Математика

DOI: 10.18721/JPM.12314
УДК 519.24

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИИ ПО ШЕННОНУ В ЗАДАЧАХ, СВЯЗАННЫХ С ЛИНЕЙНОЙ РЕГРЕССИЕЙ

Ю.А. Пичугин

Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, Российская Федерация

В статье рассматривается использование количества информации по Шеннону (SIQ) в задачах, связанных с линейной регрессией. Показано, что SIQ, содержащееся в компонентах отклика относительно стохастических параметров, выражается через информационную матрицу Фишера, является выпуклым функционалом на множестве компонент отклика и при достаточно большом масштабе параметров эквивалентно использованию *D*-критерия в задачах планирования эксперимента. Определено SIQ относительно постоянных параметров регрессии. Рассмотрена альтернативная постановка задачи оптимального планирования эксперимента (OEP) и проанализирована ее связь с традиционной постановкой. Рассмотрена задача информационного упорядочивания данных при использовании регрессии на базис главных компонент. Предложены алгоритмы, учитывающие ценность информации при наличии частичных пропусков данных.

Ключевые слова: количество информации по Шеннону, линейная регрессия, стохастический параметр регрессии

Ссылка при цитировании: Пичугин Ю.А. Особенности использования информации по Шеннону в задачах, связанных с линейной регрессией// Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 12. № 3. С. 164–176. DOI: 10.18721/JPM.12314

THE SHANNON INFORMATION QUANTITY IN THE TASKS ASSOCIATED WITH LINEAR REGRESSION: USAGE PATTERN

Yu.A. Pichugin

Saint-Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russian Federation

The article discusses the use of the Shannon information quantity (SIQ) in the tasks associated with linear regression. It has been shown that the SIQ contained in the response components with respect to stochastic parameters is expressed through the Fisher information matrix, is a convex functional on the set of response components, and is equivalent to the use of the *D*-criterion in the problems of planning the experiment at a sufficiently large scale of parameters. The SIQ relatively constant regression parameters were determined. An alternative formulation of the optimal experiment planning (OEP) problem was considered and its relation to the traditional formulation was analyzed. The problem of information ordering of data using regression on the basis of principal components was considered. Some algorithms taking into account the value of information in the presence of partial data gaps were proposed.

Keywords: Shannon information quantity, linear regression, stochastic regression parameter

Citation: Pichugin Yu.A., The Shannon information quantity in the tasks associated with linear regression: usage pattern, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (3) (2019) 164–176. DOI: 10.18721/JPM.12314

Введение

Объектом исследования настоящей работы является информация по Шеннону. Кратко перечислим основные этапы развития этого понятия.

Первоначально определенное Клодом Элвудом Шенноном количество информации (SIQ -Shannon information quantity) в передаваемом сообщении тесным образом связано с понятием энтропии. Средняя энтропия $H(\xi)$, или информация, передаваемая в сообщении ξ посредством n символов (значений)

$$\{x_1, x_2, ..., x_n\},$$

появляющихся с вероятностями

$$\{P_\xi(1), P_\xi(2), ..., P_\xi(n)\},$$

следует выражению [1]:

$$H(\xi) = -\sum_{i=1}^{n} P_\xi(i) \cdot \log_2 P_\xi(i) = \sum_{i=1}^{n} P_\xi(i) H_i,$$

где $H_i(\xi) = -\log_2 P_\xi(i)$ есть частная энтропия.

В качестве основания логарифма, в принципе, может быть взято любое число $a > 1$, которое задает масштаб.

Если имеется другая случайная величина η, принимающая m значений, т.е.

$$\{y_1, y_2, ..., y_m\},$$

с вероятностями

$$\{P_\eta(1), P_\eta(2), ..., P_\eta(m)\},$$

то, согласно Шеннону, количество информации $I(\xi, \eta)$, содержащееся в сообщении ξ относительно сообщения η или в η относительно ξ (здесь имеет место симметрия), выражается как

$$I(\xi, \eta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P_{\xi\eta}(i, j) \times \log_a \frac{P_{\xi\eta}(i, j)}{P_\xi(i) P_\eta(j)},$$

где $P_{\xi\eta}(i, j)$ – вероятность того, что ξ принимает значение x_i, а η принимает значение y_j, соответственно.

При $P_{\xi\eta}(i, j) = 0$ считается, что соответствующий член суммы равен нулю.

В своей работе И.М. Гельфанд и А.М. Яглом [2] показали, что в случае, когда $\boldsymbol{\xi}$ и $\boldsymbol{\eta}$ есть два случайных гауссовских вектора, то количество информации $I(\boldsymbol{\xi}, \boldsymbol{\eta})$, содержащееся в векторе $\boldsymbol{\xi}$ относительно вектора $\boldsymbol{\eta}$ и наоборот, равно

$$I(\boldsymbol{\xi}, \boldsymbol{\eta}) = -\frac{1}{2} \log_a \det\left(\mathbf{I} - \mathbf{V}_{\xi\eta} \mathbf{V}_\eta^{-1} \mathbf{V}_{\eta\xi} \mathbf{V}_\xi^{-1}\right), \qquad (1)$$

где $\mathbf{V}_\xi$, $\mathbf{V}_\eta$ – матрицы взаимных ковариаций компонент векторов $\boldsymbol{\xi}$ и $\boldsymbol{\eta}$, соответственно; $\mathbf{I}$ – единичная матрица соответствующей размерности; $\mathbf{V}_{\xi\eta}$ – матрица взаимных ковариаций компонент вектора $\boldsymbol{\xi}$ и вектора $\boldsymbol{\eta}$,

$$\mathbf{V}_{\eta\xi} = \mathbf{V}_{\xi\eta}^T,$$

где T – оператор транспонирования.

При этом в работе [2] была показана независимость SIQ от масштаба переменных и инвариантность относительно линейных преобразований векторов $\boldsymbol{\xi}$ и $\boldsymbol{\eta}$. Последнее позволяет представить формулу (1) в виде

$$I(\boldsymbol{\xi}, \boldsymbol{\eta}) = -\log_a \left|\sin\alpha_1 \sin\alpha_2 ... \sin\alpha_h\right|, \qquad (2)$$

где $\cos\alpha_j = r_j$ ($j = 1,2,...,h$) есть не что иное, как геометрическая интерпретация коэффициента корреляции r_i между парой независимых (главных) компонент этих векторов. При этом $h = \min(n, m)$. Формула (1), очевидно, позволяет применять SIQ в многомерных задачах.

Из общей массы прикладных работ с применением SIQ отметим работу О.М. Покровского [3], где SIQ применяется к решению задачи оптимизации спутниковых наблюдений на основе модели линейной регрессии. Особенность этой работы состоит в том, что по своему характеру она весьма близка к задаче оптимального планирования регрессионного эксперимента (опуская слово «регрессионного», приходим к аббревиатуре OEP – optimal experiment planning). Но в работе [3] автор рассматривает случай стохастических параметров регрессии (см. далее раздел «Связь SIQ с информационной матрицей Фишера»), отступая и от традиционной постановки задачи OEP в целом (см. еще далее раздел «Традиционная и альтернативная формулировки задачи оптимального планирования эксперимента»).

Цель настоящей работы – показать связи SIQ с информационной матрицей Фишера и с критериями OEP в альтернативной постановке задачи, развивая при этом идеи работ [2, 3], а также раскрыть новые возможности применения этого математического аппарата в различных областях прикладной сферы.

Связь SIQ с информационной матрицей Фишера

В классической постановке задачи регрессионного анализа мы имеем следующую модель:

$$\mathbf{y} = \mathbf{F\theta} + \boldsymbol{\varepsilon}, \tag{3}$$

где $\mathbf{y}$ – вектор измерений размерности n; $\boldsymbol{\theta}$ – вектор параметров, подлежащих оценке, размерности m; $\mathbf{F}$ – $(n \times m)$-матрица $(n \geq m)$; $\boldsymbol{\varepsilon}$ –вектор случайных ошибок.

Здесь предполагается, что компоненты вектора $\boldsymbol{\varepsilon}$ подчиняются многомерному нормальному распределению, имеют нулевые средние значения, взаимно независимы и имеют одинаковую дисперсию σ^2, т. е.

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

где $\mathbf{0}$ – нулевой вектор.

Следовательно,

$$\mathbf{y} \sim N(\mathbf{F\theta}, \sigma^2\mathbf{I}),$$

т. е. $\mathbf{V}_y = \mathbf{V}_\varepsilon$.

Кроме этого, будем предполагать, что $\mathbf{F}$ – матрица полного ранга.

Информационная матрица Фишера вектора параметров регрессии $\boldsymbol{\theta}$, которая представляет собой производное от информации, содержащейся в компонентах отклика относительно этих параметров регрессии, определяется следующей формулой [4]:

$$\mathbf{M} = -\mathrm{E}\left\{\frac{\partial^2 L}{\partial\theta_i\theta_j}\right\},$$

где E – оператор математического ожидания, L – логарифм правдоподобия, т. е.

$$L = \ln\left\{\left(2\pi\right)^{-\frac{n}{2}} \det^{\frac{-1}{2}}\left(\mathbf{V}_\varepsilon\right)\times\right.$$

$$\left.\times\exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{F\theta}\right)^T \mathbf{V}_\varepsilon^{-1}\left(\mathbf{y}-\mathbf{F\theta}\right)\right]\right\}.$$

Отсюда следует, что

$$\mathbf{M} = \sigma^{-2}\mathbf{F}^T\mathbf{F} = \mathbf{V}_{\hat{\theta}}^{-1}, \tag{4}$$

где

$$\mathbf{V}_{\hat{\theta}} = \sigma^2\left(\mathbf{F}^T\mathbf{F}\right)^{-1} \tag{5}$$

есть матрица взаимных ковариаций OLS-оценок (OLS – ordinary least squares) компонент вектора параметров регрессии $\boldsymbol{\theta}$

Различные выпуклые функционалы от информационной матрицы $\mathbf{M}$ (см. далее) могут служить мерой информации, содержащейся в компонентах отклика относительно параметров регрессии (3). Вектор OLS-оценок этих параметров регрессии в данном случае выражается как

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{y} \tag{6}$$

(см. монографию [5]).

При нарушении предположений относительно ковариационной матрицы остатков регрессии, когда эта матрица имеет произвольную структуру, т. е.

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

вместо формул (4) – (6) мы имеем, соответственно, следующие выражения:

$$\mathbf{M} = \mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{F} = \mathbf{V}_{\hat{\theta}}^{-1}, \tag{7}$$

$$\mathbf{V}_{\hat{\theta}} = \left(\mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{F}\right)^{-1}, \tag{8}$$

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{y}. \tag{9}$$

Формула (9) известна как обобщенный метод OLS-оценивания.

Из формул (5), (8) следует, что вектор OLS-оценки $\boldsymbol{\theta}$ имеет нормальное распределение

$$\hat{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\theta}, \sigma^2(\mathbf{F}^T\mathbf{F})^{-1}\right)$$

или

$$\hat{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\theta}, (\mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{F})^{-1}\right)$$

(в зависимости от структуры матрицы $\mathbf{V}_\varepsilon$).

Каждое из этих распределений определяет соответствующую функцию правдоподобия. Вычисленные по этим функциям правдоподобия информационные матрицы Фишера вектора $\boldsymbol{\theta}$, как нетрудно проверить, совпадают с матрицами, которые даются формулами (4) и (7), соответственно.

Предположим, что параметры регрессии имеют стохастическую природу и подчиняются многомерному нормальному распределению, т. е.

$$\boldsymbol{\theta} \sim N(\mathrm{E}\boldsymbol{\theta}, \mathbf{V}_\theta).$$

Тогда для OLS-оценок компонент вектора параметров регрессии (3) $\boldsymbol{\theta}$, несмотря на то, что распределение вектора $\mathbf{y}$ будет уже иным, по-прежнему верны формулы (5), (6) или (8), (9), в зависимости от структуры матрицы $\mathbf{V}_\varepsilon$ а также не изменится вид вероятностного распределения вектора $\hat{\boldsymbol{\theta}}$ и информационной матрицы Фишера (см. формулы (4), (7)). Однако какой-либо выпуклый функционал от этой матрицы будет служить мерой информации относительно вектора параметров $\boldsymbol{\theta}$, содержащейся в его OLS-оценке. Именно эта информационная матрица (как более инвариантная) имеется в виду при рассмотрении случая стохастических параметров, для которого имеет место следующая теорема.

Теорема 1. *Количество информации* $I(\mathbf{y}, \boldsymbol{\theta})$*, содержащееся в компонентах отклика в модели* (3) *относительно стохастических, нормально распределенных параметров, при условии, что* $\mathbf{F}$ *есть матрица полного ранга, связано с информационной матрицей Фишера* $\mathbf{M}$ *формулой*

$$I(\mathbf{y}, \boldsymbol{\theta}) = \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{M}\mathbf{V}_\theta\right). \quad (10)$$

Доказательство. В предположениях стохастических параметров регрессии (3) имеем следующие равенства:

$$\mathbf{V}_y = \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T + \mathbf{V}_\varepsilon,$$

$$\mathbf{V}_{y\theta} = \mathbf{F}\mathbf{V}_\theta \text{ и } \mathbf{V}_{\theta y} = \mathbf{V}_\theta\mathbf{F}^T.$$

Используя эти выражения, преобразуем формулу (1):

$$I(\mathbf{y}, \boldsymbol{\theta}) = -\frac{1}{2}\log_a \det\left(\mathbf{I} - \mathbf{F}\mathbf{V}_\theta\mathbf{V}_\theta^{-1}\mathbf{V}_\theta\mathbf{F}^T\mathbf{V}_y^{-1}\right) =$$

$$= -\frac{1}{2}\log_a \det\left(\mathbf{V}_y\mathbf{V}_y^{-1} - \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T\mathbf{V}_y^{-1}\right) =$$

$$= -\frac{1}{2}\log_a \det\left(\left[\mathbf{V}_y - \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T\right]\mathbf{V}_y^{-1}\right) =$$

$$= -\frac{1}{2}\log_a \det\left(\mathbf{V}_\varepsilon\mathbf{V}_y^{-1}\right) = \frac{1}{2}\log_a \det\left(\mathbf{V}_\varepsilon^{-1}\mathbf{V}_y\right) =$$

$$= \frac{1}{2}\log_a \det\left(\mathbf{V}_\varepsilon^{-1}\left[\mathbf{V}_\varepsilon + \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T\right]\right) =$$

$$= \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{V}_\varepsilon^{-1}\mathbf{F}\mathbf{V}_\theta\mathbf{F}^T\right) \overset{(*)}{=}$$

$$= \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{F}\mathbf{V}_\theta\right) =$$

$$= \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{M}\mathbf{V}_\theta\right).$$

В равенстве (*) мы использовали алгебраическое тождество

$$\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A}),$$

которое очевидно в случае квадратных матриц и обратимости хотя бы одной из двух.

Действительно, пусть матрица $\mathbf{A}$ обратима. Тогда, умножая под знаком определителя левую часть этого равенства слева на $\mathbf{A}^{-1}$ а справа на $\mathbf{A}$, получаем тождество. Такое преобразование, именуемое преобразованием подобия, как известно, не меняет значения определителя.

Пусть теперь $\mathbf{A}$ и $\mathbf{B}$ – прямоугольные матрицы; $\mathbf{A}$ – матрица размерности $n \times m$, $\mathbf{B}$ – матрица $m \times n$ ($n > m$).

Согласно формулировке теоремы, мы вправе предполагать, что $\mathbf{A}$ – матрица полного ранга. Не умаляя общности, будем считать, что первые m строк матрицы $\mathbf{A}$ линейно независимы. Достраиваем прямоугольные матрицы $\mathbf{A}$ и $\mathbf{B}$ до квадратных размерности $n \times n$:

$$\mathbf{A}^* = (\mathbf{A}|\mathbf{T}) \text{ и } (\mathbf{B}^*)^T = (\mathbf{B}^T|\mathbf{T}),$$

где блок $\mathbf{T}$ имеет размерность $n \times (n - m)$. Заполним блок $\mathbf{T}$ по следующему правилу:

$$t_{i+m,j} = a^*_{i+m,j+m} = \delta \ \ (\delta \neq 0),$$

$$\text{если } i = j, \ j = 1, 2, \ldots, n - m;$$

остальные элементы $\mathbf{T}$ заполняем нулями.

Тогда для матриц $\mathbf{A}^*$ и $\mathbf{B}^*$ требуемое тождество выполняется (см. выше). Далее, устремляя δ к нулю, имеем требуемое тождество для исходных матриц $\mathbf{A}$ и $\mathbf{B}$.

Теорема 1 полностью доказана.

Формула (10) и ее вывод были ранее представлены автором данного исследования в работах [6, 7].

В предположении стохастических параметров регрессии (3), в процессе вывода формулы (10), обнаруживается другая формула, а именно

$$I(\mathbf{y}, \boldsymbol{\theta}) = \frac{1}{2}\log_a \det\mathbf{V}_y - \frac{1}{2}\log_a \det\mathbf{V}_\varepsilon.$$

Это равенство очень хорошо раскрывает суть SIQ в модели (3) при стохастических параметрах.

Традиционная и альтернативная формулировки задачи оптимального планирования эксперимента

Традиционная формулировка. В первоначальной (и ставшей традиционной) формулировке задача OEP в условиях моа дели (3) сводится к вычислению множества величин [8, 9]:

$$\mathbf{p}=\{p_1, p_2, ..., p_n\},$$

$$p_i \geq 0;\ i = 1, 2, ..., n;\ \sum_{i=1}^{n} p_i = 1, \quad (11)$$

именуемого планом.

Эти величины интерпретируются как относительные частоты измерения компонент отклика (отсюда название – план), аналогично закону дискретного распределения вероятностей. Нечто подобное имеет место в теории матричных игр при использовании смешанных стратегий [10], откуда, по-видимому, и заимствована эта идея. В теории OEP величины, составляющие план, используются для преобразования информационной матрицы, которое имеет следующий вид (в рамках указанной теории):

$$\mathbf{M} \to \mathbf{M}(\mathbf{p}) = (\mathbf{F}^T \mathbf{P}\mathbf{F}), \quad (12)$$

где $\mathbf{P} = \mathrm{diag}\,(p_1, p_2, \ldots, p_n)$.

При этом на план налагается требование максимизации некоторого выпуклого функционала (критерия) от этой преобразованной матрицы (см. ниже).

Таким образом, в процессе решения задачи осуществляется переход от начальных значений $p_i = 1/n$ ($i = 1,2, \ldots, n$) к оптимальным, при которых некоторый критерий от преобразованной (см. выше) информационной матрицы достигает максимума, т. е.

$$\mathrm{crit}\mathbf{M}(\mathbf{p}) = \mathrm{crit}\ \mathbf{F}^T\mathbf{P}\mathbf{F} \to \max.$$

В качестве критериев (функционалов от $\mathbf{M}(\mathbf{p})$) рассматриваются определитель (D-критерий), след (A-критерий), минимальное собственное значение матрицы $\mathbf{M}(\mathbf{p})$ (E-критерий) и иные выпуклые функционалы от этой матрицы [8, 9].

Уточняющее замечание. Как нетрудно было заметить выше, в OEP-теории под информационной матрицей понимается не в точности информационная матрица Фишера, равная $\sigma^{-2}\mathbf{F}^T\mathbf{F}$ (см. формулу (4)), а пропорциональная матрица $\mathbf{F}^T\mathbf{F}$ [8, 9] или $\mathbf{F}^T\mathbf{P}\mathbf{F}$ (см. выше), что для задачи максимизации функционалов от этой матрицы не имеет принципиального значения. С другой стороны, из регрессионного анализа, описанного, например, в монографии [5], следует, что переход к матрице $\mathbf{F}^T\mathbf{P}\mathbf{F}$ имеет смысл лишь тогда, когда используется оценка параметров

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{F}^T \mathbf{P}\mathbf{F}\right)^{-1} \mathbf{F}^T \mathbf{P}\mathbf{y},$$

известная как обобщенная OLS-оценка (см. формулу (9)), или, при диагональной структуре $\mathbf{P}$, как взвешенная OLS-оценка, что, собственно, мы здесь и имеем. Из теоремы же Гаусса – Маркова, доказательство которой приводится в книге [9], следует, что корректность применения такой оценки имеет место лишь в случае, когда, во-первых, $\mathbf{V}_\varepsilon \neq \sigma^2\mathbf{I}$, а во-вторых, когда матрица $\mathbf{P}$ обратно пропорциональна матрице $\mathbf{V}_\varepsilon$ т. е.

$$\mathbf{P} = c\mathbf{V}_\varepsilon^{-1}\ (c > 0).$$

Однако упоминания об этих условиях в теории OEP не обнаруживается [8, 9]. Остается лишь предполагать, что поспешное отбрасывание множителей σ^2 и σ^{-2} в формулах для $\mathbf{V}_{\hat{\theta}}$ и $\mathbf{M}$, а иногда и ε в некоторых регрессионных уравнениях (см. [9]), привели к тому, что преобразование (12) оказалось не связанным со структурой ковариационной матрицы погрешностей регрессии (3), несмотря на то, что конечной целью является именно оценка параметров.

Альтернативная формулировка. Не менее практически значимой является следующая формулировка.

Будем рассматривать различные собственные подмножества вида

$$\mathbf{q} \subset \{1, 2, ..., n\}$$

(card$\mathbf{q}$ означает количество элементов множества $\mathbf{q}$).

Пусть $\mathbf{y}_q$ есть вектор размерности card$\mathbf{q}$ в котором содержатся компоненты $\mathbf{y}$ только с номерами из множества $\mathbf{q}$; пусть $\mathbf{F}_q$ есть матрица, которая содержит строки только с номерами из множества $\mathbf{q}$.

Используя вектор $\mathbf{y}_q$, мы также можем вычислить OLS-оценку вектора $\boldsymbol{\theta}$, т. е.

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{F}_q^T \mathbf{F}_q\right)^{-1} \mathbf{F}_q^T \mathbf{y}_q. \quad (13)$$

Тогда ковариационная матрица вектора $\hat{\boldsymbol{\theta}}$ следует выражению

$$\mathbf{V}_{\hat{\theta}} = \sigma^2 \left(\mathbf{F}_q^T \mathbf{F}_q\right)^{-1},$$

а информационная матрица Фишера –

$$\mathbf{M}_q = \sigma^{-2}\mathbf{F}_q^T \mathbf{F}_q. \quad (14)$$

В случае $\mathbf{V}_\varepsilon \neq \sigma^2\mathbf{I}$ формулы (13) и (14) изменятся аналогично формулам (4) и (6) (см. формулы (7) и (9)). Для фиксированного значения card$\mathbf{q} = k$ ($k < n$) выберем такое подмножество **q,** которое доставляет максимум некоторого выпуклого функционала от информационной матрицы $\mathbf{M}_q$.

Такая формулировка оказывается вполне актуальной, когда ставится задача экономии средств наблюдения с минимальной потерей точности воспроизведения всех компонент **y**. Действительно, при подстановке OLS-оценки вектора $\boldsymbol{\theta}$ по формуле (13) в модель (3) мы заинтересованы в том, чтобы все значения отклика воспроизводились с минимальной потерей точности, что, в принципе, и должна обеспечивать минимизация дисперсии OLS-оценки вектора $\boldsymbol{\theta}$.

Сравнение двух формулировок. Покажем, что альтернативная формулировка сводится к традиционной путем наложения соответствующих ограничений. Действительно, пусть $p_i = 1/k$ при условии $i \in \mathbf{q}$ и $p_i = 0$ при условии $i \notin \mathbf{q}$, что не нарушает основных требований к плану (11). Тогда информационная матрица по традиционной формулировке задачи OEP следует выражению

$$\mathbf{M}(\mathbf{p}) = \mathbf{F}^T\mathbf{P}\mathbf{F} = \frac{1}{k}\mathbf{F}_q^T\mathbf{F}_q = \frac{\sigma^2}{k}\mathbf{M}_q,$$

где $\mathbf{M}_q$ определяется формулой (14), т. е. является информационной матрицей Фишера.

Однако при этом мы приходим к оценке параметров, выраженной формулой (13):

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{F}^T\mathbf{P}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{P}\mathbf{y} =$$

$$= k\left(\mathbf{F}_q^T\mathbf{F}_q\right)^{-1}\frac{1}{k}\mathbf{F}_q^T\mathbf{y} = \left(\mathbf{F}_q^T\mathbf{F}_q\right)^{-1}\mathbf{F}_q^T\mathbf{y}.$$

Данная оценка не вступает в противоречие с теоремой Гаусса – Маркова.

Очевидно, что использование, например, *D*-критерия для решения поставленной таким образом задачи, возможно лишь при $k \geq m$, когда

$$\det\left(\mathbf{F}_q^T\mathbf{F}_q\right) \neq 0.$$

Поэтому при альтернативной формулировке наиболее удобными оказываются такие критерии, как *A*-критерий, *E*-критерий (см. выше) и т. п. Однако согласно замечанию Дж. Себера, высказанному в монографии [5], критерии, которые не зависят от масштаба величин, выступают предпочтительными. Как уже было отмечено выше, SIQ от масштаба переменных не зависит (см. формулу (2)). С другой стороны, использование SIQ подразумевает, что параметры регрессии имеют стохастическую природу. Ниже будет показано, что в альтернативной формулировке задачи OEP, SIQ можно использовать и в случае, когда параметры регрессии (3) являются постоянными величинами.

Свойства SIQ как критерия OEP в альтернативной формулировке

Из формул (10) и (14) следует, что SIQ, содержащееся в компонентах вектора $\mathbf{y}_q$ относительно компонент $\boldsymbol{\theta}$, имеющих стохастическую природу, выражается как

$$I\left(\mathbf{y}_q, \boldsymbol{\theta}\right) = \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{M}_q\mathbf{V}_\theta\right). \quad (15)$$

Формула (15) позволяет определить SIQ, содержащееся в векторе $\mathbf{y}_q$ относительно вектора параметров регрессии $\boldsymbol{\theta}$, и в случае, когда параметры регрессии (3) не обладают стохастической природой, т. е. являются постоянными величинами.

Определение. *Если в модели (3) параметры есть постоянные величины, то* $I(\mathbf{y}_q, \boldsymbol{\theta})$ *выражается формулой*

$$I\left(\mathbf{y}_q, \boldsymbol{\theta}\right) = \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{M}_q\mathbf{V}_{\hat{\theta}}\right) = \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{M}_q\mathbf{M}^{-1}\right). \quad (16)$$

Комментарии.

1. В этом определении мы приравняли $I(\mathbf{y}_q, \boldsymbol{\theta})$ к $I(\mathbf{y}_q, \hat{\boldsymbol{\theta}})$, т. е. свели к случаю, когда оба вектора являются стохастическими, как того требует формула (1).

2. Если $a = 2$ (единицей информации в двоичной системе счисления является бит), то тогда

$$0 \leq I\left(\mathbf{y}_q, \boldsymbol{\theta}\right) \leq \frac{m}{2} \quad (m = \dim\boldsymbol{\theta}).$$

Далее на подмножествах вида **q** (см. выше) определим следующий функционал:

$$\Phi(\mathbf{q}) = I(\mathbf{y}_q, \boldsymbol{\theta}).$$

Тогда имеет место следующая теорема.

Теорема 2. *Функционал* $\Phi(\mathbf{q})$ *обладает свойством выпуклости, т. е. выполняется неравенство*

$$\Phi(\alpha\mathbf{p} + \beta\mathbf{q}) \geq \alpha\Phi(\mathbf{p}) + \beta\Phi(\mathbf{q}),$$

где $\alpha, \beta \geq 0$; $\alpha + \beta = 1$; $\mathbf{p}, \mathbf{q} \subseteq \{1, 2, ..., n\}$.

Доказательство. Преобразуем формулу (15) следующим образом:

$$I\left(\mathbf{y}_q,\boldsymbol{\theta}\right)=\frac{1}{2}\log_a \det\left(\mathbf{V}_\theta^{-1}+\mathbf{M}_q\right)+ \\ +\frac{1}{2}\log_a \det \mathbf{V}_\theta. \quad (17)$$

Докажем требуемое теоремой неравенство для функционала $\Psi(\mathbf{q})$:

$$\Psi(\mathbf{q})=\frac{1}{2}\log_a\det\left(\mathbf{V}_\theta^{-1}+\mathbf{M}_q\right). \quad (18)$$

Пусть, как и предполагается, $\mathbf{p},\mathbf{q}\subseteq\{1,2,\ldots,n\}$. Тогда, согласно определению $\Psi(\mathbf{q})$, имеем:

$$\Psi\left(\alpha\mathbf{p}+\beta\mathbf{q}\right)=$$

$$=\frac{1}{2}\log_a\det\left(\mathbf{V}_\theta^{-1}+(\alpha\mathbf{M}_p+\beta\mathbf{M}_q)\right)=$$

$$=\frac{1}{2}\log_a\det\left((\alpha+\beta)\mathbf{V}_\theta^{-1}+(\alpha\mathbf{M}_p+\beta\mathbf{M}_q)\right)=$$

$$=\frac{1}{2}\log_a\det\left(\alpha(\mathbf{V}_\theta^{-1}+\mathbf{M}_p)+\right.$$

$$\left.+\beta(\mathbf{V}_\theta^{-1}+\mathbf{M}_q)\right)\geq^{(**)}$$

$$\geq\frac{1}{2}\log_a\left(\det{}^\alpha(\mathbf{V}_\theta^{-1}+\mathbf{M}_p)\times\right.$$

$$\left.\times\det{}^\beta(\mathbf{V}_\theta^{-1}+\mathbf{M}_q)\right)=$$

$$=\alpha\frac{1}{2}\log_a\det\left(\mathbf{V}_\theta^{-1}+\mathbf{M}_p\right)+$$

$$+\beta\frac{1}{2}\log_a\det\left(\mathbf{V}_\theta^{-1}+\mathbf{M}_q\right)=$$

$$=\alpha\Psi\left(\mathbf{p}\right)+\beta\Psi\left(\mathbf{q}\right).$$

Здесь неравенство (* *) следует из общего неравенства

$$\det(\alpha\mathbf{A}+\beta\mathbf{B})\geq\det\mathbf{A}^\alpha\cdot\det\mathbf{B}^\beta;$$

общее неравенство выполняется при указанных ограничениях относительно α и β в случае симметричных, положительно определенных матриц $\mathbf{A}$ и $\mathbf{B}$ (доказательство представлено в книге [9]), каковыми являются матрицы $\mathbf{V}_\theta^{-1}$, $\mathbf{M}_q$ и, следовательно, их сумма.

Таким образом, мы имеем неравенство

$$\Psi\left(\alpha\mathbf{p}+\beta\mathbf{q}\right)\geq\alpha\Psi\left(\mathbf{p}\right)+\beta\Psi\left(\mathbf{q}\right).$$

Прибавляя к обеим частям этого неравенства член

$$\frac{1}{2}\log_a\det\mathbf{V}_\theta,$$

получаем следующее неравенство:

$$\Psi\left(\alpha\mathbf{p}+\beta\mathbf{q}\right)+\frac{1}{2}\log_a\det\mathbf{V}_\theta\geq$$

$$\geq\alpha\Psi\left(\mathbf{p}\right)+\beta\Psi\left(\mathbf{q}\right)+(\alpha+\beta)\frac{1}{2}\log_a\det\mathbf{V}_\theta.$$

С учетом формулы (17) имеем требуемое неравенство для функционала $\Phi(\mathbf{q})$.

Теорема 2 полностью доказана.

Замечание. Из формулы (16) и доказательства теоремы 2 видно, что функционал

$$\Phi(\mathbf{q}) = I\,(\mathbf{y}_q,\,\boldsymbol{\theta})$$

обладает свойством выпуклости и в случае постоянных параметров регрессии, когда количество информации определено формулой (16).

Возвратимся к случаю стохастических параметров регрессии. Здесь имеет место следующая, почти очевидная теорема, связывающая $I\,(\mathbf{y}_q,\,\boldsymbol{\theta})$ с D-критерием (определителем от информационной матрицы, см. выше). Следует отметить, что теория ОЕР традиционно не рассматривает случая стохастических параметров, однако и не содержит на него прямого запрета.

Теорема 3. *При стохастических параметрах регрессии, имеющих достаточно большой масштаб, максимизация* SIQ, *содержащегося в компонентах отклика относительно параметров регрессии, эквивалентна D-критерию.*

Доказательство. Пусть $\mathbf{p},\mathbf{q}\subset\{1,2,\ldots,n\}$ и card$\mathbf{p}$ = card$\mathbf{q}$. Из формулы (17) следует, что и здесь есть возможность рассматривать функционал $\Psi(\mathbf{q})$, определенный формулой (18), вместо функционала

$$\Phi(\mathbf{q}) = I\,(\mathbf{y}_q,\,\boldsymbol{\theta}).$$

Вполне очевидно, что при достаточно большом масштабе стохастических параметров регрессии, а, соответственно, и малости элементов матрицы $\mathbf{V}_\theta^{-1}$, мы непременно будем иметь эквивалентность следующих неравенств:

$$\det\left(\mathbf{M}_p+\mathbf{V}_\theta^{-1}\right)>\det\left(\mathbf{M}_q+\mathbf{V}_\theta^{-1}\right)\Leftrightarrow$$

$$\Leftrightarrow\det\left(\mathbf{M}_p\right)>\det\left(\mathbf{M}_q\right),$$

что, с учетом монотонности логарифма, и означает требуемое.

Теорема 3 полностью доказана.

Следует отметить, что для основания к применению SIQ в альтернативной формулировке задачи ОЕР как в случае стохастических, так и в случае постоянных параметров регрессии (3) вполне достаточно формул (15), (16) и утверждения теоремы 2.

Если решать задачу ОЕР в альтернативной формулировке для последовательных значений k (см. выше) от $k = 1$ до $k = n$, то мы получим последовательность компонент отклика в порядке убывания информационного вклада, а также возрастающую последовательность значений количества информации I_j. Последние удобно представлять, исключая зависимость от выбора основания логарифма a, в следующем виде:

$$I_j := (I_j/I(\mathbf{y},\boldsymbol{\theta}))100\% \ (j = 1,2, \ldots, n),$$

где $I(\mathbf{y},\boldsymbol{\theta})$ – полное количество информации, когда $\mathbf{q} = \{1, 2, ..., n\}$.

Аргументы в пользу целесообразности постановки и решения такой задачи даны в конце следующего раздела.

Использование регрессии на базис главных компонент в информационном упорядочивании и учет ценности информации при наличии пропусков данных

В большинстве практических задач мы, в лучшем случае, имеем лишь выборку наблюдений некоторого вектора

$$\{\mathbf{y}_j, j = 1,2, \ldots, N\}.$$

Обычно такую выборку представляют в виде так называемой выборочной матрицы $\mathbf{Y}$ размерности $n \times N$. Таким образом, столбцы матрицы $\mathbf{Y}$ есть реализации случайного вектора $\mathbf{y}$. Предположим, что вектор $\mathbf{y}$ подчиняется многомерному нормальному распределению с параметрами $\boldsymbol{\theta}_y$ и $\mathbf{V}_y$ т. е.

$$\mathbf{y} \sim N(\boldsymbol{\theta}_y, \mathbf{V}_y).$$

Вычислим оценки параметров распределения

$$\widehat{\boldsymbol{\theta}}_y = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_j,$$

$$\widehat{\mathbf{V}}_y = \frac{1}{N-1}\sum_{j=1}^{N} (\mathbf{y}_j - \widehat{\boldsymbol{\theta}}_y)(\mathbf{y}_j - \widehat{\boldsymbol{\theta}}_y)^T. \quad (19)$$

Далее вычислим ортогональную матрицу $\mathbf{Q}$ такую, что выполняется равенство

$$\mathbf{Q}^T \widehat{\mathbf{V}}_y \mathbf{Q} = \widehat{\boldsymbol{\Lambda}} = \mathrm{diag}(\widehat{\lambda}_1, \widehat{\lambda}_2, ..., \widehat{\lambda}_n)$$

$$\text{и } \widehat{\lambda}_1 \ge \widehat{\lambda}_2 \ge ... \ge \widehat{\lambda}_n.$$

Определим вектор размерности m ($m < n$) в виде

$$\mathbf{z} = \mathbf{Q}^T_{(m)}(\mathbf{y} - \widehat{\boldsymbol{\theta}}_y),$$

где матрица $\mathbf{Q}_{(m)}$ содержит только первые m столбцов матрицы $\mathbf{Q}$.

Компоненты вектора $\mathbf{z}$ обладают свойством взаимной независимости, называются выборочными главными компонентами и интерпретируются как скрытые факторы.

Матрица

$$\widehat{\mathbf{V}}_z = \widehat{\boldsymbol{\Lambda}}_{(m)} = \mathrm{diag}(\widehat{\lambda}_1, \widehat{\lambda}_2, ..., \widehat{\lambda}_m)$$

– суть оценка матрицы взаимных ковариаций компонент $\mathbf{z}$.

Слово *выборочные* обычно опускается, но при этом следует помнить, что истинные главные компоненты мы бы имели лишь при известных параметрах распределения $\boldsymbol{\theta}_y$ и $\mathbf{V}_y$.

Наиболее корректный выбор размерности вектора $\mathbf{z}$ ($m = \mathrm{dim}\mathbf{z}$) связан с проверкой статистической гипотезы

$$\mathrm{H}: \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_m \ge \lambda_{m+1} = \lambda_{m+2} = \ldots = \lambda_n.$$

Однако существуют и относительно простые способы определения m [11].

Перейдем к центрированным значениям вектора $\mathbf{y}$ т. е.

$$\mathbf{y} := \mathbf{y} - \widehat{\boldsymbol{\theta}}_y,$$

и рассмотрим регрессию

$$\mathbf{y} = \mathbf{F}\mathbf{z} + \boldsymbol{\varepsilon}, \quad (20)$$

где $\mathbf{F} = \mathbf{Q}_{(m)}$

Предположим, что вектор остатков $\boldsymbol{\varepsilon}$ обладает теми же свойствами, что и в модели (3), т. е.

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

Модель (20) отличается от модели (3) прежде всего тем, что параметры регрессии в первом случае (в модели (20)), вообще говоря, стохастические. Эта модель позволяет, как отмечалось выше, решать различные практические и исследовательские задачи, связанные с информационным упорядочиванием, осуществляя последовательный отбор компонент отклика с максимизацией SIQ на каждом шаге отбора, т. е.

$$I\left(\mathbf{y}_q, \mathbf{z}\right) = \frac{1}{2}\log_a \det\left(\mathbf{I} + \mathbf{M}_q \widehat{\boldsymbol{\Lambda}}_{(m)}\right)$$

или какого-либо из классических критериев, например tr$\mathbf{M}_q$ [12].

Кроме того, модель (20) допускает различные подходы к определению ценности информации при наличии пропусков из-за нестабильности источников информации.

Пусть исходная выборочная матрица $\mathbf{Y}$ содержит пропуски по отдельным компонентам. Построим матрицу $\mathbf{N}$ той же размерности, что и $\mathbf{Y}$ ($n \times N$), в которой каждому пропущенному измерению будет соответствовать 0, а наличествующему (не пропущенному) – 1. Тогда каждый элемент n_{ij} (i = 1,2, …, n; j = 1,2, …, N) матрицы $\mathbf{N}$ равен нулю или единице. Пропуски в матрице $\mathbf{Y}$ также заполним нулями. Вычислим средние значения компонент вектора $\mathbf{y}$ по формуле

$$\overline{y}_i = \sum_{j=1}^{N} y_{ij} \Bigg/ \sum_{j=1}^{N} n_{ij} \;\; (i = 1, 2, \ldots, n)$$

и перейдем к центрированной матрице $\tilde{\mathbf{Y}}$, где

$$y_{ij} := y_{ij} - \overline{y}_i, \text{ если } n_{ij} \neq 0,$$

а значения, которым соответствует n_{ij} = 0, оставим без изменений, т. е. нулями.

Алгоритм 1. Теперь, если по каким-либо причинам исследователь предпочитает стабильные источники информации перед нестабильными, то оценку ковариационной матрицы компонент вектора $\mathbf{y}$ следует вычислять по формуле

$$\tilde{\mathbf{V}}_y = \frac{1}{N} \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^T. \quad (21)$$

При отсутствии пропусков и делителе, равном N – 1, оценка (21) совпадает с оценкой (19). После вычисления матрицы $\mathbf{F} = \mathbf{Q}_{(m)}$ и осуществления информационного упорядочивания компонент вектора $\mathbf{y}$, номера компонент отклика, соответствующие нестабильной информации (пропускам), с большой вероятностью окажутся в конце этой последовательности (см. выше, конец предыдущего раздела). Оценки отдельных элементов матрицы взаимных ковариаций по формуле (21) при наличии пропусков подобны оценкам автоковариаций, которые предложили Г. Дженкинс и Д. Ваттс в своей известной книге [13]. В этих оценках делитель, независимо от величины сдвига, а, соответственно, и от числа слагаемых, равен длине ряда N, что, как показано в работе [13], не приводит к завышению оценок длины волны. В нашем случае такой постоянный делитель снизит ценность нестабильной информации.

Алгоритм 2. Однако может оказаться так, что именно нестабильные источники информации представляют для исследователя наибольшую ценность. В этом случае элементы ковариационной матрицы следует оценивать по формуле

$$\left[\tilde{\mathbf{V}}_y\right]_{ij} = \left[\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T\right]_{ij} \Big/ \left[\mathbf{N}\mathbf{N}^T\right]_{ij} \quad (i, j = 1, 2, \ldots, n) \quad (22)$$

при $[\mathbf{N}\mathbf{N}^T]_{ij} \neq 0$ и $[\tilde{\mathbf{V}}_y]_{ij} = 0$ в противном случае.

Здесь $[*]_{ij}$ – оператор взятия элемента матрицы из i-й строки и из j-го столбца.

Формула (22) обеспечивает равенство делителя количеству ненулевых членов соответствующей суммы. При таком вычислении оценки ковариационной матрицы номера компонент отклика с пропусками данных могут оказаться и в начале информационно-упорядоченной последовательности [14].

Здесь следует напомнить, что вычисление SIQ связано со смещенным оцениванием и при гауссовском распределении имеет геометрическую интерпретацию (см. формулу (2)). Поэтому в формуле (19) можно было использовать и делитель, равный N, а в модели (20) следует использовать смещенную оценку σ^2 [11].

Замечание. Единственным серьезным критическим замечанием относительно SIQ можно считать следующее: SIQ как мера количества информации никак не учитывает субъективную ценность информации для потребителя.

Предложенные выше алгоритмы в какой-то мере решают эту проблему, поскольку их можно использовать и в самом общем случае при вычислении SIQ по формуле (1). При этом, вычисляя оценки матриц $\mathbf{V}_\xi$, $\mathbf{V}_\eta$, и $\mathbf{V}_{\xi\eta}$, мы должны построить и использовать соответствующие матрицы наличия данных: $\mathbf{N}_\xi$ и $\mathbf{N}_\eta$ (см. выше). Очевидно, что в случае высокой ценности нестабильной информации (алгоритм 2) при вычислении оценки $\mathbf{V}_{\xi\eta}$ в качестве делителей (см. формулу (22)) следует использовать соответствующие элементы матрицы $\mathbf{N}_\xi \mathbf{N}_\eta^T$. Комбинируя алгоритмы 1 и 2, можно реализовывать любые субъективные предпочтения относительно ценности информации по отдельным компонентам исследуемых векторов.

Прежде чем перейти к обсуждению и подведению итогов, отметим, что использование регрессии на базис главных компонент, а также выполнение на основе этой регрессии информационного упорядочивания (см. окончание предыдущего раздела), позволило автору решить ряд прикладных задач. Перечислим некоторые, наиболее важные из них:

оптимальное комплексирование данных спутниковых и наземных наблюдений [12];

выявление ядра коррупционной структуры [14];

определение зон, связанных с динамической неустойчивостью циркуляции атмосферы [15];

оптимизация экологического мониторинга [16];

оптимизация контроля микроэлектронного производства [17].

Достаточно широкий диапазон применимости такого подхода к решению прикладных задач, несомненно, подтверждает практическую значимость данной работы.

Обсуждение основных результатов проведенного исследования

Обозревая ситуацию в целом, следует прежде всего отдать должное исследованиям О.А. Покровского (см. раздел «Введение» в настоящей статье и его работу [3]). Он обнаружил, что и случай стохастических параметров регрессии, и альтернативная постановка задачи OEP являются актуальными как для исследовательских, так и для практических задач. О.А. Покровскому принадлежит (как было отмечено во введении) приоритет в использовании SIQ в задаче оптимизации спутникового зондирования атмосферы (см. также «Введение» и работу [3]). Однако тема связи SIQ с информационной матрицей Фишера (см. формулы (10), (15), (16)), равно как и другие вопросы, рассмотренные в данной работе, им не разбирались.

Относительно традиционной постановки задачи теории OEP отметим, что эта постановка породила ряд интересных математических результатов, связанных с тем, что такой подход позволяет подключить к решению задачи средства математического анализа при переходе к непрерывным планам. Поэтому теоретически обоснованный ответ на уточняющее замечание в разделе «Традиционная и альтернативная формулировки задачи оптимального планирования эксперимента» настоящей статьи сделал бы эти результаты более ценными для практических задач.

В альтернативной же формулировке задачи OEP, несмотря на общую простоту и прозрачность, остается невыясненным следующий вопрос. При фиксированном значении

$$\mathrm{card}\mathbf{q}=k_* \ (k_* < n),$$

очевидно, существует собственное подмножество $\mathbf{q}$, которое дает максимум критерия оптимальности

$$\max\mathrm{crit}\mathbf{M}_q \ (\mathrm{crit}\mathbf{M}_q = \mathrm{tr}\mathbf{M}_q,\ I(\mathbf{y}_q,*) \text{ и т.п.}).$$

Однако нет уверенности, как и строгого доказательства, что, решая эту задачу последовательным наращиванием значения k от 1 до k_* (см. окончание раздела «Свойства SIQ как критерия OEP в альтернативной формулировке») с максимизацией выбранного критерия на каждом шаге последовательного отбора, мы придем именно к этому значению $\max\mathrm{crit}\mathbf{M}_q$.

Аналогичная ситуация имеет место в задаче построения оптимальной регрессии [5]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k^*} x_{k^*} + \varepsilon,$$

когда при фиксированном количестве k^*, из множества потенциальных регрессоров $\mathbf{x}(\mathrm{card}\mathbf{x} > k^*)$ мы выбираем регрессоры, которые обеспечивают максимум значения коэффициента детерминации R^2.

Здесь задачу также можно решать последовательным отбором регрессоров, максимизируя на каждом шаге коэффициент детерминации R^2 [5], и точно так же остается открытым вопрос о том, придем ли мы таким путем к максимальному значению R^2.

Достаточно правдоподобным выглядит предположение, что при нормальном распределении мы можем рассчитывать на положительный ответ. На практике же эту проблему в обоих случаях обычно решают посредством процедуры включения и отбрасывания, останавливая процесс при обнаружении повторных шагов [5].

Заключение

Проведенное нами исследование позволило получить следующие результаты.

1. Доказано, что SIQ, содержащееся в компонентах отклика относительно стоха-

стических параметров,

во-первых, непосредственно связано с информационной матрицей Фишера;

во-вторых, является выпуклым функционалом на множестве компонент отклика;

в-третьих, при достаточно большом масштабе параметров эквивалентно использованию *D*-критерия в задаче планирования эксперимента.

2. Определено SIQ относительно постоянных параметров регрессии, где SIQ также является выпуклым функционалом на множестве компонент отклика.

3. Проанализировано соотношение традиционной и альтернативной постановок задачи планирования эксперимента.

4. Рассмотрено использование регрессии на базис главных компонент в задаче информационного упорядочивания данных.

5. Предложены алгоритмы, учитывающие субъективную ценность информации при наличии частичных пропусков данных.

И наконец, отметим, что перечисленные выше работы [3, 12, 14 – 17] в полной мере демонстрируют целесообразность использования модели регрессии со стохастическими параметрами при альтернативной постановке задачи OEP, а, в частности, работы [14 – 17] со всей полнотой демонстрируют практическую значимость результатов, полученных в настоящей статье.

СПИСОК ЛИТЕРАТУРЫ

1. **Shannon C.E., Weaver W.** A mathematical theory of communication, Urbana University Press, Chicago, 1949. (Перевод на русский язык см. в Сборнике переводов «Теория передачи электрических сигналов при наличии помех» под ред. С.А. Железнова. М.: Изд-во иностранной литературы, 1953. 288 с.)

2. **Гельфанд И.М., Яглом А.М.** О вычислении количества информации о случайной функции, содержащейся в другой такой функции // Успехи математических наук. 1957. Т. 12. № 1 (73). С. 3–52.

3. **Покровский О.М.** Об оптимальных условиях косвенного зондирования атмосферы // Известия АН СССР. Физика атмосферы и океана. 1969. Т. 5. № 12. С. 1324–1326.

4. **Кендал М.Дж., Стьюарт А.** Статистические выводы и связи. М.: Наука, 1973. 899 с.

5. **Себер Дж.** Линейный регрессионный анализ. М.: Мир, 1980. 456 с.

6. **Пичугин Ю.А.** О связи количества информации по Шеннону с информационной матрицей Фишера в задаче планирования регрессионного эксперимента // Межвуз. сб. науч. трудов «Информатика – исследования и инновации». Вып. 3. СПб.: РГПУ им. А.И. Герцена, 1999. С. 32–36.

7. **Пичугин Ю.А.** Замечания к отбору данных в задачах, связанных с линейной множественной регрессией // Заводская лаборатория. Диагностика материалов. 2002. № 5. С. 61–62.

8. Математическая теория планирования эксперимента. Под ред. С.М. Ермакова. М.: Наука. Гл. ред. физ.-мат. лит-ры, 1983. 392 с.

9. **Ермаков С.М., Жиглявский А.А.** Математическая теория оптимального эксперимента. М.: Наука. Гл. ред. физ.-мат. лит-ры, 1987. 320 с.

10. **Фон Нейман Дж., Моргенштерн О.** Теория игр и экономическое поведение. М.: Наука. Гл. ред. физ.-мат. лит-ры, 1970. 707 с.

11. **Пичугин Ю.А.** Замечания к использованию главных компонент в математическом моделировании // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2018. Т. 11. № 3. С. 74–89.

12. **Пичугин Ю.А., Покровский О.М.** О методе комплексирования наземной и спутниковой метеорологической информации // Исследование Земли из космоса. 1992. № 6. С. 25–31.

13. **Дженкинс Г., Ваттс Д.** Спектральный анализ и его приложения. Выпуск 1. М.: Мир, 1971. 316 с.; Выпуск 2. М.: Мир, 1972. 287 с.

14. **Pichugin Yu. A., Malafeyev O.A., Rylow D., Zaitseva I.** A statistical method for corrupt agents detection // International Conference on Numerical Analysis and Applied Mathematics (ICNAAM 2017). AIP Conf. Proc. 1978. Pp. 100014-1–100014-4.

15. **Пичугин Ю.А.** География динамической неустойчивости циркуляции атмосферы в Северном полушарии (моделирование и анализ) // Известия Русского географического общества. 2005. Т. 137. № 3. С. 12–16.

16. **Пичугин Ю.А.** Экологический мониторинг и методы многомерной математической статистики // Астраханский вестник экологического образования. 2012. № 2. С. 101–105.

17. **Гусман Ю.А., Пичугин Ю.А., Смирнов А.О.** Информативность по Шеннону в контроле микроэлектронной продукции // Вопросы радиоэлектроники. 2018. № 10. С. 6–10.

Статья поступила в редакцию 22.05.2019, принята к публикации 14.06.2019.

СВЕДЕНИЯ ОБ АВТОРЕ

ПИЧУГИН Юрий Александрович – *доктор физико-математических наук, профессор Института инноватики и базовой магистерской подготовки Санкт-Петербургского государственного университета аэрокосмического приборостроения, Санкт-Петербург, Российская Федерация.*

190000, Российская Федерация, г. Санкт-Петербург, Большая Морская ул., 61.

yury-pichugin@mail.ru

REFERENCES

1. **Shannon C.E., Weaver W.,** A mathematical theory of communication, Urbana University Press, Chicago, 1949.

2. **Gelfand I.M., Yaglom A.M.,** Computation of the amount of information about a stochastic function contained in another such function, Uspekhi Matematichtskih Nauk. 12 (1(73)) (1957) 3–52.

3. **Pokrovsky O.M.,** Ob optimalnyh usloviyah kosvennogo zondirovaniya atmosfery [On the optimum conditions of the atmosphere indirect sensing], Reports of AS USSR, Physics of Atmosphere and Ocean. 5 (12) (1969) 1324–1326.

4. **Kendall M.G., Stuart A.,** The advanced theory of statistics, Vol. 2: Inference and relationship, Charles Griffin & Company limited, London, 1958 –1966.

5. **Seber G.A.F.,** Linear regression analysis, John Wiley & Sons, New York, London, Sydney, Toronto, 1977.

6. **Pichugin Yu.A.,** O svyazi kolichestva informatsii po Shennonu s informatsionnoy matritsey Fishera v zadache planirovaniya regressionnogo experimenta [On the relation of the amount of Shannon's information with Fisher's information matrix in the problem of regression experiment planning], The Issue of Inter-University Scientific Articles "Informatics – Research and Innovation", No. 3, Saint-Petersburg, Herzen University (1999) 32–36.

7. **Pichugin Yu.A.,** Notes on selecting data in the tasks associated with linear multiple regression, Industrial Laboratory. Diagnostics of materials. 5(2002) 61–62.

8. Matematicheskaya teoriya planirovaniya experimenta [Mathematical theory of experiment planning], Ed. S.M. Ermakov, Nauka, Moscow, 1983.

9. **Ermakov S.M., Zhiglyavskiy A.A.,** Matematicheskaya teoriya optimalnogo experimenta [Mathematical theory of optimal experiment], Nauka, Moscow, 1987.

10. **Von Neumann J., Morgenstern O.,** Theory of games and economic behavior, Princeton University Press, Princeton, 1953.

11. **Pichugin Yu.A.,** Notes on using the principal components in the mathematical simulation, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 11 (3) (2018) 74–89.

12. **Pichugin Yu.A., Pokrovsky O.M.,** On the method of complexing the ground and spaceborne meteorologic information, Remote Sensing. (6) (1992) 25–31.

13. **Jenkins G.M., Watts D.G.,** Spectral analysis and its applications, Holden-Day, San Fracisco – Cambridge – London – Amsterdam (1969).

14. **Pichugin Yu.A., Malafeyev O.A., Rylow D., Zaitseva I.,** A statistical method for corrupt agents detection, International Conference on Numerical Analysis and Applied Mathematics (ICNAAM 2017), AIP Conf. Proc. (1978) 100014-1–100014-4.

15. **Pichugin Yu.A.,** Geografiya dinamicheskoy neustoychivosti tsirkulyatsii atmosfery v Severnom polusharii (modelirovanie i analiz) [Geography of dynamic instability of atmospheric circulation in the Northern hemisphere (simulation and analysis)], Reports of Russian Geographical Society. 137 (3) (2005) 12–16.

16. **Pichugin Yu.A.,** Ekologicheskiy monitoring i metody mnogomernoy matematicheskoy statistiki [Environmental monitoring and methods of multivariate mathematical statistics], Astrakhan Bulletin of Environmental Education (2) (2012) 101–105.

17. **Gusman Yu.A., Pichugin Yu.A., Smirnov A.O.,** Shannon's informative value for control of microelectronic productions, Voprosy Radioelektroniki [Radio Electronics Issues]. (10) (2018) 6–10.

Received 22.05.2019, accepted 14.06.2019.

THE AUTHOR

PICHUGIN Yury A.
Saint-Petersburg State University of Aerospace Instrumentation
61 Bolshaya Morskaya St., St. Petersburg, 190000, Russian Federation
yury-pichugin@mail.ru