Научный вестник НГТУ. – 2013. – № 1(50)

УДК 519.233.22

Устойчивое оценивание параметров модели по многомерным неоднородным неполным данным*

Д.В. ЛИСИЦИН

В работе развивается теория оптимального оценивания неизвестных параметров статистической модели по многомерным неоднородным данным. Данные представляются в числовой форме, но не обязательно являются числовыми, например, могут быть качественными или разнотипными. Оценки обеспечивают устойчивость к отклонению распределения наблюдений от постулированного. В основе теории лежит использование подхода Ф. Хампеля, связанного с функцией влияния, и подхода А.М. Шурыгина, связанного с байесовским точечным засорением распределения. Отдельное внимание уделяется случаю неполных данных, получены условия, при которых механизм порождения пропусков можно игнорировать.

Ключевые слова: оценивание параметров, робастность, функция влияния, неоднородные данные, неполные данные, разнотипные данные.

ВВЕДЕНИЕ

При изучении сложных объектов их состояние может описываться большим числом характеристик. Нередко эти характеристики измерены в разных шкалах, тогда наблюдения за изучаемым объектом являются разнотипными [1]. Часто выделяют непрерывные, дихотомические, счетные, номинальные, порядковые и смешанные (полунепрерывные) переменные.

Если в наблюдениях характеристик не всегда можно зафиксировать их значения, то данные являются неполными, содержат пропуски. С такой ситуацией часто сталкиваются при моделировании многомерных данных [2–4]. Однако к задаче моделирования по данным с пропусками может быть сведен более широкий круг задач. Например, при наличии цензурированных выборок может использоваться моделирование разнотипных переменных по неполным данным [2]. При некотором сочетании условий переменная вообще может быть не определена, как, например, в двухчастной модели полунепрерывных данных, где также используется моделирование разнотипных переменных по неполным данным [5].

Если постулирована параметрическая модель, то оценивание ее параметров может производиться по методу максимального правдоподобия [2]. Однако, в условиях отклонения реального распределения переменных от постулированного (модельного), такие оценки часто оказываются неустойчивыми.

Для решения этой проблемы разработаны различные подходы, приводящие к устойчивым (робастным) процедурам [6–10]. Тем не менее, теория робастности применяется, в основном, при моделировании непрерывных случайных величин. Устойчивым методам моделирования дискретных и разнотипных переменных уделяется меньше внимания, за исключением, пожалуй, лишь дихотомических и счетных переменных, а также случая цензурированных выборок.

Для случая пропусков в многомерных данных в [11–13] предлагаются робастные методы оценивания параметров сдвига и масштаба модели многомерной, главным образом, нормаль-

* *Статья получена 19 октября 2012 г.*
Работа выполнена при частичной поддержке гранта Президента РФ (№ МД–2690.2008.9)

ной случайной величины, при этом базируются данные методы на эвристических суждениях. Можно получить робастные оценки параметров модели, используя метод максимального правдоподобия с переходом к стьюдентовскому распределению (см., например, [14]).

В целом следует отметить, что современные подходы к устойчивому моделированию часто носят эвристический характер и являются специфическими для переменных конкретного типа или конкретного типа модели.

В данной работе рассматриваются достаточно общие методы оптимального оценивания неизвестных параметров модели по многомерным неоднородным данным. Данные представляются в числовой форме, но не обязательно являются числовыми. Так, числовая кодировка обычно используется для значений качественных (номинальных, порядковых) переменных. В более общем случае данные могут быть разнотипными.

Получаемые методы оценивания являются устойчивыми к отклонению распределения наблюдений от модельного. В основе представленной теории лежит использование подхода Ф. Хампеля [7], связанного с функцией влияния, и подхода А.М. Шурыгина [8], связанного с байесовским точечным засорением модельного распределения.

Если подход Ф. Хампеля развит на многопараметрический случай для данных достаточно общей природы, то А.М. Шурыгин развил теорию для количественных данных и однопараметрических задач (или задач, приводимых к однопараметрическим), а также для случая оценивания параметров уравнения регрессии. Теория устойчивого оценивания на базе подходов Ф. Хампеля и А.М. Шурыгина для многопараметрических задач в случаях неоднородных количественных (в том числе, счетных) и качественных данных была развита нами в [9, 10, 15–18].

Отдельное внимание в работе уделяется случаю неполных данных, для которого теория робастного оценивания развивается впервые. В том числе получены условия, при которых механизм порождения пропусков можно игнорировать.

Также в работе расширяются концепции введенных А.М. Шурыгиным условно оптимальных и стойких оценок, обсуждаются свойства инвариантности оценок.

1. МОДЕЛЬ, ОЦЕНИВАНИЕ ПАРАМЕТРОВ, ПОКАЗАТЕЛИ КАЧЕСТВА

Пусть независимые n-мерные случайные величины $\zeta_i = (\zeta_{i1},\ldots,\zeta_{in})^T$, $i = 1,...,N$, имеют модельные распределения с функциями $G_i(z_i \mid \phi)$, $z_i \in R^n$, плотностями $g_i(z_i \mid \phi)$ относительно некоторой σ-конечной меры μ и вектором параметров ϕ размера m. Обозначим $\Gamma_i = \{z_i : g_i(z_i \mid \phi) > 0\}$.

M-оценку $\hat{\phi}$ вектора параметров модели ϕ определим по наблюдениям $\tilde{\zeta}_i$, $i = 1,\ldots,N$, случайных величин ζ_i, $i = 1,...,N$, путем решения системы оценочных уравнений [19]

$$\sum_{i=1}^{N} \psi_i\left(\tilde{\zeta}_i, \hat{\phi}\right) = 0, \tag{1}$$

где $\psi_i(\tilde{\zeta}_i, \hat{\phi})$ – векторная оценочная функция, удовлетворяющая условию

$$\mathrm{E}\psi_i(z_i, \phi) = 0, \; i = 1,\ldots,N, \tag{2}$$

E – оператор математического ожидания.

Следствием (2) является асимптотическая несмещенность оценки [6]. Будем рассматривать решения системы уравнений (1), дающие состоятельные оценки.

При некоторых условиях регулярности оценка $\hat{\phi}$ является асимптотически нормальной с разбросом [19]

$$V(\psi) = \mathrm{M}_1^{-1}\,\mathrm{M}_2\left[\mathrm{M}_1^{-1}\right]^T, \tag{3}$$

где $\psi = (\psi_1^T,\ldots,\psi_N^T)^T$, $\mathrm{M}_1 = -\sum_{i=1}^{N}\frac{\partial}{\partial\tilde{\phi}^T}\mathbf{E}\psi_i(z_i,\tilde{\phi})\Big|_{\tilde{\phi}=\phi} = \sum_{i=1}^{N}\int_{R^n}\psi_i(z_i,\phi)\frac{\partial g_i(z_i\,|\,\phi)}{\partial\phi^T}d\mu$ – невырожденная матрица, $\mathrm{M}_2 = \sum_{i=1}^{N}\mathbf{E}\psi_i(z_i,\phi)\psi_i^T(z_i,\phi)$.

В теории робастного оценивания конструируются оценки, имеющие высокое качество не только при постулируемом распределении ошибок, но и при отклонении от него [7–10]. Одним из показателей качества оценки в теории робастности является функция влияния.

Функциональному аналогу уравнения (1)

$$\sum_{i=1}^{N}\int_{R^n}\psi_i(z_i,\phi[G])dG_i(z_i\,|\,\phi) = 0,$$

где $G = (G_1,\ldots,G_N)$, соответствует оценка $\hat{\phi} = \phi[G]$ как функционал от G. Введем модель точечного засорения распределения некоторого i-го наблюдения. Для этого заменим функцию распределения $G_i(z_i\,|\,\phi)$ функцией засоренного распределения $G_i^*(z_i\,|\,\phi) = (1-t)G_i(z_i\,|\,\phi) + t\Delta_{z_i^*}(z_i)$, где $\Delta_{z_i^*}(z_i)$ – вероятностная мера, приписывающая точке z_i^* единичную массу, и обозначим $G^* = (G_1,\ldots,G_{i-1},G_i^*,G_{i+1},\ldots,G_N)$. Тогда функция влияния будет определяться формулой

$$IF_i(z_i^*,\psi) = \lim_{t\to 0}\frac{\phi[G^*]-\phi[G]}{t}.$$

Можно показать, повторяя выкладки из [7], что в рассматриваемом случае для M-оценок при некоторых условиях регулярности функция влияния будет иметь вид

$$IF_i(z_i^*,\psi) = \mathrm{M}_1^{-1}\psi_i(z_i^*,\phi).$$

В соответствии с моделью байесовского точечного засорения засоряющая точка z_i^* является случайной величиной, имеющей в серии выборок плотность $s_i(z_i\,|\,\phi)$, $z_i \in R^n$, относительно меры μ, причем $\Gamma_i \subseteq \Sigma_i$, где $\Sigma_i = \{z_i : s_i(z_i\,|\,\phi) > 0\}$, а функция влияния определена на Σ_i (в отличие от классического случая [7], где она определена на Γ_i).

Показателем качества оценок является матрица

$$U_s(\psi) = \sum_{i=1}^{N}\mathbf{E}_{s_i}\left[IF_i(z_i,\psi)IF_i^T(z_i,\psi)\right] = \sum_{i=1}^{N}\int_{R^n}IF_i(z_i,\psi)IF_i^T(z_i,\psi)s_i(z_i\,|\,\phi)d\mu, \tag{4}$$

где $\mathbf{E}_{s_i}$ – математическое ожидание по плотности $s_i(z_i\,|\,\phi)$.

Показатель (4) можно представить в виде

$$U_s(\psi) = \mathrm{M}_1^{-1}\,\mathrm{M}_{2,s}\left[\mathrm{M}_1^{-1}\right]^T, \tag{5}$$

где $\mathrm{M}_{2,s} = \sum_{i=1}^{N} \int_{R^n} \psi_i(z_i,\phi)\psi_i^T(z_i,\phi)s_i(z_i \mid \phi)d\mu$.

Заметим, что формула (3) может рассматриваться как частный случай формулы (5) при $s_i(z_i \mid \phi) = g_i(z_i \mid \phi)$.

Укажем на одну особенность рассматриваемого подхода. Часто привлекательными свойствами обладают оценки, связанные с показателем (5) при функциях $s_i(z_i \mid \phi)$, которые не удовлетворяют условию нормировки и, более того, могут не быть интегрируемыми функциями [10].

Например, с точки зрения теории робастности представляет интерес случай

$$s_i(z_i \mid \phi) = 1, \tag{6}$$

не требующий, в отличие от общего случая, для получения показателя (5) каких-либо сильных предположений о механизме засорения. Однако для непрерывных случайных величин, имеющих значения на вещественной прямой или полупрямой, и счетных случайных величин функция (6) не является интегрируемой.

Получить интерпретацию показателя (5) в таком случае можно, обобщив результаты из [20].

Будем рассматривать оценочные функции ψ как элементы пространства $L_2(s,W)$ со скалярным произведением

$$\langle \psi, \varphi \rangle = \sum_{i=1}^{N} \int_{R^n} \psi_i^T(z_i,\phi)W\varphi_i(z_i,\phi)s_i(z_i \mid \phi)d\mu = \int_{R^n} \psi^T(z,\phi)S(z,\phi)\otimes W\varphi(z,\phi)d\mu,$$

где $s = (s_1,\ldots,s_N)^T$, $W = W(\phi)$ – некоторая симметричная положительно определенная матрица размера $m \times m$, $S(z,\phi) = \begin{bmatrix} s_1(z \mid \phi) & & 0 \\ & \ddots & \\ 0 & & s_N(z \mid \phi) \end{bmatrix}$, $\otimes$ – символ кронекерова произведения.

Тогда функционал

$$\Psi\left[U_s(\psi)\right] = \mathrm{tr}\left[W U_s(\psi)\right], \tag{7}$$

где tr – след матрицы, можно представить как квадрат $L_2(s,W)$ – нормы функции влияния: $\mathrm{tr}\left[W U_s(\psi)\right] = \|IF\|^2$, где $IF = (IF_1^T,\ldots,IF_N^T)^T$. А u-й элемент v-го столбца (5) есть функционал $\left[U_s(\psi)\right]_{uv} = \langle IF, B_{uv}IF \rangle$, где $B_{uv} = I_N \otimes \left[W^{-1}E_{uv}\right]$, I_N – единичная матрица размера $N \times N$, E_{uv} – матрица размера $m \times m$, u-м элементом v-го столбца которой является единица, а остальные элементы – нулевые.

Поскольку для ряда дальнейших рассуждений необходимо сохранить интерпретацию функций $s_i(z_i \mid \phi)$ как плотностей, будем допускать нарушение свойств нормировки и интегрируемости плотностей $s_i(z_i \mid \phi)$.

2. ОПТИМАЛЬНЫЕ ОЦЕНКИ

Для получения оптимальной оценочной функции будем минимизировать с учетом ограничений (2) некоторый непрерывно дифференцируемый функционал Ψ от матрицы $U_s(\psi)$ (не обязательно вида (7)):

$$\min_{\psi} \Psi\left[U_s(\psi)\right]. \tag{8}$$

В работе [10] показано, что для случая непрерывных случайных величин в области $\Gamma_i = \Sigma_i$ необходимому условию экстремума удовлетворяет функция

$$\psi^*_{s,i}(z_i,\phi) = C\left\{\frac{\partial}{\partial\phi}\ln g_i(z_i \mid \phi) + \beta_i\right\}\frac{g_i(z_i \mid \phi)}{s_i(z_i \mid \phi)}, \tag{9}$$

где $C = C(\phi)$ – невырожденная матрица, с точностью до которой определяются оценочные функции, $\beta_i = \beta_i(\phi)$ – константа, определяемая из условия (2). Легко показать, повторяя выкладки из [10], что для рассматриваемого нами случая справедлив этот же результат (в том числе в области $\Sigma_i \setminus \Gamma_i$, где оценочная функция нулевая). Заметим, однако, что функция (9) не обязательно существует.

Для получения дальнейших результатов ограничимся случаем функционала вида (7).

Поскольку оценочная функция определена с точностью до умножения на невырожденную матрицу, решение задачи (8) не единственно, поэтому введем следующую нормировку:

$$M_1 = I_m. \tag{10}$$

В результате задача получения оптимальной оценочной функции принимает вид

$$\min_{\psi} \|\psi\|^2$$

с ограничениями (2), (10).

Данная задача без учета условий регулярности для функций ψ является выпуклой экстремальной задачей со строго выпуклым оптимизируемым функционалом и квадратичным функционалом Лагранжа, квадратичный член которого положительно определен; в результате единственным решением задачи в области Σ_i является функция (9), если она существует [21, 22]. Матрица C в (9) обеспечивает выполнение условия (10) и с соответствующей ему матрицей множителей Лагранжа, которую обозначим Λ, связана соотношением $C = -W^{-1}\Lambda$ (с точностью до несущественного скалярного положительного сомножителя). Если данное решение удовлетворяет условиям регулярности, то оно является оптимальной оценочной функцией. Заметим, что решение существенно от матрицы W не зависит.

Перейдем к задачам условно оптимального оценивания. Для случая скалярного параметра и однородных непрерывных данных в [8] рассматривалась оптимизация показателя $V(\psi)$ при фиксированном значении показателя $U_1(\psi)$ (показателя $U_s(\psi)$ с плотностью (6)), оптимизация показателя $U_1(\psi)$ при фиксированном значении $V(\psi)$, а также оптимизация показателя, явно задающего компромисс между $V(\psi)$ и $U_1(\psi)$ (также в [8] рассмотрен случай регрессии). В [7] рассматривается задача оптимизации $V(\psi)$ при ограничении сверху на модуль функции влияния, получаемая оценка называется оптимальной робастной (аналогичные задачи рассматриваются и для ряда случаев с многомерным параметром).

Обобщим формулировку задачи условно оптимального оценивания из [8]. Будем рассматривать задачу оптимизации показателя $\Psi\left[U_{s_1}(\psi)\right]$ при ограничениях (2), (10) и

$$\Psi\left[U_{s_2}(\psi)\right] \le D \tag{11}$$

с некоторыми не равными функциями $s_{1i}(z_i \mid \phi)$, $s_{2i}(z_i \mid \phi)$ и $D = D(\phi)$.

Пусть существует функция (9) с матрицей C, обеспечивающей выполнение условия (10), и плотностью

$$s_i(z_i \mid \phi) = s_{1i}(z_i \mid \phi) + \gamma s_{2i}(z_i \mid \phi), \tag{12}$$

где множитель $\gamma = \gamma(\phi) \ge 0$ определяется из условий (11) и $\gamma \cdot \left\{\Psi\left[U_{s_2}(\psi)\right] - D\right\} = 0$, в области $\Sigma_{1,\ldots,2,i}(\gamma) = \left\{z_i : s_{1i}(z_i \mid \phi) + \gamma s_{2i}(z_i \mid \phi) > 0\right\}$. Ограничиваясь рассмотрением функционала (7) и не учитывая условия регулярности для функций ψ, получим выпуклую экстремальную задачу в $L_2(as_1 + bs_2, W)$, где функции $as_{1i}(z_i \mid \phi) + bs_{2i}(z_i \mid \phi)$ положительны в областях $\Sigma_{1,\ldots,2,i}(\gamma)$ и равны нулю вне их, со строго выпуклым оптимизируемым оптимизируемым функционалом и квадратичным функционалом Лагранжа, квадратичный член которого положительно определен; в результате единственным решением задачи является указанная функция (9) [21, 22]. Если данное решение удовлетворяет условиям регулярности, то оно является оптимальной оценочной функцией.

Заметим, что величину D удобно определять, исходя из ограничения на относительную характеристику устойчивости. Такая характеристика имеет вид

$$\mathrm{stb}_s(\psi) = \Psi\left[U_s(\psi_s^*)\right] \Big/ \Psi\left[U_s(\psi)\right]$$

для некоторой плотности s и положительного функционала Ψ [10]. В рассматриваемом случае можно выбирать $0 < \mathrm{stb}_{s_2}(\psi) < 1$.

Условно оптимальные оценки представляют собой компромисс между показателями $\Psi\left[U_{s_1}(\psi)\right]$ и $\Psi\left[U_{s_2}(\psi)\right]$ – при улучшении одного из них второй ухудшается. Однако такое семейство оценок не является полным: могут существовать оценки, которые имеют бóльшие значения показателей $\Psi\left[U_{s_1}(\psi)\right]$ и/или $\Psi\left[U_{s_2}(\psi)\right]$. Такие оценки задаются формулой (9) с плотностью (12) при некоторых отрицательных значениях параметра γ. Они являются дополнительными членами семейства условно оптимальных оценок и имеют свойства устойчивости, выраженные в большей или, наоборот, меньшей степени по сравнению со стандартными условно оптимальными оценками. На практике такие оценки могут быть полезны при неудачном выборе функций s_1, s_2, когда стандартные условно оптимальные оценки либо слишком устойчивы (с неоправданно малой эффективностью), либо, наоборот, недостаточно устойчивы. Например, в рамках теории робастности уместен выбор функции s_1 в виде (6), а функции s_2, равной модельной плотности (см. также [8]); тогда, не выходя за пределы этой теории, можно получить «сверхустойчивые» оценки, если устойчивость оценки, оптимальной при функции s в виде (6), оказывается недостаточной.

Рассмотрим возможный вариант экстремальной задачи, приводящий к получению дополнительных условно оптимальных оценок. Будем оптимизировать показатель $\Psi\left[U_{s_1}(\psi)\right]$ при ограничениях (2), (10) и

$$\Psi\left[U_{s_2}(\psi)\right] = D, \tag{13}$$

где $D = D(\phi) > \Psi\left[U_{s_2}(\psi_{s_1}^*)\right]$, $\Psi\left[U_{s_2}(\psi_{s_1}^*)\right] < \infty$.

Пусть существует функция (9) с матрицей C, обеспечивающей выполнение условия (10), плотностью (12) и множителем $\gamma = \gamma(\phi)$, определяемым из условия (13), в области $\Sigma_{12,i}(\gamma)$, причем выполняются условия $s_{1i}(z_i \mid \phi) + \gamma s_{2i}(z_i \mid \phi) \geq 0$ для всех $z_i \in R^n$ и $\Gamma_i \subseteq \Sigma_{12,i}(\gamma)$. Вновь ограничимся рассмотрением функционала (7) и не будем учитывать условия регулярности для функций ψ. Получим при некоторых условиях гладкую экстремальную задачу в $L_2(as_1 + bs_2, W)$, где функции $as_{1i}(z_i \mid \phi) + bs_{2i}(z_i \mid \phi)$ положительны в областях $\Sigma_{1,\ldots,2,i}(\gamma)$ и равны нулю вне их, с квадратичным функционалом Лагранжа, квадратичный член которого строго положителен; в результате единственным решением задачи является указанная функция (9) [22, 23]. Если данное решение удовлетворяет условиям регулярности, то оно является оптимальной оценочной функцией.

В [10] для случая непрерывных данных рассмотрен ряд формулировок задач оптимизации при явном компромиссе между $V(\psi)$ и $U_s(\psi)$. Данные задачи легко распространяются на рассматриваемый нами случай.

Составим компромиссный показатель

$$U_{s_1}(\psi) + kU_{s_2}(\psi) = U_s(\psi),$$

где $k = k(\phi) > 0$, $s_i(z_i \mid \phi) = s_{1,i}(z_i \mid \phi) + ks_{2,i}(z_i \mid \phi)$. Оптимальные оценочные функции тогда находим, решая задачу (8).

Можно составить компромиссный функционал

$$\frac{1}{\mathrm{stb}_{s_1}(\psi)} + \frac{k}{\mathrm{stb}_{s_2}(\psi)} = \frac{1}{\Psi\left[U_{s_1}(\psi^*_{s_1})\right]}\Psi\left[U_{s_1}(\psi)\right] + \frac{k}{\Psi\left[U_{s_2}(\psi^*_{s_2})\right]}\Psi\left[U_{s_2}(\psi)\right],$$

минимизация которого эквивалентна минимизации функционала $\Psi\left[U_{s_1}(\psi)\right] + \tilde{k}\Psi\left[U_{s_2}(\psi)\right]$ с $\tilde{k} = k\,\Psi\left[U_{s_1}(\psi^*_{s_1})\right] \Big/ \Psi\left[U_{s_2}(\psi^*_{s_2})\right]$. Если функционал Ψ линейный (например, вида (7)), то имеем

$$\Psi\left[U_{s_1}(\psi)\right] + \tilde{k}\Psi\left[U_{s_2}(\psi)\right] = \Psi\left[U_{s_1}(\psi) + \tilde{k}U_{s_2}(\psi)\right] = \Psi\left[U_s(\psi)\right]$$

с

$$s_i(z_i \mid \phi) = s_{1,i}(z_i \mid \phi) + \tilde{k}s_{2,i}(z_i \mid \phi). \qquad (14)$$

В результате вновь приходим к задаче (8).

Аналогично, оптимизация компромиссного функционала

$$\frac{1-\kappa}{\mathrm{stb}_{s_1}(\psi)} + \frac{\kappa}{\mathrm{stb}_{s_2}(\psi)},$$

где $\kappa = \kappa(\phi)$, $0 < \kappa < 1$, при линейном функционале Ψ приводит к задаче (8), (14) с $\tilde{k} = \dfrac{\kappa}{1-\kappa}\Psi\left[U_{s_1}(\psi^*_{s_1})\right] \Big/ \Psi\left[U_{s_2}(\psi^*_{s_2})\right]$.

Если же во введенных компромиссах выбирать $k < 0$, $\kappa < 0$ или $\kappa > 1$ при $s_i(z_i \mid \phi) \geq 0$ для всех $z_i \in R^n$, то будем получать дополнительные члены семейства компромиссных оценок.

Показатель (5) зависит от плотности $s_i(z_i \mid \phi)$, которая неизвестна на практике. Один из путей решения этой проблемы – использование максиминной формулировки [8]. Для случая одномерного параметра она имеет вид

$$s_* = \arg\max_{s \in S} \min_{\psi} U_s(\psi).$$

Если множество S совпадает с множеством модельных плотностей, то оценка с оценочной функцией $\psi^*_{s_*}$ называется стойкой [8]. При векторном параметре стойкие оценки строились в [8], исходя из оптимизации показателя (5), построенного для каждого параметра в отдельности. Однако такой подход может затруднить интерпретацию получаемого решения при совместном оценивании параметров, когда для разных параметров получаются разные плотности s_*.

Полученное выше многомерное решение позволяет ввести многомерные стойкие оценки, когда оптимизируемым является функционал Ψ от матричного показателя (5):

$$s_* = \arg\max_{s \in S} \min_{\psi} \Psi\left[U_s(\psi)\right]. \qquad (15)$$

В рамках такой постановки задачи получается одна плотность s_*, поэтому она более согласована со смыслом решаемой задачи. Однако здесь появляется проблема выбора вида функционала Ψ, так как при разных функционалах получаются разные плотности s_*.

В связи с неоднозначностью выбора параметризации плотности распределения – параметры могут быть введены в модель до некоторой степени произвольно – важным является свойство инвариантности оценок к преобразованиям параметров. У оценок по методу максимального правдоподобия такое свойство известно [19]. В [24] для случая непрерывных данных свойство инвариантности доказано и по отношению к оценкам, оптимальным при байесовском точечном засорении, в условиях, когда параметры двух параметризаций связаны взаимно однозначной дифференцируемой функцией. В этом случае соответствующие оценочные функции оказываются эквивалентными (отличаются одна от другой несущественным матричным сомножителем C). Легко показать, повторяя выкладки из [24], что для рассматриваемого нами случая справедливы эти же результаты.

В максиминной формулировке (15) инвариантность к преобразованию параметров отсутствует, так как оказываются различными плотности s_*, а значит, и оценочные функции $\psi^*_{s_*}$. Не являются инвариантными также условно оптимальные оценки и компромиссные оценки, построенные с использованием показателя $\mathrm{stb}_s(\psi)$, для них оказываются различными величины γ и $\tilde{k}$, что приводит к различным функциям s.

Причиной отсутствия свойства инвариантности в этих случаях является изменение показателя (5) при переходе от одной параметризации к другой [24]. В связи с этим в [24] предложены функционалы Ψ, инвариантные к репараметризации. Примечательно, что основной инвариантный функционал имеет вид (7), где в качестве матрицы W может использоваться матрица, обратная к показателю (5) при некоторой фиксированной оценочной функции и некоторой фиксированной плотности s (эти оценочная функция и плотность не обязательно должны быть как-то связаны между собой). Показатель $\mathrm{stb}_s(\psi)$ обладает свойством инвариантности также при использовании функционала в виде определителя, поэтому условно оптимальные и компромиссные оценки, построенные с использованием $\mathrm{stb}_s(\psi)$, для данного функционала также обладают свойством инвариантности.

3. ОЦЕНИВАНИЕ ПО НЕПОЛНЫМ ДАННЫМ

Применим развитую в п. 1, 2 теорию к случаю неполных данных.

Если в векторе $\tilde{\zeta}_i$ значение каждого его элемента может либо присутствовать, либо отсутствовать, то всего имеется 2^n структур пропусков. Однако на практике может быть допустимо меньшее число структур пропусков, причем разное для разных наблюдений. Предположим, что для i-го наблюдения допустимо M_i структур пропусков; перенумеруем их некоторым образом. Вслед за [2] будем считать номера структур пропусков наблюдений случайными величинами. Обозначим ρ_i – такую случайную величину для i-го наблюдения, $\tilde{\rho}_i$ – ее наблюдаемое значение, а соответствующий аргумент в плотностях, оценочных функциях и т.п. будем обозначать r_i.

Для r_i-й структуры пропусков введем векторы $\zeta_{i,\mathrm{obs}}^{r_i}$ и $\zeta_{i,\mathrm{mis}}^{r_i}$, состоящие соответственно из наблюдаемых и отсутствующих элементов вектора ζ_i и имеющие плотности относительно σ-конечных мер $\mu_{i,\mathrm{obs}}^{r_i}$ и $\mu_{i,\mathrm{mis}}^{r_i}$, таких, что мера μ является их произведением.

В результате реально нам доступны векторы $\tilde{\zeta}_{i,\mathrm{obs}}$, $i=1,...,N$, наблюдений случайных векторов $\zeta_{i,\mathrm{obs}}^{r_i}$, $i=1,...,N$. Таким образом, выборку составляют векторы $\left(\tilde{\zeta}_{i,\mathrm{obs}}^T,\tilde{\rho}_i\right)^T$, $i=1,...,N$.

Поскольку наблюдения в неполной выборке принадлежат различным подпространствам исходного пространства, непосредственно применить разработанный в п. 1, 2 подход невозможно. Перейти к единому пространству можно следующим образом.

Введем произвольные n-мерные случайные векторы η_i, независимые от случайных векторов ζ_i и величин ρ_i, с плотностями $g_i^{\eta}(z_i)$, $z_i \in R^n$, относительно меры μ (эти плотности не зависят от вектора ϕ). Для r_i-й структуры пропусков введем случайные векторы $\eta_{i,\mathrm{mis}}^{r_i}$, состоящие из элементов η_i, которые имеют те же номера, что и пропущенные элементы в векторе ζ_i. Векторы $\eta_{i,\mathrm{mis}}^{r_i}$ имеют плотности $g_i^{\eta}(z_{i,\mathrm{mis}}^{r_i})$ относительно мер $\mu_{i,\mathrm{mis}}^{r_i}$. Дополним выборку фиктивными наблюдениями $\tilde{\eta}_{i,\mathrm{mis}}$ случайных векторов $\eta_{i,\mathrm{mis}}^{r_i}$. В результате выборку составляют $(n+1)$-мерные векторы $\left(\tilde{\zeta}_{i,\mathrm{obs}}^T,\tilde{\eta}_{i,\mathrm{mis}}^T,\tilde{\rho}_i\right)^T$, $i=1,...,N$.

К такой модели теория из п. 1, 2 может быть применена.

Случайные векторы $\left(\left(\zeta_{i,\mathrm{obs}}^{\rho_i}\right)^T,\left(\eta_{i,\mathrm{mis}}^{\rho_i}\right)^T,\rho_i\right)^T$ имеют плотности

$$g_i(z_{i,\mathrm{obs}}^{r_i},z_{i,\mathrm{mis}}^{r_i},r_i\mid\phi)=g_i(z_{i,\mathrm{obs}}^{r_i},r_i\mid\phi)g_i^{\eta}(z_{i,\mathrm{mis}}^{r_i}).$$

В рамках модели байесовского точечного засорения распределения случайных векторов $\left(\left(\zeta_{i,\mathrm{obs}}^{\rho_i}\right)^T,\left(\eta_{i,\mathrm{mis}}^{\rho_i}\right)^T,\rho_i\right)^T$ введем соответствующие им плотности засоряющих значений равными

$$s_i\left(z_{i,\mathrm{obs}}^{r_i},z_{i,\mathrm{mis}}^{r_i},r_i\mid\phi\right)=s_i\left(z_{i,\mathrm{obs}}^{r_i},r_i\mid\phi\right)g_i^{\eta}\left(z_{i,\mathrm{mis}}^{r_i}\right).$$

Заметим, что такой выбор не накладывает на модель реально наблюдаемых случайных векторов $\left(\left(\zeta_{i,\mathrm{obs}}^{\rho_i}\right)^T, \rho_i\right)^T$ каких-либо ограничений.

Имеем следующие результаты.

Оптимальная оценочная функция (9) не зависит от $z_{i,\mathrm{mis}}^{r_i}$ и имеет вид

$$\psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right) = C\left\{\frac{\partial}{\partial\phi}\ln g_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i \mid \phi\right) + \beta_i\right\}\frac{g_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i \mid \phi\right)}{s_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i \mid \phi\right)}.$$

Величина β_i определяется из условия

$$\sum_{r_i=1}^{M_i}\int\limits_{R^{n_{r_i}}}\psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right)g_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i \mid \phi\right)d\mu_{i,\mathrm{obs}}^{r_i} = 0,$$

где n_{r_i} – размер вектора $\zeta_{i,\mathrm{obs}}^{r_i}$.

Показатель (5) определяется матрицами следующего вида:

$$\mathrm{M}_1 = \sum_{i=1}^{N}\sum_{r_i=1}^{M_i}\int\limits_{R^{n_{r_i}}}\psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right)\frac{\partial g_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i \mid \phi\right)}{\partial\phi^T}d\mu_{i,\mathrm{obs}}^{r_i},$$

$$\mathrm{M}_{2,s} = \sum_{i=1}^{N}\sum_{r_i=1}^{M_i}\int\limits_{R^{n_{r_i}}}\psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right)\psi_i^T\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right)s_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i \mid \phi\right)d\mu_{i,\mathrm{obs}}^{r_i}.$$

В результате и показатель (5), и оптимальная оценочная функция определяются только моделью реально наблюдаемых случайных векторов $\left(\left(\zeta_{i,\mathrm{obs}}^{\rho_i}\right)^T, \rho_i\right)^T$.

Полученное решение зависит от механизма порождения пропусков – распределения случайной величины ρ_i. В некоторых случаях переменная ρ_i является мешающей и заниматься ее моделированием нежелательно. Найдем условия, при которых механизм порождения пропусков можно игнорировать.

В методе максимального правдоподобия для игнорирования механизма порождения пропусков достаточно наложить условие ОС – «отсутствующие данные отсутствуют случайно» [2] (англоязычная аббревиатура MAR – missing at random). Представим совместную плотность случайных величин ζ_i и ρ_i в виде

$$g_i(z_i, r_i \mid \phi) = g_i(z_i \mid \phi_1)g_i(r_i \mid z_i, \phi_2),$$

где $\phi = (\phi_1^T, \phi_2^T)^T$, причем векторы ϕ_1, ϕ_2 раздельны в том смысле, что параметрическое пространство ϕ есть прямое произведение параметрических пространств для ϕ_1 и ϕ_2. Тогда условие ОС можно записать в виде

$$g_i(r_i \mid z_i, \phi_2) = g_i(r_i \mid z_{i,\mathrm{obs}}^{r_i}, \phi_2).$$

Если выполнено данное условие, то функция правдоподобия для вектора основных параметров ϕ_1 строится на основе плотностей случайных величин $\zeta_{i,\mathrm{obs}}^{r_i}$.

Однако легко убедиться, что наложение условия ОС на плотности $g_i(z_i, r_i \mid \phi)$ и $s_i(z_i, r_i \mid \phi)$ в нашем случае не избавляет от необходимости моделирования случайных величин ρ_i.

Более жестким по сравнению с условием ОС является условие ОПС – «отсутствующие данные отсутствуют случайно, присутствующие данные присутствуют случайно» [2] (англоязычная аббревиатура MCAR – missing completely at random), когда случайная величина ρ_i не зависит от случайного вектора ζ_i, т.е. справедливо представление

$$g_i(r_i \mid z_i, \phi_2) = g_i(r_i \mid \phi_2).$$

Наложим условия ОПС на плотности $g_i(z_i, r_i \mid \phi)$, $s_i(z_i, r_i \mid \phi)$ и, поскольку мы желаем получить решение, не зависящее от механизма порождения пропусков, предположим, что распределение ρ_i не зависит от оцениваемых параметров модели ϕ, т. е. справедливо $\phi \equiv \phi_1$, $g_i(r_i \mid \phi_2) = g_i(r_i)$, $s_i(r_i \mid \phi_2) = s_i(r_i)$.

Выпишем вид матриц, определяющих показатель (5), и оптимальное решение (9) для этого случая.

Имеем

$$\mathrm{M}_1 = \sum_{i=1}^{N} \sum_{r_i=1}^{M_i} g_i(r_i) \int\limits_{R^{n_{r_i}}} \psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right) \frac{\partial g_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right)}{\partial \phi^T} d\mu_{i,\mathrm{obs}}^{r_i}, \tag{16}$$

$$\mathrm{M}_{2,s} = \sum_{i=1}^{N} \sum_{r_i=1}^{M_i} s_i(r_i) \int\limits_{R^{n_{r_i}}} \psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right) \psi_i^T\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right) s_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right) d\mu_{i,\mathrm{obs}}^{r_i}, \tag{17}$$

$$\psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right) = C\left\{\frac{\partial}{\partial \phi} \ln g_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right) + \beta_i\right\} \frac{g_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right) g_i(r_i)}{s_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right) s_i(r_i)}, \tag{18}$$

где $g_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right)$ – плотность распределения $\zeta_{i,\mathrm{obs}}^{r_i}$, $s_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right)$ – плотность засоряющих значений, соответствующая $\zeta_{i,\mathrm{obs}}^{r_i}$, вектор β_i определяется из условия

$$\sum_{r_i=1}^{M_i} g_i(r_i) \int\limits_{R^{n_{r_i}}} \psi_i\left(z_{i,\mathrm{obs}}^{r_i}, r_i, \phi\right) g_i\left(z_{i,\mathrm{obs}}^{r_i} \mid \phi\right) d\mu_{i,\mathrm{obs}}^{r_i} = 0. \tag{19}$$

Вновь приходим к выводу, что наше решение зависит от механизма порождения пропусков. Чтобы освободиться от необходимости моделирования случайной величины ρ_i, нужно, помимо условия ОПС, наложить некоторые дополнительные условия. Рассмотрим их.

Чтобы освободиться от отношения $g_i(r_i)/s_i(r_i)$ в оценочной функции достаточно наложить условие

$$g_i(r_i) = s_i(r_i). \tag{20}$$

В модели байесовского точечного засорения это будет означать, что распределение ρ_i искажается в соответствии с модельным распределением. Однако, поскольку в итоге мы будем избавлены от необходимости моделирования распределения ρ_i, с практической точки зрения данное условие не будет сколько-нибудь ограничивающим.

Чтобы освободиться от учета распределения ρ_i при определении вектора β_i, на оценочную функцию необходимо наложить более жесткое условие, чем (19), а именно

$$\int\limits_{R^{n_{r_i}}} \psi_i\left(z_{i,\text{obs}}^{r_i}, r_i, \phi\right) g_i\left(z_{i,\text{obs}}^{r_i} \mid \phi\right) d\mu_{i,\text{obs}}^{r_i} = 0 \tag{21}$$

для каждой r_i-й структуры пропусков. Из него, очевидно, следует выполнение условия (19). В результате для каждой r_i-й структуры пропусков будет определен собственный вектор $\beta_i^{(r_i)}$ из условия (21), а оценочная функция примет вид

$$\psi_i\left(z_{i,\text{obs}}^{r_i}, r_i, \phi\right) = C\left\{\frac{\partial}{\partial\phi}\ln g_i\left(z_{i,\text{obs}}^{r_i} \mid \phi\right) + \beta_i^{(r_i)}\right\}\frac{g_i\left(z_{i,\text{obs}}^{r_i} \mid \phi\right)}{s_i\left(z_{i,\text{obs}}^{r_i} \mid \phi\right)}. \tag{22}$$

Легко проверить, что данная оценочная функция будет оптимальной в задачах из п. 2 с ограничением (21) вместо (2).

Хотя оптимальная оценочная функция теперь не зависит от распределения ρ_i, показатель (5), в том числе и частный случай (3), определяется с его использованием. На практике можно использовать определение (5) условно по структурам пропусков (см. также [25]), т.е.

$$\mathrm{M}_1(R) = \sum_{i=1}^{N}\int\limits_{R^{n_{\tilde{\rho}_i}}} \psi_i\left(z_{i,\text{obs}}^{\tilde{\rho}_i}, \tilde{\rho}_i, \phi\right)\frac{\partial g_i\left(z_{i,\text{obs}}^{\tilde{\rho}_i} \mid \phi\right)}{\partial\phi^T} d\mu_{i,\text{obs}}^{\tilde{\rho}_i}, \tag{23}$$

$$\mathrm{M}_{2,s}(R) = \sum_{i=1}^{N}\int\limits_{R^{n_{\tilde{\rho}_i}}} \psi_i\left(z_{i,\text{obs}}^{\tilde{\rho}_i}, \tilde{\rho}_i, \phi\right)\psi_i^T\left(z_{i,\text{obs}}^{\tilde{\rho}_i}, \tilde{\rho}_i, \phi\right) s_i\left(z_{i,\text{obs}}^{\tilde{\rho}_i} \mid \phi\right) d\mu_{i,\text{obs}}^{\tilde{\rho}_i}, \tag{24}$$

где $R = \left(\tilde{\rho}_1, ..., \tilde{\rho}_N\right)^T$.

Заметим, что номера структур пропусков можно считать детерминированными величинами с рассмотрением решений, условных по пропускам. В этом случае будет справедливой формула (18), а также выражения (16), (17), (19) при

$$g_i(r_i) = s_i(r_i) = \delta_{r_i,\tilde{\rho}_i},$$

где δ_{uv} – символ Кронекера. Отсюда можно сделать вывод, что оценочная функция (22) при ограничении (21), полученная в условиях игнорирования механизма порождения пропусков, является оптимальной при показателе (5), определяемом условно по пропускам, т.е. формулами (23), (24).

ЗАКЛЮЧЕНИЕ

Развиваемый в работе подход к устойчивому оцениванию может быть использован для моделирования широкого множества типов многомерных числовых, или сводящихся к таковым, данных, в том числе разнотипных и неполных. При этом главное, что необходимо сде-

лать для его применения – это постулировать плотность распределения реально наблюдаемых случайных величин, в том числе (при необходимости) постулировать механизм порождения пропусков. Плотность распределения засоряющих значений часто определяется на основе модельной плотности.

Полученные в работе решения обоснованы как оптимальные, что соответствует логике классической теории робастности.

Важной частью работы является выяснение условий игнорирования механизма порождения пропусков. Эти условия являются более сильными, чем для метода максимального правдоподобия, но практически приемлемыми.

Поскольку изложение теории велось на достаточно абстрактном уровне, примеров ее применения не приведено. Тем не менее, на основе данной теории нами сконструированы устойчивые оценки для ряда моделей, в том числе регрессионной модели с разнотипным откликом (частного случая модели из [4]), регрессионной модели при наличии частично-группированных данных (см., например, [26]), многооткликовой нормальной регрессии при наличии пропущенных данных и игнорировании механизма порождения пропусков. Однако их описание должно стать темой отдельных работ.

СПИСОК ЛИТЕРАТУРЫ

[1] **Лбов Г.С.** Логические решающие функции и вопросы статистической устойчивости решений / Г.С. Лбов, Н.Г. Старцева. – Новосибирск: Изд-во Ин-та математики, 1999.

[2] **Литтл Р.Дж.А.** Статистический анализ данных с пропусками / Р.Дж.А. Литтл, Д.Б. Рубин. – М.: Финансы и статистика, 1991.

[3] **Little R.J.A.** Maximum likelihood estimation for mixed continuous and categorical data with missing values / R.J.A. Little, M.D. Schluchter // Biometrika. – 1985. – Vol. 72. – P. 497–512.

[4] **Лисицин Д.В.** Оценивание параметров многофакторной модели при наличии разнотипных откликов / Д.В. Лисицин // Научный вестник НГТУ. – Новосибирск, 2005. – № 1(19). – С. 11–20.

[5] **Javaras K.N.** Multiple imputation for incomplete data with semicontinuous variables / K.N. Javaras, D.A. Dyk van // J. Am. Statist. Assoc. – 2003. – Vol. 98. – P. 703–715.

[6] **Смоляк С.А.** Устойчивые методы оценивания: (Статистическая обработка неоднородных совокупностей) / С.А. Смоляк, Б.П. Титаренко. – М.: Статистика, 1980.

[7] **Робастность** в статистике. Подход на основе функций влияния / Ф. Хампель, Э. Рончетти, П. Рауссеу, В. Штаэль. – М.: Мир, 1989.

[8] **Шурыгин А.М.** Прикладная стохастика: робастность, оценивание, прогноз / А.М. Шурыгин. – М.: Финансы и статистика, 2000.

[9] **Денисов В.И.** Методы построения многофакторных моделей по неоднородным, негауссовским, зависимым наблюдениям / В.И. Денисов, Д.В. Лисицин. – Новосибирск: Изд-во НГТУ, 2008.

[10] **Лисицин Д.В.** Об оценивании параметров модели при байесовском точечном засорении / Д.В. Лисицин // Доклады АН ВШ РФ. – 2009. – № 1(12). – С. 41–55.

[11] **Little R.J.A.** Editing and imputing for quantitative survey data / R.J.A. Little, P.J. Smith // J. Amer. Statist. Assoc. – 1987. – Vol. 82. – P. 58–68.

[12] **Cheng T.-C.** High breakdown estimation of multivariate location and scale with missing observations / T.-C. Cheng, M.-P. Victoria-Feser // British J. Math. Statist. Psych. – 2002. – Vol. 55. – P. 317–335.

[13] **Copt S.** Fast algorithms for computing high breakdown covariance matrices with missing data / S. Copt, M.-P. Victoria-Feser // Theory and Applications of Recent Robust Methods / Hubert M. et al., eds. – Basel: Birkhauser, 2004. – P. 71 – 82.

[14] **Little R.J.A.** Robust estimation of the mean and covariance matrix from data with missing values / R.J.A. Little // Appl. Statist. – 1988. – Vol. 37. – P. 23–38.

[15] **Калинин А.А.** Робастное оценивание параметров регрессионных моделей с качественным откликом / А.А. Калинин, Д.В. Лисицин // Рос. науч.-техн. конф. «Информатика и проблемы телекоммуникаций», Новосибирск, 21 – 22 апр., 2011.: Материалы конф. – Новосибирск, 2011. – Т. 1. – С. 69–72.

[16] **Kalinin A.A.** Robust estimation of qualitative response regression models / A.A. Kalinin, D.V. Lisitsin // «Applied Methods of Statistical Analysis. Simulations and Statistical Inference» – AMSA'2011, Novosibirsk, 20 – 22 September, 2011.: Proceedings of the International Workshop. – P. 303–309.

[17] **Довгаль С.Ю.** Робастные методы оценивания параметров регрессионной модели со счетным откликом / С.Ю. Довгаль, Д.В. Лисицин // Рос. науч.-техн. конф. «Информатика и проблемы телекоммуникаций», Новосибирск, 21 – 22 апр., 2011.: Материалы конф. – Новосибирск, 2011. – Т. 1. – С. 64–67.

[18] **Dovgal S.Yu.** Robust estimation of count response regression models / S.Yu. Dovgal, D.V. Lisitsin // «Applied Methods of Statistical Analysis. Simulations and Statistical Inference» – AMSA'2011, Novosibirsk, 20 – 22 September, 2011.: Proceedings of the International Workshop. – P. 318–321.

[19] **Боровков А.А.** Математическая статистика / А.А. Боровков. – Новосибирск: Наука; Изд-во Ин-та математики, 1997.

[20] **Лисицин Д.В.** Оценивание при байесовском точечном засорении: связь с подходом Хампеля и минимаксная оценка / Д.В. Лисицин // Сб. науч. тр. НГТУ. – Новосибирск: НГТУ, 2011. – Вып. 3(65). – С. 61–66.

[21] **Магарил-Ильяев Г.Г.** Выпуклый анализ и его приложения / Г.Г. Магарил-Ильяев, В.М. Тихомиров. – М.: Едиториал УРСС, 2003.

[22] **Ванько В.И.** Вариационное исчисление и оптимальное управление / В.И. Ванько, О.В. Ермошина, Г.Н. Кувыркин. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2006.

[23] **Галеев Э.М.** Оптимизация: теория, примеры, задачи / Э.М. Галеев. – М.: Едиториал УРСС, 2002.

[24] **Лисицин Д.В.** Свойства инвариантности при оценивании параметров модели в условиях байесовского точечного засорения / Д.В. Лисицин // Доклады АН ВШ РФ. – 2010. – № 1(14). – С. 18–25.

[25] **Никифоров А.М.** Методы анализа данных с пропусками и их свойства. Программное обеспечение статистической обработки неполных данных / А.М. Никифоров // Статистический анализ данных с пропусками / Литтл Р.Дж.А., Рубин Д.Б. – М.: Финансы и статистика, 1991. – С. 284–332.

[26] **Денисов В.И.** Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов. В 2-х ч. / В.И. Денисов, Б.Ю Лемешко, Е.Б. Цой. – Новосиб. гос. техн. ун-т. – Новосибирск, 1993.

Лисицин Даниил Валерьевич, доктор технических наук, профессор кафедры прикладной математики Новосибирского государственного технического университета. Основное направление научных исследований – методы построения многофакторных моделей по статистическим данным. Имеет 90 публикаций, в том числе 1 монографию. E-mail: dalis2@yandex.ru

Lisitsin D.V.

Robust estimation of model parameters in presence of multivariate nonhomogeneous incomplete data

In paper the theory of an optimum estimation of unknown parameters of statistical model in presence of multivariate nonhomogeneous data develops. Outcomes are represented in the numerical form, but not necessarily are numerical, for example, they can be qualitative or mixed. Estimators provide robustness to deviation of observations distribution from postulated distribution. The basis of the theory is constructed with use of F. Hampel's approach connected with influence function and with use of A.M. Shurygin's approach connected with Bayesian dot contamination of distributions. The separate consideration is given to case of the incomplete data; conditions which make the missing-data mechanism ignorable are obtained.

Key words: parameter estimation, robustness, influence function, nonhomogeneous data, incomplete data, mixed outcomes.