Робастное энтропийное оценивание характеристик рандомизированных моделей данных*

А. Ю. Попков, Ю. С. Попков

Аннотация. Рассматриваются рандомизированные модели данных и вводятся их функциональные и числовые характеристики. Оценивание этих характеристик производится в условиях недостаточных объемов данных, измеряемых с шумами с известными областями определения. Определен принцип робастного энтропийного оценивания и сформированы два класса математических моделей, реализующие этот принцип. Исследованы структурные свойства получаемых оценок для линейной и степенной рандомизированных моделей данных.

Ключевые слова: *рандомизированные модели данных, робастность, энтропия, мультипликативные алгоритмы.*

Введение

Проблема оценивания характеристик моделей по наблюдениям их входа и выхода (моделей данных) является по-прежнему актуальной, несмотря на обилие методов, развитых в математической статистике [1], [2], эконометрике [3], финансовой математике [4], теории распознавания образов [5] и других научных дисциплинах.

Методы, разработанные в указанных научных дисциплинах, базируются на двух фундаментальных гипотезах, относящихся к *моделям* и к *данным*. Относительно моделей предполагается, что они имеют вполне определенные (будем называть их детерминированными), но с неизвестными значениями, и не доступные прямому измерению характеристики (параметры). Что касается данных, то их статистические свойства декларируются, а количество предполагается достаточным для получения оценок параметров моделей.

Обоснования возможностей применение этих гипотез при решении прикладных задач оказываются весьма затруднительным, а часто и невозможным. Если иметь в виду данные, то их всегда *недостаточно* и к тому же достоверность их вызывает серьезные сомнения. Параметризация модели, т. е. обозначение в ней определенного набора параметров, возникает из компромисса между уровнем знаний об исследуемом объекте, сложностью их математической формализации и достаточностью данных. В этой ситуации гипотеза о детерминированности характеристик модели кажется излишне жесткой и не отвечающей реальным ситуациям. Кроме того, нужно иметь в виду, что оценка значений характеристик (параметров) модели осуществляется по наблюдаемым конечным массивам входных и выходных данных, объем которых и их достоверность, как правило, невелики. Если относиться к массивам данных как к случайным объектам, то и оценки параметров приобретают свойства случайных переменных.

Поэтому естественным образом возникает предложение рассматривать параметры модели как случайные величины. Это предложение приводит к трансформации модели данных с детерминированными параметрами в модель данных со случайными параметрами, которую будем называть *рандомизированной моделью данных* (*РМД*). В РМД случайные параметры характеризуются функциями плотности распределения вероятностей (ПРВ), принадлежащими заданному классу. Поэтому для РМД нужно определять, используя данные, не оценки значений параметров, а оценки функции ПРВ этих параметров.

При этом следует иметь ввиду, что кроме недостаточности объемов массивов входных и выходных данных, элементы как входных, так и выходных массивов содержат случайные ошибки, которые также характеризуются функциями ПРВ из соответствующих классов для входных и выходных данных. Функции ПРВ случайных ошибок также неизвестны, и их нужно как-то определить.

Данная работа ориентирована на развитие методов оценивания функций плотности распределения вероятностей параметров рандомизированных моделей данных и измерительных шумов **в условиях недостаточного объема данных**. Предлагается строить оценки указанных ПРВ, максимизируя их сум-

* Работа поддержана РФФИ (проект № 11-07-00065).

марную энтропию при заданных РМД, классах ПРВ и моделях наблюдений. Преимущества энтропийного критерия для построения оценок ПРВ основываются на общих методологических принципах и интерпретациях, которые связывают энтропию с одной стороны, с информацией и энергией (по крайней мере для физических систем), а с другой — с мерой неопределенности. Получаемое при максимизации энтропии решение трактуется как решение, наилучшее при максимальном уровне неопределенности [6]. Если принять за свойство робастности уровень неопределенности, измеряемый значениями энтропийной функции, то ее максимизация дает робастные оценки указанных ПРВ. Эти оценки не зависят от других элементов из заданных классов ПРВ [7].

1. Рандомизированные модели данных и их числовые характеристики

Рассмотрим статический параметризованный объект, про который известны двумерный массив входных данных — матрица Z, с размерностью $(s \times n)$, где s — количество измерений входа, n — количество параметров объекта; и одномерный массив выходных данных — вектор $\mathbf{u}$, размерности s.

Связь между указанными массивами характеризуется **рандомизированной моделью данных (РМД)**, которая состоит из

— модели связи «вход Z — выход $\mathbf{y}$», описываемой заданным s-мерным оператором $\mathbf{F}$ с параметрами $\mathbf{a}$ в виде:

$$\mathbf{y} = \mathbf{F}[Z, \mathbf{a}]; \qquad (1)$$

— модели измерений:

$$Z = X + \eta, \qquad \mathbf{v} = \mathbf{y} + \xi, \qquad (2)$$

где X — матрица идеальных измерений входа.

В этих выражениях:

$\mathbf{a}$ — случайный вектор параметров с независимыми компонентами a_i, $i = \overline{1, n}$, которые принимают значения в интервалах $\mathcal{A}_i$, $i = \overline{1, n}$;

$(s \times n)$-матрица η имеет независимые случайные элементы η_{ji}, имитирующие шумы измерений входа, принимающие значения в интервалах $\mathcal{E}_{ji}$, $j = \overline{1, s}$, $i = \overline{1, n}$;

s-мерный вектор ξ имеет независимые случайные компоненты ξ_j, имитирующие шумы измерений выхода, принимающие значения в интервалах $\mathcal{Q}_j$, $j = \overline{1, s}$.

Будем различать два класса РМД в зависимости от того, как характеризуются ее случайные компоненты.

1. РМД-PWQ. В РМД данного класса ее случайные компоненты (параметры и шумы) предполагаются вещественными числами в соответствующих интервалах, на которых определены функции плотности распределения вероятностей (ПРВ):

– для случайных параметров

$$P(\mathbf{a}) = \prod_{i=1}^{n} p_i(a_i), \qquad a_i \in \mathcal{A}_i = [a_i^-, a_i^+]; \quad (3)$$

– для шумов измерений входа

$$W(\eta) = \prod_{j=1}^{s} \prod_{i=1}^{n} w_{ji}(\eta_{ji}), \qquad \eta_{ji} \in \mathcal{E}_{ji} = [\eta_{ji}^-, \eta_{ji}^+]; \qquad (4)$$

– для шумов измерений выхода

$$Q(\xi) = \prod_{j=1}^{s} q_j(\xi_j), \qquad \xi_j \in \Xi_j = [\xi_j^-, \xi_j^+]. \quad (5)$$

Указанные ПРВ случайных параметров и шумов подлежат оцениванию, используя наблюдения X, $\mathbf{u}$, РМД (1), (2) и априорную информацию об указанных ПРВ. Последняя формализуется как в терминах априорных ПРВ параметров $\hat{P}(\mathbf{a})$ и шумов $\hat{W}(\eta)$, $\hat{Q}(\xi)$, так и в описаниях классов ПРВ.

Рассмотренная РМД генерирует ансамбль случайных векторов $\mathbf{v}$. Поэтому для учета наблюдений необходимо использовать числовые характеристики этого ансамбля. В качестве таковых воспользуемся моментами компонент вектора выхода РМД $\mathbf{v}$:

$$\mathbf{m}^{(k)} = \{\mathcal{M}(v_1^k), \dots, \mathcal{M}(v_s^k)\}, \qquad (6)$$

где k — порядок момента,

$$\mathcal{M}(v_j^k) = \int\limits_{\mathbf{a}\in\mathcal{A}),\eta\in\mathcal{E},\xi\in Xi} \big(F_j[X+\eta,\mathbf{a}]+\xi_j\big)^k \, d\mathbf{a}d\eta d\xi, \quad j = \overline{1, s}. \qquad (7)$$

Далее будут использоваться моменты первого порядка, т. е. средние значения компонент выхода РМД:

$$\bar{\mathbf{v}} = \int\limits_{\mathbf{a}\in\mathcal{A}),\eta\in\mathcal{E},\xi\in Xi} (\mathbf{F}[X+\eta,\mathbf{a}]+\xi) \, d\mathbf{a}d\eta d\xi. \qquad (8)$$

2. РМД-pwq. В РМД данного класса ее случайные компоненты (параметры и шумы) также предполагаются вещественными числами, принадлежность которых к соответствующему интервалу характеризуется некоторой вероятностью.

Так, параметры $a_1, \dots, a_n$ принимают значения в интервалах $\mathcal{A}_1, \dots, \mathcal{A}_n$ с вероятностями $p_1, \dots, p_n$ соответственно. Вероятности $p_i \in [0, 1]$, $i = \overline{1, n}$.

Аналогично, элементы матрицы шумов измерений входа ηji принимают значения в интервалах $\mathcal{E}ji$ с вероятностями $w_{ji} \in [0,1]$ соответственно; и компоненты шумов измерений выхода ξ_j принимают значения в интервалах Ξ_j с вероятностями $q_j \in [0,1]$ соответственно. Здесь $j = \overline{1,s}$, $i = \overline{1,n}$. Априорные значения указанных вероятностей обозначим $\hat{p}_i$, $\hat{w}_{ji}$, $\hat{q}_j$.

РМД рассматриваемого класса, так же как и РМД-PWQ, воспроизводит ансамбль случайных векторов $\mathbf{v}$, генерируя случайные параметры с вероятностями $\mathbf{p}$, компоненты шумов измерений выхода с вероятностями $\mathbf{q}$ и элементы матрицы шумов измерения входа с вероятностями W.

В качестве числовой характеристики ансамбля, генерируемого РМД рассматриваемого класса, будем использовать, следуя [3], квазимомент первого порядка, который основан на следующем преобразовании переменных (параметров и компонент шумов):

$$\mathbf{a} = \mathbf{a}^- + L_{\mathbf{a}}\mathbf{p}, \quad \eta = \eta^- + L\eta \bullet W, \quad \xi = \xi^- + L_\xi \mathbf{q}, \quad (9)$$

где

$L_{\mathbf{a}}, L_\xi$ — диагональные матрицы, элементы которых длины соответствующих интервалов:

$$L_{\mathbf{a}}^i = a_i^+ - a_i^-, \quad L_\xi^j = \xi^+ - \xi^-, \quad i = \overline{1,n}, \quad j = \overline{1,s}; \quad (10)$$

L_η — матрица длин интервалов $\mathcal{E}_{ji}$:

$$L_\eta^{ji} = \eta_{ji}^+ - \eta_{ji}^-, \qquad i = \overline{1,n},\ j = \overline{1,s}; \quad (11)$$

$\bullet$ — знак поэлементного умножения.

Преобразование (9) можно интерпретировать как замена случайных параметров и шумов их «квазисредними» значениями.

Подставляя (9) в (1), (2) получим выражение для квазимомента первого порядка выхода РМД-pwq через вероятности $\mathbf{p}, W, \mathbf{q}$:

$$\tilde{\mathbf{v}} = \mathbf{F}\left[X + (\eta^- + L\eta \bullet W), \mathbf{a}^- + L_{\mathbf{a}}\mathbf{p}\right] + \xi^- + L_\xi \mathbf{q}. \quad (12)$$

2. Принципы робастного энтропийного оценивания

Поскольку объектом оценивания являются функции плотности распределения вероятностей, то введем не функцию, а *функционал правдоподобия*, максимизация которого по всем функциям ПРВ из заданного класса и при наличии информации о входе и выходе РМД определит наилучшую оценку.

Рассмотрим процедуру формирования функционала правдоподобия для оценивания ПРВ параметров РМД (1), (2) [1].

[1] Для ПРВ шумов вводится аналогично.

Обозначим $P(\mathbf{a} \mid \mathbf{u}, Z)$ — условная ПРВ параметров $\mathbf{a}$ при наблюденных выходе $\mathbf{u}$ и входе Z. Следуя [2], введем логарифмическое отношение правдоподобия (ЛОП) в следующем виде:

$$\varphi_{\mathbf{u},Z}(\mathbf{a}) = \ln\left(\frac{P(\mathbf{a} \mid \mathbf{u}, Z)}{P^0(\mathbf{a})}\right). \quad (13)$$

Отсюда видно, что ЛОП является неслучайной функцией случайного аргумента (параметров) при фиксированных входе Z и выходе $\mathbf{u}$.

1. Функционал правдоподобия $\mathcal{L}$ определим, используя операцию осреднения функции ЛОП (13):

$$\mathcal{L}[P(\mathbf{a} \mid \mathbf{u}, Z)] = \int\limits_{\mathbf{a}\in\mathcal{A}} P(\mathbf{a} \mid \mathbf{u}, Z)\varphi_{\mathbf{u},Z}(\mathbf{a})d\mathbf{a}. \quad (14)$$

Нетрудно заметить, учитывая выражение функции φ (13), что функционал правдоподобия, взятый со знаком минус, представляет собой обобщенный энтропийный функционал Больцмана ([9], [10], [11]). Последний имеет много интерпретаций. В частности, она трактуется как «расстояние» между ПРВ $P(\mathbf{a} \mid \mathbf{u}, Z)$ и $P^0(\mathbf{a})$ [12], или как мера неопределенности ПРВ $P(\mathbf{a} \mid \mathbf{u}, Z)$ [13] по отношению к ПРВ $P^0(\mathbf{a})$ [14], или как мера **робастности**, т. е. степени инвариантности ПРВ $P(\mathbf{a} \mid \mathbf{u}, Z)$ по отношению к наблюдениям [8]. Воспользуемся последней трактовкой функционала энтропии для построения оценок ПРВ.

Учитывая, что компоненты вектора параметров и шумов независимые (3)–(5), определим функционал энтропии для указанных ПРВ в виде:

$$\begin{aligned}
&\mathcal{H}[P(\mathbf{a}, W(\eta), Q(\xi) \mid \mathbf{u}, Z)] = \\
&= -\sum_{i=1}^{n} \int\limits_{a_i\in\mathcal{A}_i} p_i(a_i) \ln \frac{p_i(a_i)}{p_i^0(a_i)} da_i - \\
&- \sum_{j=1}^{s}\sum_{i=1}^{n} \int\limits_{\eta_{ji}\in\mathcal{E}_{ji}} w_{ji}(\eta_{ji}) \ln \frac{w_{ji}(\eta_{ji})}{w_{ji}^0(\eta_{ji})} d\eta_{ji} - \\
&- \sum_{j=1}^{s} \int\limits_{\xi_j\in Xi_j} q_j(\xi_j) \ln \frac{q_j(\xi_j)}{q_j^0(\xi_j)} d\xi_j. \quad (15)
\end{aligned}$$

Здесь $\mathcal{A}_i, \mathcal{E}_{ji}, \Xi_j$ — множества определения компонента вектора параметров a_i, элементов матрицы шумов измерений входа η_{ji} и компонент вектора шумов измерений выхода ξ_j соответственно.

Если функции $p_i(a_i), w_{ji}(\eta_{ji}), q_j(\xi_i)$ — дискретные и их ординаты представляют собой вероятности попадания случайных величин a_i, η_{ji}, ξ_j в соответствующие множества $\mathcal{A}_i, \mathcal{E}_{ji}, \Xi_j$, то функционал энтропии трансформируется в функцию — обобщенную информационную энтропию Больцмана [11], которая принимает следующий вид:

$$H[\mathbf{p}, W, \mathbf{q} \mid \mathbf{u}, Z] = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{p_i^0} - \\ - \sum_{j=1}^{s}\sum_{i=1}^{n} w_{ji} \ln \frac{w_{ji}}{w_{ji}^0} - \\ - \sum_{j=1}^{s} q_j \ln \frac{q_j}{q_j^0}. \qquad (16)$$

Напомним, что в равенства (15), (16) входят условные ПРВ и условные вероятности при фиксированном наблюдаемом входе Z и выходе $\mathbf{u}$ исследуемого объекта.

Поэтому оценки ПРВ должны определяться с учетом наблюдений $Z, \mathbf{u}$ и связей между ними, описываемой моментными характеристиками РМД-PWQ (6), (7) или квазимоментом первого порядка (квазисреднего) для РМД-pwq (12).

2. $\mathcal{S}^k_{PWQ}$-робастная энтропийная оценка функций плотности распределения вероятностей параметров $p_i(a_i)\ (i = \overline{1,n})$ и функций плотности распределения вероятностей компонент шумов измерений $w_{ji}(\eta_j i), q_j(\xi_j)\ i = \overline{1,n},\ j = \overline{1,s})$ определяется решением следующей задачи:

$$\mathcal{H}[P(\mathbf{a}, W(\eta), Q(\xi) \mid \mathbf{u}, Z)] \Rightarrow \max, \qquad (17)$$

при условии, что плотности распределения вероятностей принадлежат к

- классу
$$\mathcal{S} = \mathcal{P} \bigcup \mathcal{Q} \bigcup \mathcal{W}, \qquad (18)$$
где $\mathcal{P}, \mathcal{Q}, \mathcal{W}$ — классы ПРВ параметров, входных и выходных шумов соответственно;
- и выполняется баланс между k-моментным выходом РМД-PWQ $\mathbf{v}$ и наблюдаемым массивом данных $\mathbf{u}$:
$$\left(\mathcal{M}\{[\mathbf{F}^{(k)}(X_0 + \eta)] + \xi^{(k)}\}\right)^{(1/k)} = \mathbf{u}. \qquad (19)$$

Функционал энтропии $\mathcal{H}[P(\mathbf{a}, W(\eta), Q(\xi) \mid \mathbf{u}, Z)]$ определяется выражением (15). В равенстве (19) $\mathbf{F}^{(k)}$ — оператор с компонентами F_j^k и $\xi^{(k)}$ — вектор с компонентами $\xi_j^k,\ j = \overline{1,s}$.

3. $\mathcal{S}^{\tilde{1}}_{pwq}$-робастная энтропийная оценка функций распределения вероятностей параметров $p_i\ (i = \overline{1,n})$, и функций плотности распределения вероятностей компонент шумов измерений $w_{ji}, q_j\ i = \overline{1,n},\ j = \overline{1,s})$ определяется решением следующей задачи:

$$H[\mathbf{p}, W, \mathbf{q} \mid \mathbf{u}, Z)] \Rightarrow \max, \qquad (20)$$

при условии, распределения вероятностей принадлежат к

- классу
$$\mathbb{S} = \mathbb{P} \bigcup \mathbb{Q} \bigcup \mathbb{W}, \qquad (21)$$
где $\mathbb{P}, \mathbb{Q}, \mathbb{W}$ — классы В параметров, входных и выходных шумов соответственно;
- и выполняется баланс между квазисредним выходом РМД-pwq $\tilde{\mathbf{v}}$ и наблюдаемым массивом выходных данных $\mathbf{u}$:
$$\mathbf{F}\left[X + (\eta^- + L\eta W), (\mathbf{a}^- + L_{\mathbf{a}}\mathbf{p}\right] + \xi^- + L_\xi \mathbf{q} = \mathbf{u}. \qquad (22)$$

Энтропия $H[\mathbf{p}, W, \mathbf{q} \mid \mathbf{u}, Z)]$ определяется выражением (16).

3. Структурные свойства оценок $\mathcal{S}^1_{PQ}$ для степенной РМД-PQ.

Рассмотрим статическую РМД следующего вида:

$$y(t) = \sum_{k=1}^{N} \langle \mathbf{x}^{(k)}(t), \mathbf{a}^k \rangle, \qquad v = y + \xi. \qquad (23)$$

В этом равенстве $\langle \bullet \rangle$ — знак скалярного произведения, $\mathbf{a}^{(1)} = \{a_1, \ldots, a_n\}$ — вектор случайных параметров с неизвестной функцией ПРВ $P(\mathbf{a})$. Векторы $\mathbf{a}^{(k)}$, $k > 1$ содержат компоненты — степени соответствующих компонент вектора $\mathbf{a}^{(1)}$, т. е. $\mathbf{a}^{(k)} = \{a_1^k, \ldots, a_n^k\}$. Переменные v, y — измеряемый и идеальный выходы РМД, ξ — аддитивный шум. Последний характеризуется соответствующей ПРВ $Q(\xi)$, которая также неизвестна.

Параметры и компоненты шума измерений выхода независимые (3), (5), и принимают значения в соответствующих интервалах. Предполагается, что вход измеряется точно, т. е. шумы $\eta = 0$.

Для построения оценок функций ПРВ $p_1(a_1)$, $\ldots$, $p_n(a_n)$ и $q(\xi_1)$, $\ldots$, $q_s(\xi_s)$ используются измерения, из которых формируется вектора $\mathbf{y}$ и $\mathbf{v}$ (2) размерности s. Предположим, что априорная информация отсутствует, т. е. $P^0(\mathbf{a}) = Q(\xi) = \mathrm{const}$.

Входными измеряемыми переменными являются векторы $\mathbf{x}^{(1)}$, $\ldots$, $\mathbf{x}^{(N)}$, из компонент которых формируются матрицы

$$X^{(k)} = \begin{pmatrix} x_{11}^{(k)} & \cdots & x_{1,n}^{(k)} \\ \cdots & \cdots & \cdots \\ x_{s,1}^{(k)} & \cdots & x_{s,n}^{(k)} \end{pmatrix}, \quad k = \overline{1,N}. \qquad (24)$$

Итак, с учетом измерений, РМД-PQ можно представить в следующем виде:

$$\mathbf{y} = \sum_{k=1}^{N} X^{(k)} \mathbf{a}^{(k)}, \\ \mathbf{v} = \mathbf{y} + \xi. \qquad (25)$$

Рассмотрим задачу (17)–(19), представив ее в следующем виде:

$$\mathcal{H}[p(\mathbf{a}), q(\xi)] = \\ = -\sum_{i=1}^{n} \int\limits_{a_i \in \mathcal{A}_i} p_i(a_i) \ln p_i(a_i) da_i - \\ - \sum_{j=1}^{s} \int\limits_{\xi_j \in \Xi_j} q_j(\xi_j) \ln q_j(\xi_j) d\xi_j \Rightarrow \max, \tag{26}$$

при ограничениях:

$$D_i[p_i(a_i)] = 1 - \int\limits_{a_i \in \mathcal{A}_i} p_i(a_i) da_i = 0,$$

$$T_j[q_j(\xi_j)] = 1 - \int\limits_{\xi \in \Xi_j} q_j(\xi_j) d\xi_j = 0,$$

$$i = \overline{1, n}, \ \ j = \overline{1, s}; \tag{27}$$

$$\Phi_j[p(\mathbf{a}), q(\xi)] = \sum_{k=1}^{N} \sum_{i=1}^{n} x_{ji}^{(k)} \int\limits_{a_i \in \mathcal{A}_i} a_i^k p_i(a_i) da_i + \\ + \int\limits_{\xi \in \Xi_j} \xi_j q_j(\xi_j) d\xi_j - u_j = 0, \quad j = \overline{1, s}. \tag{28}$$

Под структурными свойствами $\mathcal{S}_{PQ}^1$-оценки понимается вид робастного семейства функций плотности распределения вероятностей параметров и шума. Оказывается, что имеет место следующий результат.

Теорема. *$\mathcal{S}_{PQ}^1$-оценки $\hat{p}_i(a_i)$, $(i = \overline{1, n})$, $\hat{q}_j(\xi_j)$, $j = \overline{1, s}$ функций плотности распределения вероятностей параметров и шумов измерений имеют следующую структуру*:

$$\hat{p}_i(a_i) \sim \pi_i \exp\,(-\sum_{k=1}^{N} \alpha_{ik} a_i^k), \qquad i = (\overline{1, n});$$

$$\hat{q}_j(\xi_j) \sim \kappa_j \exp\,(-\omega_j \xi_j), \qquad j = (\overline{1, s}), \tag{29}$$

где π_i, α_i и κ_j, ω_j — коэффициенты.

Доказательство этого утверждения основано на необходимых условиях оптимальности в задаче (26)–(28) максимизации функционала энтропии $\mathcal{H}[p(\mathbf{a}), q(\xi)]$ (26) при ограничениях функционального типа (27), (28). Эти условия формулируются в терминах вариаций $\delta\mathcal{L}$ функционала Лагранжа $\mathcal{L}$:

$$\mathcal{L}[p(\mathbf{a}), q(\xi) \,|\, \lambda, \varepsilon, \theta] = \mathcal{H}[p(\mathbf{a}), q(\xi)] + \\ + \langle \lambda, \mathbf{D}[p(\mathbf{a})] \rangle + \langle \varepsilon, \mathbf{T}[q(\xi)] \rangle + \langle \theta, \Phi[p(\mathbf{a}), q(\xi)] \rangle. \tag{30}$$

В этом выражении $\lambda, \varepsilon, \theta$ — множители Лагранжа.

Представим функции $p(\mathbf{a})$ и $q(\xi)$ виде линейной комбинации функций $\hat{p}(\mathbf{a}), \hat{q}(\xi)$, являющихся решением задачи (26)–(28), и произвольных функций $h(\mathbf{a}) \in \mathcal{P}$ и $l(\xi) \in \mathcal{Q}$:

$$p_i(a_i) = \hat{p}_i(a_i) + \beta_i h_i(a_i), \\ q_j(\xi_j) = \hat{q}_j(\xi_j) + \gamma_j l_j(\xi_j), \\ i = \overline{1, n}, \ \ j = \overline{1, s}, \tag{31}$$

где β_i, γ_j — коэффициенты. Тогда необходимые условия стационарности функционала Лагранжа по прямым переменным — функциям $p(\mathbf{a}), q(\xi)$ можно сформулировать в следующем виде:

$$\frac{\partial \mathcal{L}}{\partial \beta_i}|_{(\beta_i = 0)} = 0, \quad \frac{\partial \mathcal{L}}{\partial \gamma_j}|_{(\gamma_j = 0)} = 0, \\ i = \overline{1, n}, \ j = \overline{1, s}. \tag{32}$$

В результате получаем систему из $n + s$ интегральных уравнений:

$$\int\limits_{a_i \in A_i} h_i(a_i)(-\ln \hat{p}_i(a_i) - 1 - \lambda_i - \\ - \sum_{j=1}^{s} \theta_j [\sum_{k=1}^{N} x_{ji}^{(k)} a_i^k) da_i = 0,$$

$$\int\limits_{\xi \in \Xi_j} l_j(\xi_j) \left(-\ln \hat{q}_j(\xi_j) - 1 - \varepsilon_j - \theta_j \xi_j\right) d\xi_j = 0, \\ i = \overline{1, n}, \ j = \overline{1, s}. \tag{33}$$

Для выполнения этих интегральных условий при любых функциях $h_i(a_i), l_j(\xi_j)$, необходимо и достаточно, чтобы:

$$-\ln \hat{p}_i(a_i) - 1 - \lambda_i - \sum_{j=1}^{s} \theta_j \sum_{k=0}^{N} x_{ji}^{(k)} a_i^k = 0, \\ -\ln \hat{q}_j(\xi_j) - 1 - \varepsilon_j - \theta_j \xi_j = 0, \\ i = \overline{1, n}, \ j = \overline{1, s}. \tag{34}$$

Отсюда $\mathcal{S}_{PQ}^1$-оценки функций плотности распределения вероятностей приобретают следующий вид:

$$\hat{p}_i(a_i) \ = \exp \left(-1 - \lambda_i - \sum_{k=1}^{N} \sum_{j=1}^{s} \theta_j x_{ji}^{(k)} a_i^k \right), \\ \hat{q}_j(\xi_j) \ = \exp \left(-1 - \epsilon_j - \theta_j \xi_j \right), \\ i = \overline{1, n}, \ j = \overline{1, s}. \tag{35}$$

Обозначая

$$\pi_i = \exp\,(-1 - \lambda_i), \ \ \alpha_{ik} = \sum_{j=1}^{s} \theta_j x_{ji}^{(k)}, \\ \kappa_j = \exp\,(-1 - \varepsilon_j), \ \ \omega_j = \theta_j, \tag{36}$$

получим утверждение теоремы. Из (35) видно, что функций плотности распределения вероятностей определены с точностью до констант (36), которые зависят от множителей Лагранжа $\lambda, \varepsilon, \theta$. Для того чтобы они стали $\mathcal{S}^1_{PQ}$-оценками нужно найти такие значения указанных множителей Лагранжа, при которых выполняются ограничения (27), (28). Однако также можно видеть, что от конкретных значений множителей Лагранжа структура оценок не зависит. Она определяется структурой РМД (23). На рис. 1–4 показаны примеры оценок ПРВ параметров. Для линейной РМД $\mathcal{S}^1_{PQ}$-оценка ПРВ параметров всегда экспоненциального типа, что иллюстрирует рис. 1. Наблюдения входа и выхода не влияют на структуру оценки: форма ПРВ всегда экспоненциальная, но с различным затуханием. Для нелинейных РМД структура $\mathcal{S}^1_{PQ}$-оценка ПРВ параметров более разнообразная и зависит от наблюдений входа и выхода. На рис. 2–4 показаны оценки ПРВ для квадратичной, квадратично-линейной и кубической РМД-PQ.

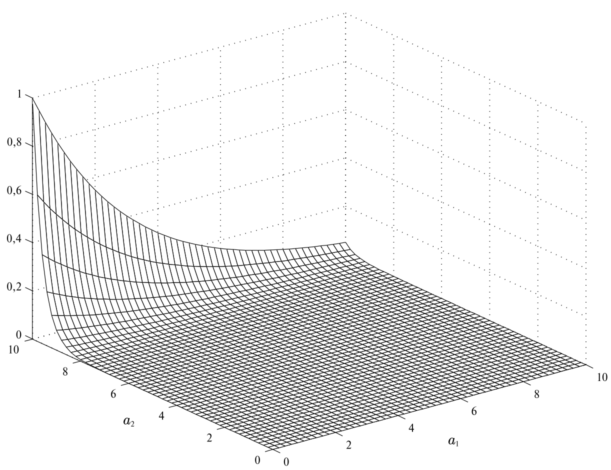

Рис. 1. $p(a_1, a_2) = \exp\left(-0{,}3a_1 + 2{,}5a_2\right)$

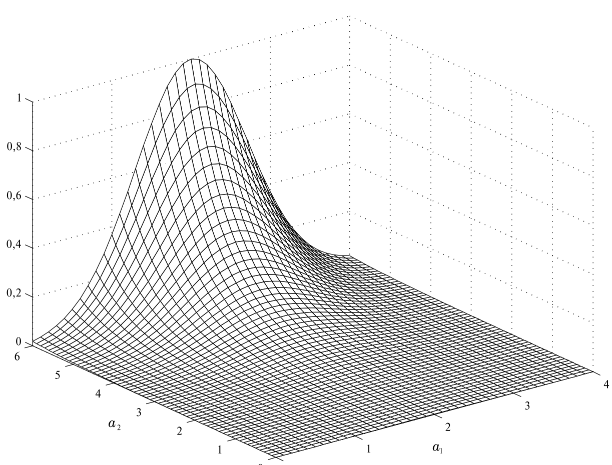

Рис. 2. $p(a_1, a_2) = exp(4a_1 - a_1^2 + a_2)$

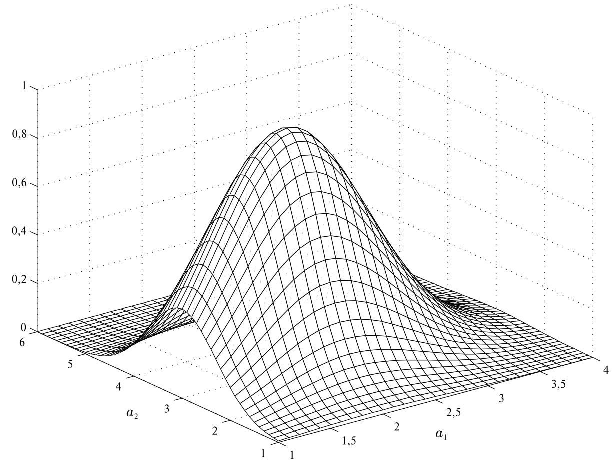

Рис. 3. $p(a_1, a_2) = exp(4a_1 - a_1^2 + 6a_2 - a_2^2)$

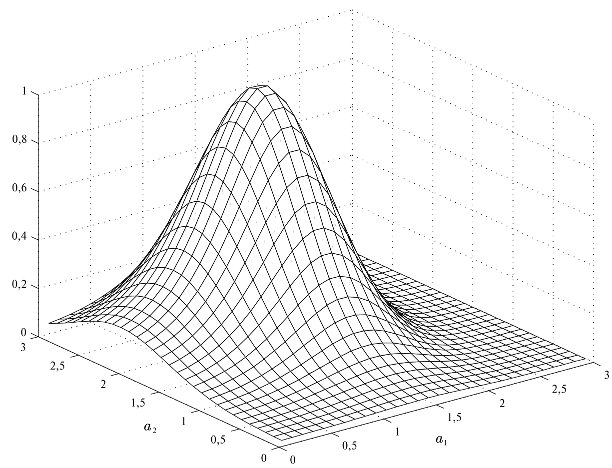

Рис. 4.
$p(a_1, a_2) = exp(0{,}5a_1 + 3{,}5a_2 + 2a_1^2 - a_1^3 - 0{,}2a_2^2 - 0{,}2a_2^3)$

4. Структурные свойства оценок $\mathcal{S}^{\tilde{1}}_{pq}$ ПРВ линейной РМД-pwq

Рассмотрим линейную РМД-pwq с шумами измерений входа и выхода:

$$\mathbf{y} = X\mathbf{a}, \qquad \mathbf{v} = \mathbf{y} + \xi, \tag{37}$$

где векторы идеального выхода $\mathbf{y}$, измеряемого выхода $\mathbf{v}$ и шума измерений выхода ξ имеют размерность s (количество наблюдений); матрица X имеют размерность $(s \times n)$, случайный вектор параметров $\mathbf{a}$ имеет размерность n. Параметры и шумы принимают значения в соответствующих интервалах, что позволяет применить преобразование (9), (12) и воспользоваться выражением для квазимомента 1-го порядка выхода РМД-pwq (37):

$$\tilde{\mathbf{v}} = XL_{\mathbf{a}}\mathbf{p} + L_\xi\mathbf{q} + \mathbf{K}(\mathbf{a}^-, \xi^-), \tag{38}$$

где

$$\mathbf{K}(\mathbf{a}^-,\xi^-) = X\mathbf{a}^- + \xi^-. \qquad (39)$$

Пусть имеется априорная информация про распределения вероятностей параметров и шумов в виде априорных распределений $\mathbf{p}^0, \mathbf{q}^0$. Тогда $\mathcal{S}_{pq}^{\tilde{1}}$-оценка есть решение следующей задачи максимизации энтропии (для класса *нормированных вероятностей*):

$$H(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{p_i^0} - \\ -\sum^{s} j = 1 q_j \ln \frac{q_j}{q_j^0} \Rightarrow \max. \qquad (40)$$

при условиях нормировки вероятностей:

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{j=1}^{s} q_j = 1; \qquad (41)$$

и баланса реальных наблюдений с квазимоментом первого порядка выхода РМД-pq:

$$\sum_{i=1}^{n} x_{ji} L_{\mathbf{a}}^i p_i + L_\xi^j q_j + K_j = u_j, \qquad j = \overline{1,s}. \qquad (42)$$

Данная задача относится к классу задач энтропийно-линейного программирования, и ее решение строится с использованием необходимых условий оптимальности в терминах функции Лагранжа:

$$L(\mathbf{p},\mathbf{q},\lambda,\varepsilon,\theta) = H(\mathbf{p},\mathbf{q}) + \lambda(1 - \sum_{i=1}^{n} p_i) + \\ + \varepsilon\left(1 - \sum_{j=1}^{s} q_j\right) + \sum_{j=1}^{s} \theta_j \left[u_j - \\ - \sum_{i=1}^{n} x_{ji} L_{\mathbf{a}}^i p_i - L_\xi^j q_j - K_j\right]. \qquad (43)$$

Имеем следующую систему уравнений, определяющие энтропийно-оптимальные вероятности $\hat{\mathbf{p}}, \hat{\mathbf{q}}$:

$$0 \leqslant \hat{p}_i(\theta) = \frac{p_i^0 \exp\left(-\sum\limits_{j=1}^{s} \theta_j \tilde{x}_{ji}\right)}{\sum\limits_{i=1}^{n} p_i^0 \exp\left(-\sum\limits_{j=1}^{s} \theta_j \tilde{x}_{ji}\right)} \leqslant 1, \\ \tilde{x}_{ji} = L_{\mathbf{a}}^i p_i, \quad i = \overline{1,n}, \qquad (44) \\ 0 \leqslant \hat{q}_j(\theta) = \frac{q_j^0 \exp\left(-\theta_j L_\xi^j\right)}{\sum\limits_{j=1}^{s} q_j^0 \exp\left(-\theta_j L_\xi^j\right)} \leqslant 1, \\ j = \overline{1,s},$$

и соответствующие им множители Лагранжа $\theta_1,\ldots,\theta_s$:

$$\Phi_j(\theta)\frac{1}{u_j}\sum_{i=1}^{n} x_{ji} L_{\mathbf{a}}^i \hat{p}_i(\theta) + L_\xi^j \hat{q}_j(\theta) + K_j = 1, \\ j = \overline{1,s}. \qquad (45)$$

Рассмотрим случай, когда вместо условия (41) введены *интервальные ограничения*:

$$0 \leqslant p_i \leqslant 1, \quad i = \overline{1,n}; \quad 0 \leqslant q_j \leqslant 1, \quad j = \overline{1,s}. \qquad (46)$$

Тогда $\mathcal{S}_{pq}^{\tilde{1}}$-оценка есть решение следующей задачи максимизации энтропии (обобщенной информационной энтропии Ферми—Дирака [11]):

$$H(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{n}\left(p_i \ln \frac{p_i}{\varrho_i^0} + (1-p_i)\ln(1-p_i)\right) - \\ -\sum^{s} j = 1\left(q_j \ln \frac{q_j}{\rho_j^0} + (1-q_j)\ln(1-q_j)\right) \Rightarrow \max, \qquad (47)$$

при условии

$$\sum_{i=1}^{n} x_{ji} L_{\mathbf{a}}^i p_i + L_\xi^j q_j + K_j = u_j, \qquad j = \overline{1,s}, \qquad (48)$$

где

$$\varrho_i = \frac{p_i^0}{1 - p_i^0}, \qquad \rho_j^0 = \frac{q_j^0}{1 - q_j^0}. \qquad (49)$$

Используя технику множителей Лагранжа, получим следующую систему уравнений для определения оценок вероятностей $\mathbf{p}, \mathbf{q}$:

$$0 \leqslant \hat{p}_i(\theta) = \frac{p_j^0 \exp\left(-\sum\limits_{j=1}^{s} \theta_j \tilde{x}_{ji}\right)}{1 + p_j^0 \exp\left(-\sum\limits_{j=1}^{s} \theta_j \tilde{x}_{ji}\right)} \leqslant 1, i = \overline{1,n}, \\ 0 \leqslant \hat{q}_j(\theta) = \frac{q_j^0 \exp\left(-\theta_j L_\xi^j\right)}{1 + q_j^0 \exp\left(-\theta_j L_\xi^j\right)} \leqslant 1, j = \overline{1,s}; \qquad (50)$$

и множителей Лагранжа $\theta_1,\ldots,\theta_s$:

$$\Phi_j(\theta) = \frac{1}{u_j - K_j}\sum_{i=1}^{n} \theta_j \tilde{x}_{ji} \hat{p}_i(\theta) + L_\xi^j \hat{q}_j(\theta) = 1, \quad \overline{1,s}. \qquad (51)$$

Поскольку оценки вероятностей $\mathbf{p}$ и $\mathbf{q}$ аналитически выражаются через множители Лагранжа θ, то фактически нужно решать только систему уравнений (45) или (51). Для этого можно использовать мультипликативный алгоритм следующего вида [11]:

$$z_j^{h+1} = z_j^h \tilde{\Phi}_j(z^h), \\ z_j = \exp(-\theta_j), \quad z_j^0 > 0, \quad j = \overline{1,s}. \qquad (52)$$

Следует отметить, что оценки вероятностей, которые получаются в результате решения задач (40)–(41), (47)–(49) могут отличаться (при определенных условиях) по величине энтропии (обобщенной энтропии Больцмана) (40).

Условия и соотношения между оптимальными значениями энтропии (40) для указанных задач определяются теоремой 2. Прежде чем формулировать ее, введем некоторые обозначения:

$$H(\mathbf{p}, \mathbf{q}) = H(\mathbf{w}), \quad \mathbf{w} = (\mathbf{p}, \mathbf{q}); \tag{53}$$

$\hat{\mathbf{w}}_1$ — оптимальные оценки в задаче (40)–(41), $\hat{\mathbf{w}}_2$ — оптимальные оценки в задаче (47)–(49);

$$\Pi = \{\mathbf{w} : \mathbf{0} \leqslant \mathbf{w} \leqslant \mathbf{1}\} \supset \tilde{\Pi} = \{\mathbf{w} : \langle \mathbf{p}, \mathbf{1} \rangle, \langle \mathbf{q}, \mathbf{1} \rangle\}. \tag{54}$$

Терема 2. *Между максимальными значениями энтропии в задачах* (40)–(41) *и* (47)–(49) *имеет место следующее соотношение*

$$H(\hat{w}_1) \leqslant H(\hat{w}_2),$$

причем равенство достигается только в случае $\hat{\mathbf{w}}_1 = \mathbf{w}_{abs}$, *где* $\mathbf{w}_{abs}$ *— точка абсолютного максимума энтропии.*

Доказательство. Энтропия (40) является строго вогнутой функцией с единственным абсолютным максимумом в точке $\mathbf{w}^*$. Обозначим $\varrho_1(\mathbf{w}^*, \mathbf{w}_1)$ и $\varrho_1(\mathbf{w}^*, \mathbf{w}_2)$ — евклидовы расстояния между точками абсолютного максимума и точками, соответствующими оптимальным оценкам в задачах (40)–(42) и (47)–(49). В силу (54) между указанными расстояниями имеет место очевидное соотношение:

$$\varrho_1(\mathbf{w}^*), \mathbf{w}_1) \geqslant \varrho_1(\mathbf{w}^*), \mathbf{w}_2). \tag{55}$$

Из строгой вогнутости энтропии (53) следует утверждение теоремы.

Литература

1. *Кендал М. Дж.*, *Стьюарт А.* Статистические выводы и связи. М.: Наука, 1973. С. 897.
2. *Крамер Г.* Математические методы статистики. М.: Мир, 1975.
3. *Golan A.*, *Judge G.*, *Miller D.* Maximum Entropy Econometrics: Robust Estimation with Limited Data. 1996. New York, John Wiley & Sons.
4. *Ширяев А. Н.* Стохастическая финансовая математика. МАИК «Наука», 2002.
5. *Вапник В. Н.*, *Червоненкис А. Я.* Теория распознавания образов. М.: Наука, 1974. С. 416.
6. *Racine J.*, *Maasoumi E.* «A versatile and robust metric entropy test of time-reversibility, and other hypotheses» // Journal of Econometrics. 2007. V. 138. P. 547–567.
7. *Huber P. J.* Robust Statistics. John Willey & Sons, 1984.
8. *Golan A.* Information and Entropy Econometrics — a Review and Synthesis // Foundation and Trends in Econometrics. 2006. V. 2. № 1–2. P. 1–145.
9. *Boltzmann L.* On the link between the second beginning of mechanical calory theory and probability theory in theorems of thermal equilibrium, 1877 // Избр. труды «Классики науки». 1984. С. 190–236.
10. *Ландау Л. Д.*, *Лифшиц Е. М.* Статистическая физика. М.: Наука, 1964.
11. *Popkov Yu. S.* Macrosystems Theory and its Applications? Lecture Notes in Control and Information Sciences, 203. Springer, London, 1995. P. 324.
12. *Kullback S.*, *Leibler R. A.* On information and Sufficiency // Ann. of Math. Statistics. 1951. V. 22(1). P. 79–86.
13. *Jaynes E. T.* Information Theory and Statistical Mechanics // Physics Review. 1957. V. 106. P. 620–630.
14. *Shannon C.* Communication Theory of Secrecy Systems // Bell System Technical Journal. 1949. V. 28(4). P. 656–715.

Попков Алексей Юрьевич. К. т. н., с. н. с. ИСА РАН. Окончил МГУ в 2002 г. Количество печатных работ: 19. Область научных интересов: математическое моделирование, высокопроизводительные вычисления, параллельные алгоритмы, распределенные вычислительные системы. E-mail: apopkov@isa.ru

Попков Юрий Соломонович. Директор ИСА РАН, д. т. н., профессор. Окончил Московский энергетический институт в 1960 г. Количество печатных работ: 165. Область научных интересов: системный анализ, математическое моделирование. E-mail: popkov@isa.ru