученные МНК-оценки удовлетворяют неравенству $\hat{\alpha}_1 + \hat{\alpha}_2 < 1$. В этом случае в качестве окончательного результата в экономике принято брать значения [11]:

$$\alpha_1^e = \frac{\hat{\alpha}_1}{\hat{\alpha}_1 + \hat{\alpha}_2}, \quad \alpha_2^e = \frac{\hat{\alpha}_2}{\hat{\alpha}_1 + \hat{\alpha}_2},$$

которые на координатной плоскости (α_1, α_2) соответствуют точке пересечения прямой, проходящей через начало координат $(0; 0)$ и точку $(\hat{\alpha}_1, \hat{\alpha}_2)$, с прямой (23).

Полученная таким образом точка (вектор)

$$\boldsymbol{\alpha}^e = (\alpha_1^e, \alpha_2^e)^T$$

является центральной проекцией из начала координат.

Если из точки $(\hat{\alpha}_1, \hat{\alpha}_2)$ опускать перпендикуляр на прямую (23) (ортогональная проекция), то получается точка

$$\boldsymbol{\alpha}^o = (\alpha_1^o, \alpha_2^o)^T$$

с координатами

$$\alpha_1^o = \frac{1 + \hat{\alpha}_1 - \hat{\alpha}_2}{2}, \quad \alpha_2^o = \frac{1 + \hat{\alpha}_2 - \hat{\alpha}_1}{2}.$$

В регрессионной модели (22) удобно перейти к центрированным значениям, когда член $\ln a$ отсутствует, что позволяет непосредственно получить (2×2) -матрицу $V_{\hat{\alpha}}$ вектора МНК-оценок $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)^T$. В противном случае эта матрица представляла бы правый нижний блок аналогичной матрицы размера (3×3) для вектора

$$(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)^T, \quad \alpha_0 = \ln a.$$

Если элементы матрицы $V_{\hat{\alpha}}$ обозначить v_{ij} ($i = 1, 2; j = 1, 2$), то для координат граничной точки максимального правдоподобия, лежащей на прямой (23), из решения (21) получаются следующие удобные формулы:

$$\begin{aligned} \alpha_1^* &= \hat{\alpha}_1 + \lambda(v_{1,1} + v_{1,2}), \\ \alpha_2^* &= \hat{\alpha}_2 + \lambda(v_{2,1} + v_{2,2}), \\ \lambda &= \frac{1 - \hat{\alpha}_1 - \hat{\alpha}_2}{v_{1,1} + v_{1,2} + v_{2,1} + v_{2,2}}. \end{aligned}$$

В этой задаче мы столкнулись с ситуацией, когда при проверке гипотезы $H_G: \alpha \in G$ область G совпадает со своей границей $G = \partial G$, так как G — это прямая (23).

Конкретный пример. В табл. 1 приведены значения индексов выпуска, фондов и труда в СССР за период 1958 — 1990 гг.

В табл. 2 представлены результаты, полученные по данным табл. 1; графически они приведены на рис. 2.

Как видно из табл. 2, точки, полученные методами проектирования, не проходят статистическую проверку, т. е. отвергаются гипотезы $H_e: \alpha = \alpha^e$ и $H_0: \alpha = \alpha^o$ при $\alpha = 0,05$. Однако для точки максимального правдоподобия ни одно из неравенств (8), (9) и (10) не выполняется, а, следовательно, гипотеза $H_e: \alpha = \alpha^e$ принимается. Таким образом, условие (23) оказывается выполненным. При $\alpha = 0,01$ все остается на своих местах.

Такой, несколько неожиданный результат объясняется тем, что матрица $V_{\hat{\alpha}}$ близка к вырожденной (т. е. и распределение вектора $\hat{\alpha}$ близко к вырожденному). На это указывают величины собственных значе-

Таблица 1

Индексы выпуска, фондов и труда в СССР за период 1958 — 1990 гг. [12]

Год	Y	K	L
1958	43,20	30,83	61,97
1959	46,45	33,94	64,19
1960	50,17	38,10	68,74
1961	53,59	41,59	73,06
1962	56,63	44,97	75,72
1963	58,90	49,79	78,16
1964	64,38	54,31	81,26
1965	68,81	60,16	85,25
1966	74,39	68,11	88,36
1967	80,85	78,00	91,24
1968	87,55	86,68	94,35
1969	91,68	93,01	97,45
1970	100,00	100,00	100,00
1971	105,65	107,84	102,88
1972	109,81	116,64	105,54
1973	119,62	125,98	108,09
1974	125,98	135,32	110,64
1975	131,73	145,69	113,30
1976	139,48	156,78	115,52
1977	145,81	167,69	117,96
1978	153,32	179,45	120,40
1979	157,10	191,56	122,62
1980	164,82	203,74	124,72
1981	173,55	216,58	126,39
1982	186,64	230,20	127,72
1983	195,43	244,73	128,71
1984	203,11	259,80	129,49
1985	206,29	274,32	130,60
1986	211,03	288,73	131,37
1987	214,41	303,50	131,49
1988	223,85	318,14	129,93
1989	229,43	333,94	127,94
1990	220,26	349,49	125,17

Обозначения: Y — выпуск продукции, K — капиталовложения, L — людские ресурсы. Данные 1970 г. приняты за 100 %.

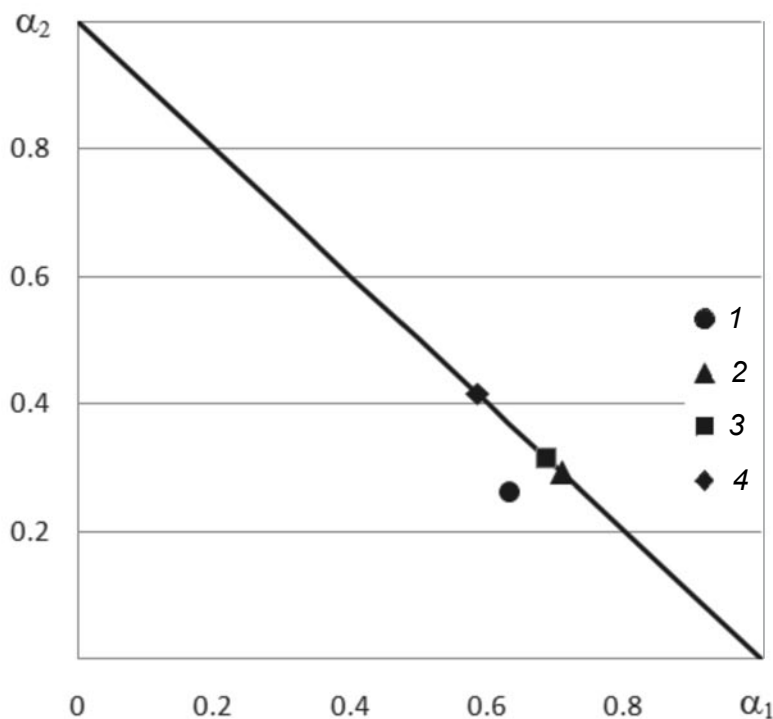


Рис. 2. Значения параметров производственной функции по данным табл. 1:
 1 – МНК-оценка; 2, 3 – центральная и ортогональная проекции;
 4 – точка максимального правдоподобия

Таблица 2

Результаты обработки данных табл. 1

Параметр производственной функции	Значение показателя					
	α_1	α_2	r	q	$ s_1 $	$ s_2 $
МНК-оценка ($n = 33$)	0,63	0,26	—	—	—	—
Центральная проекция	0,71	0,29	252,86	126,43	0,08	15,9
Ортогональная проекция	0,69	0,31	169,83	84,91	0,39	13,03
Точка максимального правдоподобия	0,58	0,42	3,07	1,53	1,74	0,18
Для $\alpha = 0,05$ $\chi_2^2(\alpha) = 5,99$, $F_{2,30}(\alpha) = 3,32$, $t_{30}^{\alpha/2} = 2,36$						

ний этой матрицы, которые отличаются на два порядка: $\lambda_1 = 8,62 \cdot 10^{-3}$; $\lambda_2 = 2,7 \cdot 10^{-5}$, а, следовательно, оси эллипсоида рассеяния (или надежности, см. раздел «Граничная точка...») отличаются на порядок. Поэтому направленный отрезок ($\hat{\alpha} \rightarrow \alpha^*$), име-

ющий координаты $(-0,046; 0,155)$, почти коллинеарен первому собственному вектору матрицы $V_{\hat{\alpha}}$ (направлению максимальной дисперсии, т. е. первому столбцу Q , (см. формулу (4)), который имеет координаты $(-0,294; 0,956)$.

Задача космической безопасности. Одна из таких задач состоит в оценке вероятности схода с орбиты космического объекта, каковыми являются спутники связи, спутники навигации и прочие. Рассмотрим в первом приближении самую простую модель движения космического объекта. Будем предполагать, в соответствии с законами Кеплера, что орбита является эллиптической и лежит в плоскости, проходящей через центр массы Земли, с которым совпадает один из фокусов эллипса орбиты. В этом случае она может быть описана в стандартных полярных координатах (ρ, φ) уравнением вида [13]

$$1/\rho = \theta_0 + \theta_1 \cos \varphi + \theta_2 \sin \varphi.$$

Параметры орбиты образуют трехмерный вектор $\theta = (\theta_0, \theta_1, \theta_2)^T$. Орбита, которая не пересекает и даже не касается сферы определенного радиуса R_G ($R_G > R_0$, R_0 – радиус Земли), будет считаться безопасной.

Предположим, что по выборке инструментальных наблюдений $\{(\rho_i, \varphi_i), i = 1, 2, \dots, n\}$ и формулам (6) получена оценка вектора параметров орбиты $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)^T$ и матрица взаимных ковариаций этой оценки $V_{\hat{\theta}}$. В данном случае элементы регрессионной модели $Y = X\theta + \varepsilon$ таковы: вектор Y имеет размерность n , X – $(n \times 3)$ -матрица, а компоненты вычисляются по формулам

$$y_i = 1/\rho_i, x_{i1} = 1, x_{i2} = \cos \varphi_i, \\ x_{i3} = \sin \varphi_i \quad (i = 1, 2, \dots, n),$$

соответственно.

Далее по формуле (21) находим граничную точку максимального правдоподобия θ_* для распределения $N(\hat{\theta}, V_{\hat{\theta}})$. В этом случае формулы, как и в приведенных выше примерах, имеют простой, но принципиально отличный, зависящий от φ вид:

$$\theta_*(\varphi) = \hat{\theta} + \lambda V_{\hat{\theta}} c(\varphi), \quad (24)$$

$$\lambda = (c_G - c(\varphi)^T \hat{\theta})(c(\varphi)^T V_{\hat{\theta}} c(\varphi))^{-1},$$

так как граница ∂G задается условием

$$c(\varphi)^T \theta = c_G,$$

где $c(\varphi) = (1, \cos \varphi, \sin \varphi)^T$, $c_G = R_G^{-1}$.

Имея значения $\theta_*(\varphi)$, можно вычислять эллипсоид надежности (см. раздел «Гра-

ничная точка...») и таким образом оценивать безопасность объекта в вероятностном смысле, или проверять гипотезу катастрофической ситуации

$$H_G : c(\varphi)^T \theta \geq c_G.$$

При этом (при верной гипотезе H_G) мы имеем полный набор статистик (4), (5) и (7):

$$s_j \sim t_{n-3} \quad (j = 1, 2, 3), \quad r \sim \chi_3^2, \quad q \sim F_{3, n-3}.$$

Зависимость точки $\theta_*(\varphi)$ (24) от угловой координаты φ дает возможность определить момент наступления опасной ситуации. Наконец, напомним, что здесь показана лишь общая идея решения задачи, в то же самое время предполагается, что данное решение незначительно изменится, если усложнить модель учетом всех необходимых и существенных факторов.

Завершая краткий обзор задач, в которых может быть использован рассмотренный здесь метод проверки сложных многомерных статистических гипотез, отметим в самых общих чертах еще два достаточно важных направления.

Задача проектирования машин и механизмов. В указанной задаче при проведении серии испытаний экспериментального образца получаем выборку вектора замеренных параметров и соответствующие статистические оценки. Это позволяет оценить надежность соответствия образца техническим требованиям, выражающимся в принадлежности вектора его параметров некоторой области. Когда параметры измеряются непосредственно, не по схеме линейной регрессии, а оценки даются осреднением, задача отличается от модели Альтмана лишь тем, что границы области требуемых значений параметров определяются самой задачей, для решения которой и предназначен проектируемый механизм.

Математическое моделирование реальных процессов. Как было отмечено во введении, применение дифференциальных уравнений сталкивается с проблемой оценки параметров по данным наблюдений за моделируемым процессом. При этом возникает вопрос о принадлежности истинных значений вектора параметров той или иной

области, определяющей поведение системы, например, о принадлежности области устойчивости решения. Примером здесь может служить не раз упомянутая выше работа [1], которая, кроме того, может служить примером случая, когда параметры модели оцениваются по системе регрессионных уравнений. При этом следует обратить внимание на уточнение относительно F -статистик, сделанное в конце раздела «Проверка простой гипотезы относительно граничной точки максимального правдоподобия».

Заключение

Очевидно, что предложенный подход к проверке сложных статистических гипотез, относящихся к вектору параметров, позволяет решать достаточно широкий класс задач математического моделирования. Предлагаемый метод проиллюстрирован как в опубликованной ранее работе [1], так

и в данной статье на ряде примеров.

Благодарности

Автор выражает глубокую признательность доктору физико-математических наук, профессору Л.А. Петросяну, декану факультета прикладной математики – процессов управления Санкт-Петербургского государственного университета, обратившему внимание автора на особую роль принципа максимального правдоподобия, а также доктору физико-математических наук, профессору О.А. Малафееву, заведующему кафедрой моделирования социально-экономических систем, и кандидату физико-математических наук А.В. Кондрашкову, доценту кафедры прикладных информационных технологий Санкт-Петербургского государственного экономического университета, за ознакомление с проблематикой примеров, рассмотренных в статье.

СПИСОК ЛИТЕРАТУРЫ

- [1] Кондрашков А.В., Пичугин Ю.А. Идентификация и статистическая проверка устойчивости модели Вольтерры // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2014. № 1(189). С. 124–135.
- [2] Пичугин Ю.А., Кондрашков А.В. Геометрические аспекты проверки статистических гипотез в задачах математического моделирования // Некоторые актуальные проблемы современной математики и математического образования. Герценовские чтения – 2014: Материалы научной конференции, 14 – 18 апреля 2014. СПб.: РГПУ им. А.И. Герцена, 2014. С. 230–236.
- [3] Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. 456 с.
- [4] Пичугин Ю.А. Вероятностные оценки риска и надежности в экономике // Научная сессия ГУАП. Сб. докл.: В 3-х ч. Ч. 1. Технические науки. СПб.: ГУАП, 2013. С. 195–196.
- [5] Пичугин Ю.А., Малафеев О.А. Об оценке риска банкротства фирмы // VI Международная конференция «Динамические системы: устойчивость, управление, оптимизация (DSSCO'13)», посвященная 95-летию со дня рождения Е.А. Барбашина, 1–5 октября 2013 г., г. Минск. С. 204–206.
- [6] Магнус Я.Р., Нейдеккер Х. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике. М.: Физматлит, 2002. 496 с.
- [7] Вернер А.Л., Кантор Б.Е., Франгулов С.А. Геометрия. Ч. 2. СПб.: Специальная литература, 1997. 322 с.
- [8] Пичугин Ю.А. Проверка сложной статистической гипотезы в линейном регрессионном анализе // Научная сессия ГУАП. Сб. докл.: В 3-х ч. Ч. 1. Технические науки. СПб.: ГУАП, 2014. С. 275–277.
- [9] Altman E. Managing Credit Risk, 2-nd ed. John Wiley and Sons, 2008.
- [10] Пичугин Ю.А., Кондрашков А.В. Многомерные критерии моделей математической экономики на основе альтмановского подхода // Развитие социально-культурной и экономической деятельности в России: региональный аспект. Материалы II Всерос. научно-практической конференции. СПб., 29–30 апреля 2014 г. СПб.: ГПА, 2014. С. 92–97.
- [11] Колемаев В.А. Математическая экономика. Учебник для вузов. 3-е изд. М.: Юнити–Диана, 2005. 399 с.
- [12] Бессонов В.А. Проблемы построения производственных функций в российской переходной экономике // В кн.: Бессонов В.А., Цухло С.В. Анализ динамики российской переходной экономики. М.: Институт экономики переходного периода, 2002. 589 с.
- [13] Пичугин Ю.А., Кондрашков А.В. Вероятностная оценка для одной из задач космической

безопасности на основе кеплеровской модели // Научная сессия ГУАП. Сб. докл.: В 3-х ч. Ч 1. Технические науки. СПб.: ГУАП, 2014. С. 273–275.

СВЕДЕНИЯ ОБ АВТОРЕ

ПИЧУГИН Юрий Александрович – доктор физико-математических наук, профессор кафедры геометрии Российского государственного педагогического университета им. А.И. Герцена. 191186, Российская Федерация, г. Санкт-Петербург, наб. р. Мойки, 48
yury-pichugin@mail.ru

Pichugin Yu.A. GEOMETRICAL ASPECTS OF TESTING THE COMPLEX STATISTICAL HYPOTHESES IN MATHEMATICAL SIMULATION.

It is well known that mathematical simulation parameters often are obtained by statistical estimating. Therefore the problem of testing the complex statistical hypotheses such as the one about an adjunct of a vector of model parameters to some domain is of current concern. This article deals with the problem in geometrical aspects. The basic theorem to solve this problem has been stated and proved. The theorem asserts that the solution can be done through testing some simple statistical hypothesis concerning a boundary point of maximum likelihood. The theorem proof is based on the use of generalized Euclidean metric and an affine transformation of parameter space. Typical examples of its use for different mathematical models are also considered. They are the following: (i) Altman's model of the economic stability and risk estimating; an estimation of specific enterprise is treated in terms of the statistical hypotheses testing; (ii) the method to refine statistical estimations of production function parameters; (iii) the statistical estimation of the space object dynamic stability is considered on the basis of Kepler's model, as well.

BOUNDARY POINT OF MAXIMUM LIKELIHOOD, LAGRANGE TASK, ELLIPSOID OF RELIABILITY, GENERALIZED EUCLIDEAN METRIC.

REFERENCES

- [1] **A.V. Kondrashkov, Yu.A. Pichugin**, On the identification and statistical testing stability of Volterra model, St. Petersburg State Polytechnical University Journal: Physics and Mathematics. 1 (189) 2014, 124–135.
- [2] **Yu.A. Pichugin, A.V. Kondrashkov**, Geometrical aspects of the statistical hypothesis testing in mathematical models, Proceedings of Scientific Conference “Herzen reading – 2014”, Herzen State Pedagogical University of Russia, St. Petersburg, 2014, 230–236.
- [3] **G.A.F. Seber**, Linear regression analysis, Mir, Moscow, 1980.
- [4] **Yu.A. Pichugin**, Veroyatnostnyye otsenki riska i nadezhnosti v ekonomike [Probabilistic risk and reliability assessment in economics], Nauchnaya sessiya GUAP: Sb. dokl.: Ch. 1. Tekhnicheskiye nauki. SPb.: GUAP, St. Petersburg, 2013, pp. 195–196.
- [1] **Yu.A. Pichugin, O.A. Malafeyev**, Ob otsenke riska bankrotstva firmy [On the assessment of the bankruptcy risk of a company]. VI Mezhdunarodnaya konferentsiya “Dinamicheskiye sistemy: ustoychivost, upravleniye, optimizatsiya” (DSSCO'13), posvyashchennaya 95-letiyu so dnya rozhdeniya Ye.A. Barbashina, 1–5 oktyabrya 2013 goda, Minsk, pp. 204–206.
- [6] **Ya.R. Magnus, Kh. Neydekker**, Matrichnoye differentsialnoye ischisleniye s prilozheniyami k statistike i ekonometrike [Matrix differential calculation with applications in statistics and econometrics]. Fizmatlit, Moscow, 2002.
- [7] **A.L. Verner, B.Ye. Kantor, S.A. Frangulov**, Geometriya, chast 2 [Geometry, part 2]. Spetsialnaya l-ra, Moscow, 1997.
- [8] **Yu.A. Pichugin**, The testing of complex statistical hypothesis in linear regression analysis, Proceedings of Scientific Conference “Nauchnaya sessiya GUAP” [Scientific session of SUAI], V. 1, Technical Sciences, St. Petersburg State University of Aerospace Instrumentation, 2014, pp. 275–277.
- [9] **E. Altman**, Managing Credit Risk, 2nd Edition. John Wiley and Sons, 2008.
- [10] **Yu.A. Pichugin, A.V. Kondrashkov**, Mnogomernyye kriterii modeley matematicheskoy ekonomiki na osnove altmanovskogo podkhoda [Multidimensional models of mathematical economics criteria based on the Altman approach], in: Razvitiye sotsialno-kulturnoy i ekonomicheskoy deyatel'nosti v Rossii: regionalnyy aspekt. Materialy II Vseross. nauchno-prakticheskoy konferentsii. (S.-Peterburg. 29-30 aprelya 2014 g.). GPA, St. Petersburg, 2014, pp. 92–97.
- [11] **V.A. Kolemeyev**, Matematicheskaya ekonomika: Uchebnik dlya vuzov [Mathematical economics. Textbook for universities]. 3rd ed.

Yuniti-Diana, Moscow, 2005.

[12] **V.A. Bessonov**, Problemy postroyeniya proizvodstvennykh funktsiy v rossiyskoy perekhodnoy ekonomike [Problems of constructing production functions in Russian transition economy], in: Bessonov V.A., Tsukhlo S.V. Analiz dinamiki rossiyskoy perekhodnoy ekonomiki. Institut ekonomiki perekhodnogo perioda, Moscow, 2002.

[13] **Yu.A. Pichugin, A.V. Kondrashkov**, The probabilistic estimate of the stability and hazard of the dynamics of a space object on the basis of Kepler's model, Proceedings of Scientific Conference "Nauchnaya sessiya GUAP" [Scientific session of SUAI], V. 1, Technical Sciences, St. Petersburg State University of Aerospace Instrumentation, 2014, pp. 273–275.

THE AUTHOR

PICHUGIN Yury A.

Herzen State Pedagogical University of Russia
48, Moyka Emb., St.Petersburg, 191186, Russian Federation
yury-pichugin@mail.ru