

УДК 519.86 (075.8)
ББК 22.18я73

**РЕЦЕНЗИЯ НА КНИГУ Ф.Ф. ПАЩЕНКО
«ВВЕДЕНИЕ В СОСТОЯТЕЛЬНЫЕ МЕТОДЫ
МОДЕЛИРОВАНИЯ СИСТЕМ.
ЧАСТЬ 2. ИДЕНТИФИКАЦИЯ НЕЛИНЕЙНЫХ
СИСТЕМ». М.: ФИНАНСЫ И СТАТИСТИКА, 2007.
288 С. ISBN 5-279-03042-2**

К.Р. Чернышев
(Учреждение Российской академии наук
Институт проблем управления им. В.А. Трапезникова
РАН, Москва)

Данная работа носит методологический характер и посвящена анализу некоторых методов идентификации и моделирования стохастических систем, представленных в одном из учебных пособий (автор – Ф.Ф. Пащенко) издательства «Финансы и статистика» (г. Москва, 2007 г.) и поддержанного Российским фондом фундаментальных исследований (проект 05-08-50336а) и Программой № 15 фундаментальных исследований Отделения энергетики, машиностроения, механики и процессов управления Российской академии наук. Рассмотренные вопросы связаны с применением теоретико-информационных методов и методов теории оптимальных систем. Показано, что проанализированные положения либо лишены содержательного смысла, либо вообще представляют собой заблуждения, что, учитывая учебный характер данной книги, следует рассматривать как целенаправленные действия, направленные на снижение уровня подготовки отечественных специалистов.

Ключевые слова: математическое моделирование, идентификация систем, оптимальность, взаимная информации, энтропия

1. Введение

Решение задачи идентификации систем всегда основано на применении тех или иных мер зависимости случайных величин (процессов), идет ли речь о представлении моделей исследуемых систем с помощью входо-выходного отображения или в пространстве состояний. Наиболее часто в качестве такой меры выступают традиционные линейные ковариационные или корреляционные зависимости, использование которых непосредственно вытекает из самой постановки задачи идентификации на основе среднеквадратического критерия. Их основным достоинством является удобство использования, включая как возможность построения явных аналитических выражений для определения искомых характеристик, так и относительную простоту построения их оценок, в том числе и на основе наблюдения зависимых данных. Однако главным недостатком мер зависимости, основанных на линейной корреляции, является, как известно, возможность их обращения в нуль даже в случае существования детерминированной зависимости между парой исследуемых переменных¹.

Именно на преодоление этого недостатка направлено использование в задачах идентификации более сложных, нелинейных, мер зависимости, примером чему является глава 7 «Определение и моделирование закономерностей по экспериментальным данным» книги [1], в которой многие результаты потребовали всестороннего рассмотрения и анализа, который и был как в работе [2], так и в целом ряде других. Однако из результатов работы [2] и последующих автором главы 7 книги [1] не было сделано соответствующих выводов, что в итоге и привело к повторению целого ряда заблуждений из главы 7 книги [1] в книгах [3, 4], являющихся двухтомным учебным пособием (рис. 1), допущенным «УМО в области прикладной математики и управления качеством» в качестве учебного пособия для студентов

¹ Райбман Н.С. (ред.) Дисперсионная идентификация. М.: Наука, 1981. 320 с.;
Rényi A. On measures of dependence // Acta Math. Hung. 1959. Vol. 10, No. 3-4. P. 441-451.

высших учебных заведений, обучающихся по направлению 230400 «Прикладная математика» специальности 230401 «Прикладная математика». При этом материал книг [3, 4], хотя и не содержит никаких новых научных результатов, подготовлен при поддержке Российского фонда фундаментальных исследований (инициативный проект 05-08-50336а), что тем более обуславливает привлечение к нему пристального внимания. Его первая часть, книга [3], подробно проанализирована в работе [5].



Рис. 1

В качестве «мостика» от первого ко второму тому данного учебного пособия, книге [4], можно рассматривать следующее утверждение. «Учитывая, что начальные условия, параметры, входящие в уравнения, описывающие систему, входные и выходные сигналы системы нельзя задать или измерить с абсолютной точностью, можно сделать вывод, что теория моделирования и построения эмпирических законов и закономерностей включает в себе «фундаментальную неопределенность», – отмечается в книге [3] на стр. 76. И на той же странице автор книги [3] делает следующий «неожиданный» вывод: «фундаментальная неопределенность», вызванная влиянием помех и погрешностей измерений параметров системы, с течением времени постепенно меняется и может перекрыть всю область возможных состояний и движений системы» (стр. 76-77). Слово «неожиданный» здесь взято в кавычки, поскольку, с одной стороны, данный вывод автора книги [3] нет возможности понять, по

крайней мере, с точки зрения норм русского языка, а, с другой стороны, нечто подобное ранее уже имело место у автора книги [3], а именно: «... энтропия замкнутых и некоторых других систем со временем возрастает и, в конечном счете, может перекрыть всю область возможных состояний и движений системы» [1, стр. 455].

Здесь и далее для наглядности и упрощения восприятия, прямые цитаты из книг [1, 3, 4] выделены серым цветом.

Очевидно в силу подобных «перекрытий» автор книги [3] ставит перед студентами проблему: «можно ли уменьшить влияние помех и погрешностей измерений на теоретические выводы и определение законов и закономерностей? Можно ли увеличить количество знаний и информации, извлекаемых из экспериментальных данных и предыдущих знаний?» (стр. 77). При этом в книге [3] «для ответа на этот вопрос ставится и решается задача построения экспериментальных зависимостей по критерию максимума информации об исследуемой системе. Рассматриваются некоторые возможности и примеры построения предфильтров знаний из экспериментальных данных на основе метода функциональных преобразований» (стр. 77).

Ответ на вопрос как «ставится и решается задача ... по критерию максимума информации», какие возможности следуют из «построения предфильтров знаний» (и, соответственно, что ждет студентов) изложены в книге [4], являющейся предметом рассмотрения настоящей работы.

2. Теоретико-информационные критерии идентификации

Среди многочисленных «общеметодологических» замечаний в книге [3] можно найти и замечание о том, что рассмотренный в книге метод «предполагает также эвристический подход. Например, использование логарифмических преобразований в качестве операторов B и C приводит к информационным критериям построения идентифицируемых моделей и применения мощного инструмента теории информации [1, 6, 7]» (стр. 301). Во-первых, нигде в [1, 6, 7] не представлены модели вида

$\log y = A \log x$, а, во-вторых, нет никаких оснований для утверждения, что данная модель приводит к применению информационного критерия идентификации. Тем не менее, поскольку «применение мощного инструмента теории информации» заняло значительное место в книге [4], анализ его изложения и составляет предмет настоящего раздела.

2.1. «ОБЩИЙ» СЛУЧАЙ

В книге [4] (раздел 2.2.1) рассматривается идентифицируемый объект, который «описывается уравнением

$$(2.1) \quad Y_t = F(X_t, t),$$

где Y_t – случайный сигнал (случайная величина) на выходе объекта, X_t – случайный сигнал (случайная величина) на входе объекта, t – время, $F: X_t \rightarrow Y_t$ – отображение, описывающее объект» (стр. 95-96 в [4]).

Оптимальная модель объекта ищется «из условия, чтобы выход модели Y_M содержал максимум информации о выходе объекта Y_t . В этом случае критерий идентификации может быть в виде:

$$(2.2) \quad L = \mathbf{M} \left\{ \log \frac{p\left(\frac{y}{y_M}\right)}{p(y)} \right\} = \max,$$

где $p\left(\frac{y}{y_M}\right)$ – условная плотность распределения выходной величины Y_t относительно выходной величины модели Y_M , $p(y)$ – безусловная плотность распределения выходной величины Y_t ,

$$(2.3) \quad \mathbf{M} \left\{ \log \frac{p\left(\frac{y}{y_M}\right)}{p(y)} \right\} = I\{Y_t, Y_M\}$$

– количество информации, содержащееся в Y_M об Y_t , где $\mathbf{M}\{\cdot\}$ – символ математического ожидания. Критерий (2.2) содержит требование максимизации этой информации.

Пусть уравнение модели имеет вид

$$(2.4) \quad Y_M = f(x),$$

где $f: R^1 \rightarrow R^1$; R^1 – одномерное пространство (числовая ось), т.е. выход модели есть некоторая функция от входного сигнала.

Тогда информация, содержащаяся в выходе модели о выходном сигнале объекта, определяется выражением

$$(2.5) \quad I\{Y, Y_M\} = I\{Y, f(x)\}.$$

Справедливы следующие утверждения.

Утверждение 2.1. Пусть идентифицируемый объект (2.1) является статическим преобразователем. Случайные величины на входе X и выходе Y описываются произвольными законами распределения с конечными моментными характеристиками $M_i \leq d < \infty$, $i = 1, 2, \dots, n$.

Пусть в качестве модели (2.4), описывающей объект, взята однозначная функция, определенная на R^1 , со значениями в R^1 : $f: R^1 \rightarrow R^1$, R^1 – пространство действительных чисел (числовая ось).

Тогда информация, содержащаяся в выходном сигнале модели о выходном сигнале объекта, удовлетворяет следующему неравенству:

$$(2.6) \quad I\{f(x), Y\} = I\{Y_M, Y\} \leq I\{X, Y\}.$$

Утверждение 2.2. Пусть выполняются условия утверждения 2.1. Пусть функция $f(\cdot)$ – взаимнооднозначная. Тогда неравенство (2.6) превращается в равенство

$$(2.7) \quad I\{Y_M, Y\} = I\{X, Y\}.$$

Следствие 2.1. Пусть выполняются условия утверждения 2.2. Пусть функция $f(\cdot)$ – непрерывная на R^1 . Тогда $f(\cdot)$ – строго монотонная функция и

$$I\{Y_M, Y\} = I\{X, Y\}.$$

Следствие 2.2. Количество информации, содержащейся в выходе модели – Y_M о выходной величине объекта – Y , достигает максимума тогда и только тогда, когда модель (2.4) описывается взаимнооднозначной функцией входной переменной X .

Следствие 2.3. Решение задачи идентификации по критерию максимума информации, содержащейся в модели (2.4) о выходной величине объекта (2.1), в классе однозначных и взаимнооднозначных функций, не единственно» (стр. 96-97 в [4]).

Таким образом, легко видеть, что процитированные утверждения из книги [4] (стр. 96-97) не представляют никакого полезного аппарата для решения задачи идентификации.

В свою очередь, что касается предложенного в книге [4] теоретико-информационного подхода к идентификации вообще, то такой подход нельзя признать сколь-нибудь конструктивным, поскольку в него изначально закладываются либо необходимость знания совместной плотности распределения выходных переменных объекта и модели, либо возможность их (выходных переменных) наблюдения. Но этот второй вариант заведомо невыполним, поскольку задача идентификации как раз и состоит в построении модели и, естественно, ее выход наблюдаться не может. Но и первый вариант не может считаться приемлемым, поскольку требует наличие такого объема априорной информации, при котором собственно задача идентификации уже должна потерять содержательный смысл.

2.2. НОРМАЛЬНОСТЬ СОВМЕСТНОГО РАСПРЕДЕЛЕНИЯ ВЫХОДОВ МОДЕЛИ И ОБЪЕКТА

Однако именно постулирование конкретного вида совместной плотности выходных переменных модели и объекта положено в основу аналитических выводов в книге [4], а именно делается предположение о *нормальности совместного распреде-*

ления выходных переменных модели и объекта. При этом в книге [4] (раздел 2.2.2) выход модели Y_M представляется «в виде

$$(2.8) \quad Y_M = \sum_{i=1}^n b_i f_i(X_i) + m,$$

где $f = (f_1, \dots, f_n)$, – вектор преобразований входных сигналов X , $X = (X_1, \dots, X_n)$ – вектор входных сигналов, m – шум с моментными характеристиками: математическое ожидание $\mathbf{M}\{m\} = 0$, дисперсия $\mathbf{D}\{m\} = D_m$ » (стр. 101 в [4]).

При этом «для упрощения выкладок» рассматривается «модель

$$(2.9) \quad Y_M = \sum_{i=1}^n b_i X_i + m.$$

Модели (2.8) и (2.9) являются эквивалентными, учитывая замену переменных

$$Z_i = f(X_i), \quad i = 1, \dots, n.$$

Таким образом, для произвольного нелинейного объекта ищется модель (2.9), оптимальная по критерию (2.2)» (стр. 102 в [4]).

Замечание. Необходимо отметить среди прочего «странность» моделей (2.8), (2.9), а именно – включение в них ненаблюдаемого шума m , что исключает возможность вычислений значений выходной переменной модели на основе наблюдения входных переменных при тех или иных значениях коэффициентов.

Естественно, в предположении нормальности совместного распределения выходов объекта и модели задача идентификации сводится на основании хорошо известного в литературе соотношений к задаче максимизации коэффициента корреляции между выходами модели и объекта. С содержательной точки зрения предположение о нормальности совместного распределения выходов модели (2.9), эквивалентной модели (2.8), и «произвольного нелинейного объекта», сделанное в книге [4], эквивалентно тому, что, например, заявив некий новый метод обращения матриц, для его применения сделать предположение

о том, что обращаемая матрица является диагональной. Следует отметить также, что предположение о нормальности совместной плотности распределения выходов модели и объекта всегда не выполняется, например, в случае идентификации тождественного преобразователя. Действительно, пусть: вход $X \sim \Phi(x)$; выход $Y \equiv X$; выход модели $Y_M \equiv X$; совместное распределение выхода объекта и выхода модели имеет вид:

$$\begin{aligned} P\{Y < y; Y_M < y_M\} &= P\{X < y; X < y_M\} = \\ &= P\{X < \min(y, y_M)\} = \Phi(\min(y, y_M)). \end{aligned}$$

Следовательно, $p(y, y_M)$ не является нормальной.

В тех же редких случаях, когда предположение о нормальности совместного распределения выходов модели и объекта выполняется (обусловлено структурой модели и объекта, вероятностными свойствами входов), в принципе исчезает сам смысл обращения к подобному информационному критерию, поскольку в данном случае достаточно использовать обычный среднеквадратический критерий (как хорошо известно, в случае нормальности совместного распределения максимальная корреляция линейна и совпадает с обычной).

Известны дополнительные аргументы в пользу научной целесообразности изложенного сначала в книге [1], а за тем и в книге [4] теоретико-информационного подхода². Эти аргументы разделены на две группы, первая из которых касается конструктивности изложенного в книге [4] подхода, а вторая – его «корректности». Однако эти аргументы нельзя признать состоятельными по следующим причинам.

Во-первых, в обоснование конструктивности изложенного в книге [3] (раздел 2.2.2) (книге [1] (раздел 7.5)) подхода рассматривается «классический» критерий идентификации из книги Н.С. Райбмана «Что такое идентификация?» вида (стр. 24, формула (3.3))

² Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

$$(2.10) \quad \mathbf{M} \left\{ \rho \left[y_t, y_t^* \right] \right\} = \min ,$$

или (стр. 25, формула (3.6))

$$\mathbf{M} \left\{ \left[y_t - y_t^* \right]^2 \right\} = \min ,$$

«где y_t^*, y_t – выходы модели и объекта соответственно, $\mathbf{M}\{\cdot\}$ – символ математического ожидания, ρ – функция потерь.

Минимизируя указанные выше критерии при заданной функции потерь, в работах по идентификации получают оценку (модель) идентифицируемой системы³. Но в книге [4] (раздел 2.2.2) такая функция потерь ρ как раз и *не задана*, поскольку имеет вид

$$(2.11) \quad \rho \left[y_t, y_t^* \right] = \log \frac{p_{y_t, y_t^*}(y_t, y_t^*)}{p_{y_t}(y_t) p_{y_t^*}(y_t^*)}$$

и включает как маргинальные $p_{y_t}(y_t)$, $p_{y_t^*}(y_t^*)$, так (что особенно принципиально) и *совместную* $p_{y_t, y_t^*}(y_t, y_t^*)$ плотности распределения выхода объекта и выхода модели соответственно. Конечно, следует заметить, что критерий (2.10) при выборе ρ в соответствии с выражением (2.11) имеет вид

$$(2.12) \quad \mathbf{M} \left\{ \rho \left[y_t, y_t^* \right] \right\} = \max .$$

При этом утверждается, что постановка задачи (2.11), (2.12) аналогична постановке задачи (2.10), «конкретизирована только функция потерь. В качестве этой функции используется количество информации, содержащейся в выходе модели о выходе

³ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

идентифицируемой системы»⁴. Однако, как видно из выражения (2.11), такая аналогия не справедлива.

Во-вторых, отмечается, что в рассматриваемом теоретико-информационном подходе «лишь незначительно усложняется ситуация, имеющая место в известных задачах идентификации, например, по критерию минимума среднеквадратической ошибки и, особенно, при использовании байесовских оценок и подходов. Поскольку для вычисления количества информации дополнительно вводится только операция вычисления логарифма от соответствующих плотностей распределения случайных величин и сигналов»⁵. Но вопрос как раз и заключается в том, что «соответствующие плотности распределения» не известны даже с точностью до параметризованного представления, как минимум это относится к совместному распределению выходов модели (которую еще только предстоит построить) и объекта. И далее указывается, «что строгое теоретическое изложение корреляционных методов идентификации также требует знания соответствующих функций и плотностей распределения входных и выходных сигналов»⁶. Но здесь подменяется сам предмет обсуждения, поскольку речь должна идти о совместном распределении $p_{y_t, y_t^*}(y_t, y_t^*)$ выходов модели и объекта.

В-третьих, приведенные ссылки на «вычислительные аспекты количества информации и энтропии» при реализации изложенных в [4] методов, эти аспекты никак не отражают, поскольку в [4] приводится следующий подход: «Рассмотрим две случайные величины X и Y , для которых в результате N независимых опытов получены следующие значения

⁴ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

⁵ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

⁶ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

$$\xi_1, \xi_2, \dots, \xi_N,$$

$$\eta_1, \eta_2, \dots, \eta_N.$$

Случайные величины X и Y , можно рассматривать как двумерную случайную величину (X, Y) .

По полученным опытным данным составлена таблица, в которой $\hat{p}_{ij}$ обозначает частоту совместного события $X = x_j$ и $Y = y_i$, и может служить оценкой соответствующей вероятности p_{ij} .

	Y_1	...	Y_i	...	Y_a	
X_1						$\hat{q}_1$
$\vdots$						$\vdots$
X_j			$\hat{p}_{ij}$			$\hat{q}_j$
$\vdots$						$\vdots$
X_b						$\hat{q}_1$
	$\hat{p}_1$	...	$\hat{p}_i$	...	$\hat{p}_a$	

Число столбцов этой таблицы, т.е. число различных значений величины Y обозначим через a , а число строк, т.е. число различных значений величины X , через b . При этом

$$\hat{p}_{ij} = \frac{m_{ij}}{N},$$

где m_{ij} – число совместных появлений значений x_j и y_i в N опытах.

$$\sum_i \sum_j \hat{p}_{ij} = 1; \quad \sum_i \hat{p}_{ij} = \hat{q}_j; \quad \sum_j \hat{p}_{ij} = \hat{p}_i.$$

Обозначим через C число различных значений двумерной величины (X, Y) , иными словами – число ненулевых клеток таблицы.

Для нахождения оценки количества информации $\hat{I}\{X, Y\}$, содержащейся в случайной величине X о случайной величине Y , воспользуемся формулой

$$I\{X, Y\} = H\{X\} + H\{Y\} - H\{X, Y\},$$

где $H\{X, Y\}$ совместная энтропия величин X и Y ⁷.

Тогда в качестве оценки количества информации естественно принять величину

$$\hat{I}\{X, Y\} = \hat{H}\{X\} + \hat{H}\{Y\} - \hat{H}\{X, Y\},$$

где

$$\hat{H}\{X\} = -\sum_j \hat{q}_j \log \hat{q}_j; \quad \hat{H}\{Y\} = -\sum_i \hat{p}_i \log \hat{p}_i;$$

$$\hat{H}\{X, Y\} = -\sum_i \sum_j \hat{p}_{ij} \log \hat{p}_{ij}.$$

Тогда

$$\hat{I}\{X, Y\} = -\sum_i \sum_j \hat{p}_{ij} \log \frac{\hat{p}_{ij}}{\hat{p}_i \hat{q}_j} \gg$$

(стр. 127-129 в [4]).

Таким образом, для применения данного метода оценивания взаимной информации совершенно необходимо *наблюдение* выборки значений выхода модели y_i^* (в то время как эта модель не известна и, соответственно, значения выхода модели ни наблюдаться, ни вычисляться не могут).

Замечание. Имеется в виду конечно выход модели, *оптимальный в смысле заданного критерия*. «Просто» выход модели, безусловно, может наблюдаться, поскольку в качестве выхода модели может (формально) быть принят результат любого априори выбранного преобразования входа. Сказанное, однако, не относится к моделям (2.8), (2.9), поскольку они содержат ненаблюдаемый шум t , и, следовательно, значение выходной переменной Y_M в формулах (2.8), (2.9) не может быть вычислено даже при заданных значениях коэффициентов b_i , $i = 1, \dots, n$.

⁷ Вентцель Е.С. Теория вероятностей. М.: Физматгиз, 1962 г.

Таким образом, все вышесказанное свидетельствует о том, что конструктивность представленного в разделе 2.2.2 книги [4] подхода к идентификации по критерию максимума количества информации не продемонстрирована и не подтверждена.

В-четвертых, выражается мнение о том, что «предположение о нормальном распределении входных и выходных сигналов объекта и модели достаточно традиционно в работах по идентификации систем»⁸ (при этом делаются ссылки на соответствующие работы Н.С. Райбмана, П. Эйкхоффа, Я.З. Цыпкина). Но такое утверждение слишком расплывчато в рамках данного обсуждения, поскольку в нем не указывается, какие именно виды распределений (совместные, маргинальные) имеются в виду. Если здесь подразумевается, в том числе, и совместное распределение выхода объекта и выхода модели, то тогда очевидно вряд ли допустимо говорить о *традиционности* предположения о совместной нормальности (совместное распределение является следствием множества факторов: структуры системы и модели, статистических особенностей входов и т.п.). Можно, в частности, записать следующее формальное выражение для совместной плотности распределения $p_{y,y^*}(y, y^*)$ выхода объекта и выхода модели (которое вытекает из хорошо известного соотношения для совместной плотности распределения преобразования случайного вектора):

$$(2.13) \quad p_{y,y^*}(y, y^*) = \int_{(z_{n+1}, \varphi(z_1, \dots, z_n)) \in C} \int_{x_1, \dots, x_n, y} p_{x_1, \dots, x_n, y}(z_1, \dots, z_n, z_{n+1}) \frac{dS_{n-1}}{\sqrt{\sum_{i_1 < i_2} \left[\frac{D(z_{n+1}, \varphi)}{D(z_{i_1}, z_{i_2})} \right]^2}}.$$

Формула (2.13) записана для модели объекта, представленной в виде

$$(2.14) \quad y^* = \varphi(x_1, \dots, x_n),$$

⁸ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

где $x_1, \dots, x_n$ – (обобщенные) входы объекта, y – выход объекта, $p_{x_1, \dots, x_n, y}(z_1, \dots, z_n, z_{n+1})$ – совместная плотность распределения входов и выхода объекта. Интеграл в правой части выражения (2.13) – поверхностный интеграл по поверхности размерности $(n-1)$, определяемой системой уравнений

$$\begin{cases} \varphi(z_1, \dots, z_n) = y^* \\ z_{n+1} = y \end{cases},$$

а

$$\frac{D(z_{n+1}, \varphi)}{D(z_{i_1}, z_{i_2})} = \begin{vmatrix} \frac{\partial z_{n+1}}{\partial z_{i_1}} & \frac{\partial z_{n+1}}{\partial z_{i_2}} \\ \frac{\partial \varphi}{\partial z_{i_1}} & \frac{\partial \varphi}{\partial z_{i_2}} \end{vmatrix}$$

– якобиан функций z_{n+1}, φ по переменным z_{i_1}, z_{i_2} .

Из формул (2.13), (2.14), в частности, следует тот хорошо известный факт, что совместное распределение выходов модели и объекта является нормальным при нормальности распределения $p_{x_1, \dots, x_n, y}(z_1, \dots, z_n, z_{n+1})$ и линейности функции $\varphi(x_1, \dots, x_n)$, описывающей модель (2.14). Поэтому в любых более общих случаях, например, рассматриваемых в рамках того же теоретико-информационного подхода в книге [4] (раздел 2.2.2) дисперсионной модели

$$(2.15) \quad Y_M = \sum_{i=1}^n b_i \left\{ Y / X_i \right\} + m,$$

или динамической модели

$$(2.16) \quad Y_M(t) = \int_{-\infty}^{\infty} g(\tau) X(t-\tau) d\tau + m(t)$$

для априорного предположения о нормальности совместного распределения выходов модели и «произвольного нелинейного объекта» нет никаких оснований, и такое предположение есть

лишь заведомое упрощение исходной задачи, ведущее к выхолащиванию ее сущности.

В-пятых, в отношении заявленных целей раздела 2.2.2 книги [4] (в основе которых лежит связь информационного и корреляционного критериев в случае совместного нормального распределения выходов объекта и модели) необходимо заметить, что хорошо известно следующее соотношение для информации многомерного нормального распределения

$$(2.17) \quad I(\mathbf{Y}, \mathbf{X}) = -\frac{1}{2} \ln \left(\frac{\det(Q_{\mathbf{Z}\mathbf{Z}})}{\det(Q_{\mathbf{Y}\mathbf{Y}})\det(Q_{\mathbf{X}\mathbf{X}})} \right).$$

В формуле (2.17) приняты следующие обозначения: $\mathbf{Z}$ – нормально распределенный случайный вектор с ковариационной матрицей $Q_{\mathbf{Z}\mathbf{Z}}$, $\dim \mathbf{Z} = n + m$, причем $\mathbf{Z} = \begin{pmatrix} \mathbf{X}^T & \mathbf{Y}^T \end{pmatrix}^T$, где $\dim \mathbf{X} = n$, $\dim \mathbf{Y} = m$, и $Q_{\mathbf{X}\mathbf{X}}$, $Q_{\mathbf{Y}\mathbf{Y}}$ – ковариационные матрицы случайных векторов $\mathbf{X}$ и $\mathbf{Y}$ соответственно. Таким образом, в разделе 2.2.2 книги [4] просто используется простейший частный случай соотношения (2.17), то есть когда $\mathbf{X}$ и $\mathbf{Y}$ являются скалярами (принимая во внимание тот факт, что этот частный случай и в книге [4], и ранее⁹ представлен в виде отдельной леммы с приведением соответствующего доказательства, естественно предположить, что формула (2.17) автору [4] не известна).

Более того, в рамках «выявления» связи между информационным и корреляционным критериями в случае нормальности совместного распределения выходов объекта и модели¹⁰, в работе [4] можно было бы «установить» и формально более общую (но столь же тривиальную) «связь», а именно – тождество информационного критерия и критерия, основанного на максими-

⁹ Дургарян И.С., Пашенко Ф.Ф. Идентификация объектов по критерию максимума количества информации // Автоматика и телемеханика. 2001. № 7. С. 91-102.

¹⁰ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

зации *обобщенной* корреляции между выходами объекта и модели, например, вида

$$\mathbf{corr}\{\varphi(y_t), y_t^*(\vartheta)\} \rightarrow \sup_{\varphi, \vartheta},$$

где $\varphi(\cdot)$ – борелевская функция, ϑ – вектор параметров модели, подлежащий идентификации.

В качестве одного из свойств рассмотренного информационного метода идентификации указывается то, что он также как и корреляционные методы позволяет получить *явные* оценки параметров, например линейных систем¹¹. Однако справедливость и этого утверждения следует поставить под сомнение. Действительно, для определения параметров b_i , $i = 1, \dots, n$ моделей (2.9), (2.15) и весовой функции $g(\tau)$ модели (2.16) в книге [4] (стр. 104-105) выписаны следующие уравнения:

$$(2.18) \quad K(X_q, Y)\lambda(b_1, \dots, b_n) = \sum_{i=1}^n b_i K(X_i, X_q) + K(X_q, m),$$

$$\lambda(b_1, \dots, b_n) = \frac{\sum_{i=1}^n \sum_{j=1}^n b_i b_j K(X_i, X_j) + 2 \sum_{i=1}^n b_i K(X_i, m) + D_m}{\sum_{i=1}^n b_i K(X_i, Y) + K(Y, m)},$$

$$q = 1, \dots, n$$

– для модели (2.9),

$$(2.19) \quad \theta_{YX_q} \lambda(b_1, \dots, b_n) = \sum_{i=1}^n b_i \theta_{YYX_i X_q} + \theta_{YmX_q},$$

¹¹ Дургарян И.С., Пашенко Ф.Ф. Ответ на статью К.Р. Чернышева «Об одной статье по идентификации стохастических объектов» // Автоматика и телемеханика. 2002. № 4. С. 185-187.

$$\lambda(b_1, \dots, b_n) = \frac{\sum_{i=1}^n \sum_{j=1}^n b_i b_j \theta_{YYX_i X_j} + 2 \sum_{i=1}^n b_i \theta_{YmX_i} + D_m}{\sum_{i=1}^n b_i \theta_{YX_i} + K(Y, m)},$$

$$q = 1, \dots, n$$

– для модели (2.15), и

$$(2.20) \quad K_{YX}(\tau) \lambda(g(\tau)) = \int_{-\infty}^{\infty} g(\tau_1) K_{XX}(\tau - \tau_1) d\tau_1 + K_{Xm}(\tau),$$

$$\lambda(g(\tau)) =$$

$$= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tau_1) g(\tau_2) K_{XX}(\tau_2 - \tau_1) d\tau_1 d\tau_2 + 2 \int_{-\infty}^{\infty} g(\tau) K_{Xm}(\tau) d\tau + D_m}{\int_{-\infty}^{\infty} g(\tau) K_{YX}(\tau) d\tau + K_{Ym}(0)},$$

– для модели (2.16).

В формуле (2.18) $K(\cdot, \cdot)$ – ковариация соответствующих случайных величин; в формуле (2.19)

$$\theta_{YYX_i X_j} = \mathbf{M} \left\{ \left(\mathbf{M} \left(\frac{Y}{X_i} \right) - \mathbf{M}Y \right) \left(\mathbf{M} \left(\frac{Y}{X_j} \right) - \mathbf{M}Y \right) \right\},$$

$$\theta_{YmX_i} = \mathbf{M} \left\{ \left(\mathbf{M} \left(\frac{Y}{X_i} \right) - \mathbf{M}Y \right) \left(\mathbf{M} \left(\frac{m}{X_i} \right) - \mathbf{M}m \right) \right\},$$

$$\theta_{YX_i} = \mathbf{M} \left\{ \left(\mathbf{M} \left(\frac{Y}{X_i} \right) - \mathbf{M}Y \right)^2 \right\}$$

– соответствующие корреляционные отношения (с точностью до нормировки) (дисперсионные коэффициенты); в формуле (2.20) $K_{\bullet\bullet}(\cdot)$ – ковариационная функция соответствующих случайных процессов. Как видно, уравнения (2.18)-(2.20) являются нели-

нейными, и их явные решения – явные выражения для оценок искомых характеристик исследуемых систем – не представлено.

Столь же неуместным выглядит и утверждение в книге [4] (стр. 105) о том, что в частном случае при отсутствии помехи и при $\lambda(b_1, \dots, b_n) = 1$ ($\lambda(g(\tau)) = 1$) уравнения (4.18)-(4.20) представляют собой соответственно нормальное уравнение, дисперсионное уравнение и уравнение Винера-Хопфа. Формально при $\lambda(\cdot) = 1$ данные уравнения, безусловно, являются таковыми, однако сама величина $\lambda(b_1, \dots, b_n) (\lambda(g(\tau)))$, представляющая собой отношение дисперсии выхода модели к ковариации выходов модели и объекта, зависит от неизвестных параметров (весовой функции) модели, являющихся решением этих уравнений, которое, в свою очередь, еще только предстоит найти. Таким образом, данное утверждение в книге [4] (стр. 105) представляет собой тавтологию и лишено содержательного смысла.

2.3. ОБ ИДЕНТИФИКАЦИИ ПО ИНФОРМАЦИОННОМУ КРИТЕРИЮ ПРИ ОГРАНИЧЕНИЯХ НА ПАРАМЕТРЫ

Все вышесказанное в отношении изложенного в книге [4] (разделы 2.2.1-2.2.2) теоретико-информационного подхода относится и к изложенному в разделе 2.2.3 книги [4] теоретико-информационному подходу к решению задачи идентификации при ограничениях на параметры модели. Более того, представленный в разделе 2.2.3 книги [4] метод решения сам по себе является неприемлемым. А именно, рассматривается объект (2.1), модель которого ищется в виде (2.9) в соответствии с критерием (2.2) при условии априорной известности того, что совместная плотность распределения выходов объекта (2.1) и модели (2.9) является нормальной. При этом на параметры модели (2.9) накладывается ограничение вида

$$(2.21) \quad \sum_{i=1}^n b_i C_i = d.$$

Такая задача с ограничениями решается с помощью метода множителей Лагранжа, что приводит к системе нелинейных уравнений, напоминающей систему (2.18), однако – более слож-

ную, решения которой b_i^* , $i = 1, \dots, n$ зависят от множителя Лагранжа λ (здесь необходимо обратить внимание на простое совпадение обозначений с коэффициентом $\lambda(b_1, \dots, b_n)$ в системе уравнений (2.18)):

$$(2.22) \quad b_i^* = b_i^*(\lambda).$$

По замыслу автора [4], полученные аналитические выражения (2.22) должны быть подставлены в уравнение ограничений (2.21) для определения множителя Лагранжа λ . Но вопрос как раз и состоит в том, что такие аналитические выражения получить не возможно, поскольку уравнения, образующие систему для определения $b_i^*(\lambda)$, представляют собой нелинейные алгебраические уравнения *четвертой* степени общего вида:

$$(2.23) \quad 0 = \lambda C_q + \frac{1}{D_Y \left(\sum_{i=1}^n D_{X_i} b_i^2 + D_m \right)^2} \times \\ \times 2 \left\{ K_{YX_q} \left(\sum_{i=1}^n K_{YX_i} b_i + K_{Ym} \right) \left(\sum_{i=1}^n D_{X_i} b_i^2 + D_m \right) - \right. \\ \left. D_{X_q} b_q \left(\sum_{i=1}^n K_{YX_i} b_i + K_{Ym} \right)^2 \right\}, \quad q = 1, \dots, n,$$

где $D_{\bullet}$ – дисперсии соответствующих случайных величин. В данной задаче идентификации с ограничениями в книге [4] (раздел 2.2.3) приняты предположения о попарной некоррелированности входных переменных X_i , $i = 1, \dots, n$ и о некоррелированности входных переменных и помехи m .

Система уравнений (2.23) получена как условие экстремума по

$$b = (b_1, \dots, b_n)^T$$

целевой функции

$$V_{\lambda}(b_1, \dots, b_n) = \frac{\left(\sum_{i=1}^n K_{YX_i} b_i + K_{Ym} \right)^2}{D_Y \left(\sum_{i=1}^n D_{X_i} b_i^2 + D_m \right)} + \lambda \left(\sum_{i=1}^n b_i C_i - d \right),$$

что уже само по себе выглядит достаточно странно, поскольку метод множителей Лагранжа для того и служит, чтобы от задачи условной оптимизации перейти к задаче безусловной оптимизации по расширенному (за счет множителей Лагранжа) набору аргументов оптимизации. В данном случае, таким расширенным набором является, естественно, $b_1, \dots, b_n, \lambda$.

Таким образом, поскольку сама возможность *аналитического* решения в общем случае *системы* полиномиальных алгебраических уравнений *четвертой степени* является крайне сомнительной (и никаких ссылок на такую возможность в книге [4] не представлено), то следует заключить, что предложенный в книге [4] (раздел 2.2.3) метод не дает никаких оснований для вывода о существовании возможности получения решения представленной задачи идентификации при наличии ограничений.

2.4. «ОБОБЩЕННАЯ» ЭНТРОПИЯ

В разделе 2.1 книги [4] констатируется, «что энтропия замкнутых и некоторых других систем со временем возрастает и, в конечном счете, может перекрыть всю область возможных состояний и движений системы. С другой стороны, отмечалось, что объем информации, содержащейся в научной теории, превосходит объем информации, содержащейся в элементарных экспериментальных данных.

Каждый из этих факторов имеет свое объяснение. Например, научные законы, хотя они основаны на эмпирических наблюдениях, изобретены человеческим воображением и открыты человеком. Таким образом, источником избыточной информации являются плоды творческой работы и знаний ученого.

Однако достойного математического объяснения указанные факты до сих пор не имеют, хотя существующее состояние тео-

рии информации и негэнтропийного принципа позволяют это сделать.

Действительно, определение энтропии в негэнтропийном принципе в статистической физике и теории информации дается с точностью до знака и постоянного множителя, выражениями вида

$$(2.24) \quad H\{Y\} = -\int_{-\infty}^{\infty} (p(y)) \log(l_Y p(y)) dy,$$

$$(2.25) \quad H\{Y\} = -\sum_1^S p_i \log p_i,$$

где l_Y обуславливает начало отсчета по шкале энтропий.

В этом случае учитываются только состояния системы и вероятности, соответствующие этим состояниям.

При этом никак не учитываются ни скорость изменения состояния, ни ускорение движения, ни скорость обмена информацией и энтропией между элементами системы и внешней средой и многое другое, что могло бы объяснить нам скорость распространения и возрастания неопределенности, избыточность информации и др.» [4] (стр. 91-92). Данный текст полностью присутствует и в [1] (стр. 455), с заменой «др.» на «многое другое».

Очевидно, с целью «достойного математического объяснения» в книге [4], стр. 92-93 представлены «понятия динамической и общей энтропии, которые позволяют учесть некоторые из приведенных выше факторов. Движения, скорости, перемещения и другие изменения системы и ее отдельных частей и элементов можно рассматривать как некоторые преобразования состояний, событий, элементов и системы в целом. Учитывая это, можно ввести следующее понятие динамической энтропии» [4] (стр. 92-93):

$$(2.26) \quad H^{\phi}\{Y\} = -\int_{-\infty}^{\infty} (p(B(y))) \log(l_Y p(B(y))) dy,$$

«где элементы $\{By\}$ образуют множество всех состояний $\{BY\}$, которое получается в результате действия произвольных преобразований B на исходную случайную величину Y » [там же].

Очевидно под влиянием критики, прозвучавшей в работе [2] в адрес книги [1], в книге [4] автор постарался исключить

Пример. Пусть преобразование стандартной гауссовской случайной величины сводится к умножению на скаляр $BY = \alpha \cdot Y$, $\alpha \in [0; 1]$. График получаемого при этом подынтегрального выражения с внесением знака «минус» под знак интеграла для различных значений $\alpha \in [0; 1]$ представлен на рис. 2. Легко видеть, что при любом $\alpha \in (0; 1]$ интеграл в (2.26) существует, и чем меньше α , тем больше значение $H^\phi\{Y\}$.

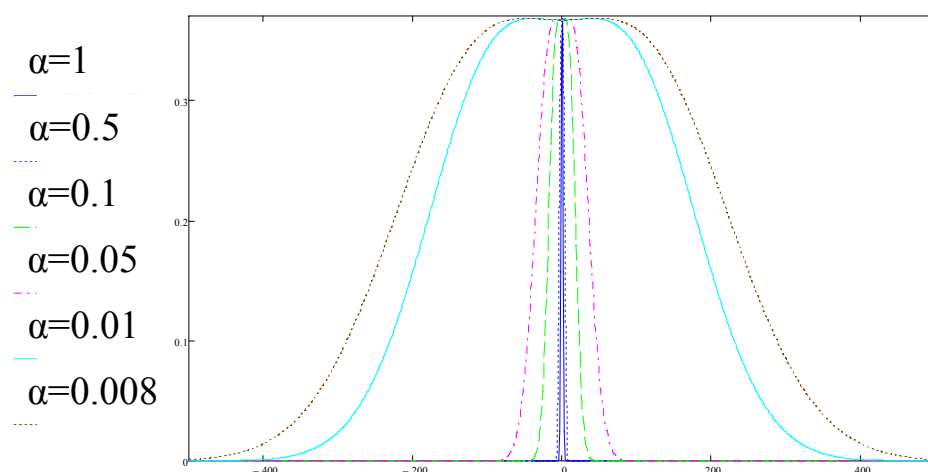


Рис. 2. График подынтегрального выражения в (2.26) с внесением знака «минус» под знак интеграла для стандартной гауссовской случайной величины Y при ее преобразовании вида $BY=\alpha Y$ для различных значений α из интервала $[0, 1]$.

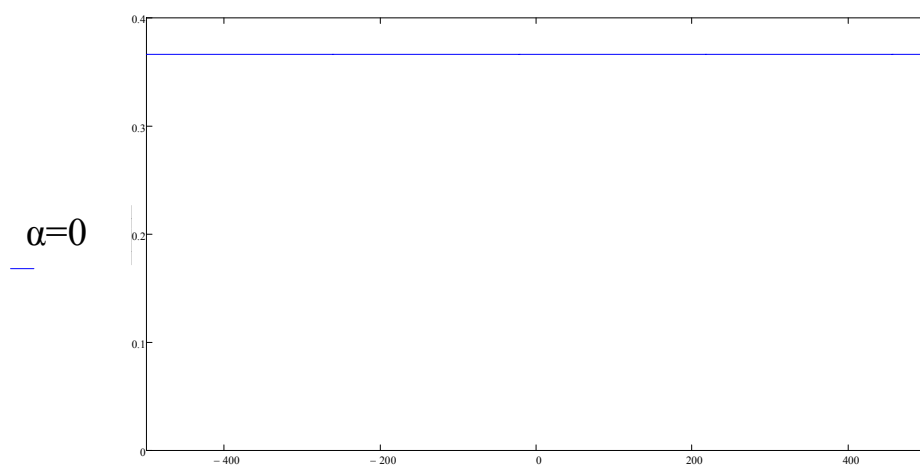


Рис. 3. График подынтегрального выражения в (2.26) с внесением знака «минус» под знак инте-

грала для стандартной гауссовской случайной величины Y при ее преобразовании вида $BY = \alpha Y$ при $\alpha \neq 0$.

При $\alpha = 0$ интеграл в (2.26) расходится (график получаемого при этом подынтегрального выражения с внесением знака «минус» по знак интеграла для $\alpha = 0$) представлен на рис. 3). Вопрос: начиная с какого значения α «динамическая энтропия» теряет смысл?

Далее в книге [3, стр. 93] представлены следующие определения:

обобщенная динамическая энтропия

$$(2.27) \quad H^{\phi^0} \{Y\} = - \int_{\{B\}} \int_{-\infty}^{\infty} (p(B(y))) \log(l_Y p(B(y))) dy d\mu(B),$$

общая энтропия

$$(2.28a) \quad H^0 \{Y\} = H \{Y\} + H^{\phi^0} \{Y\},$$

максимальная энтропия

$$(2.28б) \quad H^{\phi^{\max}} \{Y\} = \max_B H^{\phi} \{Y\}.$$

В приведенных формулах « $\mu(B)$ – мера, определенная на множестве всех отображений (функций) $\{B\}$. При этом множество $\{B\}$ не содержит тождественных преобразований» [4, стр. 93]. Причем о такой существенной детали, как, например, определение меры $\mu(B)$ в формуле (2.27) в книге [4] умалчивается. Но в книге [4] (стр. 94) отмечается, что максимальная энтропия не превосходит обобщенную динамическую и общую энтропии:

$$H^{\phi^{\max}} \{Y\} \leq H^{\phi^0} \{Y\}, \quad H^{\phi^{\max}} \{Y\} \leq H^0 \{Y\},$$

и что введенные понятия энтропии обладают всеми свойствами обычной энтропии.

В книге [4] отмечается, что определение обобщенной динамической энтропии «в случае, если множество $\{B\}$ включает в себя тождественные преобразования, обобщает определение общей энтропии. Однако выражение (2.28a) более привлекательно с той точки зрения, что в этом случае мы можем придать

некоторый физический смысл энтропии. Так по аналогии с энергией обычную энтропию (2.24), (2.25) мы можем называть потенциальной энтропией. Введенная же динамическая энтропия является как бы аналогом кинетической энергии» [4] (стр. 93).

«Аналогично вводятся понятия обобщенной условной $H^\phi\{Y/X\}$ и совместной $H^\phi\{Y, X\}$ энтропии, а также обобщенной информации $I^\phi\{Y, X\}$. На основе этих понятий можно ввести и другие аналогичные характеристики систем» [4] (стр. 94).

Но уже приведенный пример (рис. 2, 3) показывает, что постулированная в книге [4] максимальная энтропия $H^{\phi \max}\{Y\} = \max_B H^\phi\{Y\}$ обращается в бесконечность. Легко показать, что и в общем случае, максимальная энтропия не существует. Действительно, согласно [4]: «Элементы $\{By\}$ образуют множество всех состояний $\{BY\}$, которое получается в результате действия произвольных преобразований B на исходную случайную величину Y ». Если так, то сузим область поиска максимума в (2.28б) до преобразований вида $BY = \alpha \cdot Y$, $\alpha \in R^1$. Тогда на основании формул (2.27), (2.28б)

$$H^{\phi \max}\{Y\} = \max_B H^\phi\{Y\} \geq \max_{\alpha \in R^1} \left\{ - \int_{-\infty}^{\infty} (p(\alpha \cdot y)) \log(l_Y p(\alpha \cdot y)) dy \right\}.$$

Quod erat demonstrandum.

Следовательно, представленный в книге [4] «обобщенный» информационно-энтропийный подход ошибочен. Можно лишь добавить, что известно множество статей¹² и, соответственно,

¹²

Abe, S. (2004). "Stability analysis of generalized entropies and q-exponential distributions", *Physica D: Nonlinear Phenomena*, vol. 193, no. 1-4, pp. 84-89.

Adimurthi, Mishra, S., and G.D. Veerappa Gowda (2007). "Conservation law with the flux function discontinuous in the space variable – II: Con-

-
- vex-concave type fluxes and generalized entropy solutions”, *Journal of Computational and Applied Mathematics*, vol. 203, no. 2, pp. 310-344.
- Almeida, M.P. (2001). “Generalized entropies from first principles”, *Physica A: Statistical Mechanics and its Applications*, vol. 300, no. 3-4, pp. 424-432.
- Bagci, G.B., and U. Tirnakli (2009). “On the way towards a generalized entropy maximization procedure”, *Physics Letters A*, vol. 373, no. 36, pp. 3230-3234.
- Canosa, N. and R. Rossignoli (2003). “Generalized entropies and quantum entanglement”, *Physica A: Statistical Mechanics and its Applications*, vol. 329, no. 3-4, pp. 371-376.
- Carranza, M.L., Acosta A., and C. Ricotta (2007). “Analyzing landscape diversity in time: The use of Renyi's generalized entropy function”, *Ecological Indicators*, vol. 7, no. 3, pp. 505-510.
- Figueiredo, A., Amato, M.A., and T. Marciano da Rocha Filho (2006). “On the statistical interpretation of generalized entropies”, *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 191-206.
- Gudkov, V., and V. Montealegre (2008). “Analysis of networks using generalized mutual entropies”, *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 11, pp. 2620-2630.
- Hotta, M. and I. Joichi (1999). “Composability and generalized entropy” *Physics Letters A*, vol. 262, no. 4-5, pp. 302-309.
- Liping Yin and Lei Guo (2009). “Fault isolation for multivariate nonlinear non-Gaussian systems using generalized entropy optimization principle”, *Automatica*, vol. 45, no. 11, pp. 2612-2619.
- Masi, M. (2006). “On the q-parameter spectrum of generalized information-entropy measures with no cut-off prescriptions”, *Physics Letters A*, vol. 357, no. 4-5, pp. 288-294.
- Mathai, A.M. and H.J. Haubold (2007a). “Pathway model, superstatistics, Tsallis statistics, and a generalized measure of entropy”, *Physica A: Statistical Mechanics and its Applications*, vol. 375, no. 1, pp. 110-122.
- Mathai, A.M. and H.J. Haubold (2007b). “On generalized entropy measures and pathways”, *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 493-500.
- Matsuba, I. and H. Takahashi (2003). “Generalized entropy approach to stable Levy distributions with financial application”, *Physica A: Statistical Mechanics and its Applications*, vol. 319, pp. 458-468.
- Nanda, A.K. and P. Paul (2006a). “Some results on generalized residual entropy”, *Information Sciences*, vol. 176, no. 1, 6, pp. 27-47.

-
- Nanda, A.K. and P. Paul (2006b). "Some results on generalized past entropy", *Journal of Statistical Planning and Inference*, vol. 136, no. 10, pp. 3659-3674.
- Pashchenko, F.F. (2007). *Introduction to consistent methods of systems modeling. Volume 2. Identification of non-linear systems*. Moscow: Finansy i statistika, 2007, 288 p. ISBN 5-279-03042-2 (in Russian)
- Peeters L.M.K. (2004). "Estimating a random-coefficients sample-selection model using generalized maximum entropy", *Economics Letters*, vol. 84, no. 1, pp. 87-92.
- Qiang Liu, Ke-Fei Cao, and Shou-Li Peng (2009). "A generalized Kolmogorov-Sinai-like entropy under Markov shifts in symbolic dynamics", *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 20, pp. 4333-4344.
- Rojdestvenski, I. and M.G. Cottam (2000). "Time rescaling and generalized entropy in relation to the internal measurement concept", *Physica A: Statistical Mechanics and its Applications*, vol. 278, no. 1-2, pp. 150-160.
- Saridis, G.N. (2001). *Entropy in Control Engineering*, Singapore: World Scientific, 148 p.
- Sasaki, K. and M. Hotta (2002). "Generalized entropy composition with different q indices: An attempt", *Chaos, Solitons & Fractals*, vol. 13, no. 3, pp. 513-519.
- Scarfone, A.M. (2005). "Canonical quantization of classical systems with generalized entropies", *Reports on Mathematical Physics*, vol. 55, no. 2, pp. 169-177.
- Shamilov, A. (2007). "Generalized entropy optimization problems and the existence of their solutions", *Physica A: Statistical Mechanics and its Applications*, vol. 382, no. 2, pp. 465-472
- Taruya, A. and M. Sakagami (2002). "Gravothermal catastrophe and Tsallis' generalized entropy of self-gravitating systems", *Physica A: Statistical Mechanics and its Applications*, vol. 307, no. 1-2, pp. 185-206.
- van Beijeren, H. (2004). "Generalized dynamical entropies in weakly chaotic systems", *Physica D: Nonlinear Phenomena*, vol. 193, no. 1-4, pp. 90-95.
- Wada, T. (2004). "Thermodynamic stabilities of the generalized Boltzmann entropies", *Physica A: Statistical Mechanics and its Applications*, vol. 340, no. 1-3, pp. 126-130.

результатов по различным обобщениям понятия энтропии, относящихся к самым разнообразным областям, но то, что представлено в книге [4] превосходит всякое воображение!

3. Неопределенность

Согласно разделу 4.3.1 книги [4]:

«Обобщенной энтропией системы (процесса) S называется выражение

$$(3.1) \quad \mathfrak{E}^\phi(S) = k \log A(S),$$

где $\{A(S)\}$ – множество всех состояний системы, которое получается в результате действия произвольных или допустимых преобразований A на множество исходных состояний системы» [4] (стр. 225). Ясно, что значения выражения (3.1) не ограничены ни снизу, ни сверху при тех или иных A из $\{A(S)\}$.

Здесь уместно задаться вопросом, что имеется в виду при многочисленных упоминаниях понятия «энтропия» в книгах [1, 4], в том смысле, что не вполне ясно, как следует понимать то или иное выражение, стоящее под знаком логарифма. Так, например, в литературе по общесистемным закономерностям имеет место следующая трактовка энтропии.

«Энтропия служит количественной мерой беспорядка в системе и определяется числом допустимых состояний системы. Такое утверждение в точности соответствует определению $\mathfrak{E} \equiv \ln S$ и означает, что энтропия есть натуральный логарифм числа допустимых состояний системы S . Чем больше у системы число допустимых состояний S , тем больше энтропия. Так как логарифм – число безразмерное, то и энтропия является безразмерной величиной. Если система может находиться только в одном единственном допустимом состоянии $S = 1$, тогда $\mathfrak{E} = \ln 1 = 0$.»¹³

С другой стороны, в разделе 4.3.1 книги [4], значительная часть которого заслуживает того, чтобы быть приведенной в

¹³ Основные аспекты системного подхода к решению сложных задач // В кн. «Системные законы и закономерности в электродинамике, природе и обществе». М.: Наука, 2001. Гл. 2. С. 143-154. ISBN 5-02-013088-5

настоящем материале практически дословно, утверждается, что «с легкой руки М. Планка была введена энтропия как логарифмическая функция множества состояний или вероятность этих состояний. При этом энтропию $\mathcal{E}$ системы S , как одну из важнейших информационных характеристик, можно представить в виде

$$(3.2) \quad \mathcal{E} = k \log S,$$

где k – постоянная Больцмана, S – множество состояний системы S . Основание логарифма в (3.2) принимается равным числу символов алфавита, с помощью которого описывается информация.

Число возможных состояний S системы функционально связано с вероятностью $P(S)$ этих состояний. Поэтому энтропию $\mathcal{E}$ системы S можно определить по формуле

$$(3.3) \quad \mathcal{E} = -k \log P(S).$$

Определения (3.2) и (3.3) равноправны, но отличаются знаком, т.к. вероятности $P(\cdot) \leq 1$.

В общем случае величина k не обязательно совпадает с постоянной Больцмана. Она связана с выбором единиц информации, единиц измерения и называется *адиабатическим инвариантом*. В качестве адиабатического инварианта в зависимости от целей исследования и постановки задачи используют и другие физические постоянные, например постоянную Планка. Адиабатический инвариант в абстрактных системах обычно принимается равным единице, при этом $k = \ln a$, где a – основание логарифма в выражениях (3.2), (3.3). Изменение числа a не изменяет количество информации о системе. В зависимости от выбора основания логарифма величина k определяет известные единицы информации, например при $a = 2$, получаем в качестве единицы информации бит.

В классической термодинамике молчаливо принимается, что энтропия есть переменная, отнесенная к единице объема системы. В таком виде энтропия не изменяется при изменении числа S элементов системы. Тем не менее, когда речь идет о динамических, хаотических или слабоформализуемых системах, когда выполняется закономерность роста энтропии, важна эн-

тропия системы в целом, т.е. $\mathcal{E} = \mathcal{E}(\mathcal{S}, t \in T)$, где T – интервал жизни системы.

На основе определений (3.2) и (3.3) строятся энтропийная теория термодинамических процессов и негэнтропийный принцип. В соответствии с последним принципом информация делится на два вида: преходящая информация, ценность которой убывает с течением времени, и непреходящая информация, которая остается ценной в течение длительного времени. Примерами преходящей информации являются краткосрочный прогноз погоды, бюллетени биржи, газетные и телевизионные новости. К непреходящей или долгосрочной ценной информации относятся научные открытия, теория и модели, законы и закономерности.

При этом информация, содержащаяся в эмпирической закономерности или законе, определяется по формуле¹⁴:

$$(3.4) \quad I = k(\ln P_0 - \ln P_1) = \mathcal{E}_0 - \mathcal{E}_1,$$

где P_0 – площадь изменения исследуемой величины, P_1 – площадь, порождаемая погрешностью измерений, $\mathcal{E}_0$ – энтропия начального состояния системы, $\mathcal{E}_1$ – энтропия конечного состояния, определяемая ошибками измерения.

Формулу (6.4) можно выразить в терминах множества состояний системы $\mathcal{S}$:

$$(3.5) \quad I = k(\ln \mathcal{S}_0 - \ln \mathcal{S}_1) = \mathcal{E}_0 - \mathcal{E}_1 \gg$$

[4] (стр. 222-223).

И далее, автору [4] понятно, «что знания и опыт исследователя и создают ту «избыточную информацию», которая наряду с информацией, полученной из выражения (3.5), и формируют закон, теорию или модель исследуемого явления или системы. С другой стороны, очевидно, что полученный закон или модель должны быть адекватны моделируемой системе, а, следовательно, и тем экспериментальным данным, которые были получены при наблюдении» [4] (стр. 224).

¹⁴ Бриллюэн Л. Наука и теория информации. М.: Гос. из-во физ.-мат. лит., 1960. 392 с.

«Этот парадокс негэнтропийного принципа» по мнению автора [4] разрешим на пути использования следующих рассуждений: «Исследователь стремится построить мысленный, математический или физический фильтр, с помощью которого он хочет извлечь максимум информации об исследуемой системе, содержащейся в экспериментальных данных и/или теоретических знаниях о системе. Обозначим такой фильтр знаний (информации) через A :

$$(3.6) \quad \mathcal{E}'_0 = \max_{\{A\}} k \ln A(\mathcal{S}_0).$$

С другой стороны, исследователь-экспериментатор стремится снизить энтропию, вносимую помехами в канал измерений, т.е.

$$(3.7) \quad \mathcal{E}'_1 = \min_{\{B\}} k \ln B(\mathcal{S}_1),$$

где B – фильтр, снижающий уровень помех, или преобразователь, фильтрующий полезный сигнал от помех.

Полагая, что множества преобразований $\{A\}$ и $\{B\}$ включают в себя тождественные преобразования, получаем очевидные неравенства:

$$(3.8) \quad \mathcal{E}'_0 \geq \mathcal{E}_0, \quad \mathcal{E}'_1 \leq \mathcal{E}_1.$$

Информация, содержащаяся в эмпирической закономерности, в этом случае определяется по формуле

$$(3.9) \quad I' = \mathcal{E}'_0 - \mathcal{E}'_1 = k \left(\max_{\{A\}} \ln A(\mathcal{S}_0) - \min_{\{B\}} \ln B(\mathcal{S}_1) \right).$$

Сравнивая выражения (3.5) и (3.9), получаем неравенство

$$I' \geq I.$$

Разность количеств информации, получаемых по выражениям (3.5) и (3.9), и представляет собой количество информации, вносимое в теорию исследователем

$$(3.10) \quad \Delta I = I' - I.$$

Используя выражения энтропии системы через плотности вероятностей состояний (3.3), получим следующую оценку информации, содержащейся в эмпирической зависимости

(3.11)

$$I' = \mathfrak{Z}'_0 - \mathfrak{Z}'_1 = -k \left(\max_{\{A\}} \ln p(A(\mathbf{S}_0)) - \min_{\{B\}} \ln p(B(\mathbf{S}_1)) \right).$$

[4] (стр. 224-225).

Приведенные цитаты из книги [4] (стр. 224-225), вызывают множество вопросов, связанных как собственно с определением энтропии, так и с истолкованием и последующим «обобщением» формулы Бриллюэна (3.4).

Во-первых, всюду в определении энтропии, восходящем к термодинамике (формула Больцмана), под знаком логарифма стоит *число* состояний системы, в то время как здесь используются термины «состояние» и «множество состояний». Следует ли эти термины понимать эквивалентные термину «число состояний»? Очевидно – да, в противном случае совершенно неясно, что следует понимать под выражением «логарифм состояния (множества состояний)».

Во-вторых, в общепринятой вероятностной формулировке энтропии никогда не предполагается простой замены числа состояний их вероятностью (плотностью вероятности), как это сделано в формулах (3.3), (3.11). Ясно, например, что в этом случае стремление к нулю вероятности (плотности вероятности) ведет к неограниченному увеличению энтропии, что противоречит содержательному смыслу. Поэтому некорректно и представленное в книге [4, стр. 226] определение обобщенной энтропии в виде

$$\mathfrak{Z}^\Phi(\mathbf{S}) = -k \log P(A(\mathbf{S})),$$

равно как некорректны и приводимые там же неравенства

$$(3.12) \quad \mathfrak{Z}^\Phi(\mathbf{S}) \geq \mathfrak{Z}'(\mathbf{S}) \geq \mathfrak{Z}(\mathbf{S}),$$

где в соответствии с вышеприведенными определениями

$$\mathfrak{Z}'(\mathbf{S}) = \max_A \{-k \log P(A(\mathbf{S}))\},$$

а первое из неравенств (3.12) выглядит просто абсурдно.

В-третьих, также требует комментария определение компонент, входящих в формулу (3.4), поскольку в формуле (3.4) Л. Бриллюэном в качестве P_0 и P_1 рассматриваются соответствен-

но начальное и конечное числа микросостояний некоторой физической системы, в которой система молекул имеет P_0 различных возможных случаев с одинаковыми априорными вероятностями, и для уменьшения числа возможных случаев до P_1 требуется информация

$$(3.13) \quad I_b = k(\ln P_0/P_1).$$

«Для установления соотношения между информацией и энтропией имеем следующую картину:

Информация	Число микросостояний	Энтропия
Начальное положение $I_b = 0$	P_0	$S_0 = k \ln P_0$
Конечное положение $I_b \neq 0$	$P_1 < P_0$	$S_1 = k \ln P_1$

Связь между уменьшением энтропии системы и требуемой информацией очевидна, так как из (3.13) следует

$$I_b = k(\ln P_0 - \ln P_1) = S_0 - S_1$$

или

$$S_1 = S_0 - I_b.^{15}$$

Что касается интерпретации в книге [4] величин P_0 и как P_1 «площадей» (наблюдения, измерения), то она, по-видимому, почерпнута из другой работы Л. Бриллюэна¹⁶, ссылка на которую в этом случае выглядела бы более уместной. В свою очередь, в рамках данной интерпретации формулы (3.4) уместно задаться вопросом, о какой собственно системе идет речь в книге [4], если при этом используются термины «начальное» (характеризуемое величиной P_0 – площадью изменения исследуемой величины) и «конечное» (характеризуемое величиной P_1 – площадью,

¹⁵ Бриллюэн Л. Наука и теория информации. М.: Гос. из-во физ.-мат. лит., 1960. С. 201.

¹⁶ Бриллюэн Л. Научная неопределенность и информация. М.: Мир, 1966. 272 с.

порождаемой погрешностью измерений) состояния системы? Если исходить из рассмотренной выше концепции разрешения «парадокса негэнтропийного принципа», предложенной в книгах [4], согласно которой «исследователь стремится построить мысленный, математический или физический фильтр, с помощью которого он хочет извлечь максимум информации об исследуемой системе, содержащейся в экспериментальных данных и/или теоретических знаниях о системе», а «с другой стороны, исследователь-экспериментатор стремится снизить энтропию, вносимую помехами в канал измерений», то привлечение понятий «начальное» и «конечное» состояния системы выглядят здесь как не имеющие никакого отношения к предмету обсуждения. Но даже если абстрагироваться от этого несоответствия, то не может не вызывать удивления формула (3.6). Она по мысли автора [4] призвана отражать стремление исследователя «извлечь максимум информации об исследуемой системе, содержащейся в экспериментальных данных и/или теоретических знаниях о системе», но максимизация энтропии как раз наоборот отражает увеличение степени неопределенности, то есть минимизацию количества информации, а не извлечение ее максимума(!). Поэтому формула (3.6) бессмысленна как в формально-теоретическом отношении (максимум логарифмической функции на всей области ее определения не существует), так и в содержательном (в контексте данного рассмотрения) отношении.

4. Оптимальность

«Пусть пара $\langle x(t), y(t), t \in T \rangle$ представляет собой совокупность случайных процессов на входе и выходе идентифицируемой системы. Пусть $A(x(t), y(t), t)$ – модель системы. В каждый фиксированный момент t модель $A(\cdot)$ можно рассматривать как функционал, отображающий наблюдаемую реализацию входного сигнала $x(s)$, $x(s)$, $s \in T_t$, $T_t = T \cap (-\infty, t)$, где T_t –

интервал наблюдения, в выходной сигнал модели
 $y_M(t) = A(x(s), s \in T_t)$.

Рассмотрим различные критерии близости выходных сигналов системы $y(t)$ и модели $y_M(t)$ вида
(4.1)

$$J(\varphi) = M(\varphi(y(t), Ax(s))) = M[\varphi(y_M(t) - y(t))] \rightarrow \min_{Y_M},$$

или

$$(4.2) \quad J(\varphi) = M[\varphi(|y_M(t) - y(t)|)] \rightarrow \min_{Y_M},$$

где φ – произвольная неотрицательная неубывающая функция, $\varphi(\cdot) \neq \text{const}$.

Возникает вопрос: «Что такое Y_M в формулах (4.1), (4.2)?». Если следовать содержанию, то, по-видимому, $Y_M = y_M(t)$, а использование обозначения Y_M – просто очередная небрежность.

Далее в книге [4] рассматриваются функционал вида
(4.3)

$$J[\varphi(\cdot)/x(s)] = M[\varphi(|y_M(t) - y(t)|)/x(s), s \in T_t] \rightarrow \min_{Y_M}$$

и « Φ – класс моделей (например, линейных, нелинейных, полупереносных) в котором ищется идентифицируемая модель системы. Требуется найти модель системы из класса Φ , оптимальную в смысле критериев (4.1), (4.2) или (4.3)» (стр. 31).

В этих рамках стр. 32 книги [4] приводятся следующие определения.

Определение 1. Модель A^0 называется оптимальной «в среднем» в классе Φ , если $A^0 \in \Phi$ и для любых $A \in \Phi$, $x(s)$ любого и критерия (4.2) или (4.1) выполняется неравенство

$$(4.4) \quad J(\varphi, A^0) \leq J(\varphi, A).$$

Определение 2. Модель A^0 называется оптимальной «в целом» в классе Φ , если $A^0 \in \Phi$ и для любых $A \in \Phi$, $x(s)$ и критерия (4.3) выполняется неравенство

$$(4.5) \quad J(\varphi, A^0/x(s)) \leq J(\varphi, A/x(s)).$$

Теорема 1. 1. Пусть случайные процессы на входе $x(s)$ и выходе $y(t)$ представляют собой гауссовские процессы. «2. Пусть условные распределения $y(t)$ относительно $x(s)$

$$(4.6) \quad F_{y/x}(u) = P(y(t) \leq u/x(s) \in T_t)$$

унимодальны и симметричны. Тогда существует оптимальная модель «в целом», при этом выходная реакция модели определяется условным математическим ожиданием

$$(4.7) \quad y_M(t) = M(y(t)/x(s), s \in T_t).$$

Справедливо и обратное утверждение.

Теорема 2. Пусть выполняется условие 1 теоремы 1. Если существует оптимальная «в целом» модель, то соответствующие условные распределения (4.6) унимодальны и симметричны.

Доказательство теорем 1, 2 следует из работ [7, 8]» [4, стр. 32].

Принимая во внимание объем «работ» [7, 8], последнее предложение в приведенной цитате следует признать самым ярким в книгах [3, 4]!

«Теорема 3. Пусть выполнено условие 2 теоремы 1. Пусть выполнены условия: а) плотность условного распределения случайного процесса (величины) $y(t)$ относительно наблюдаемых значений $x(s)$ — $p_{y/x}(u)$ при $u > u_0 = M(y(t)/x(s), s \in T_t)$ строго убывающая функция; функция $\varphi_0(E)$ является строго возрастающей.

Тогда существует и единственная оптимальная «в целом» модель идентифицируемой системы:

$$J(\varphi_0, A^0/x(s)) = \min_A J(\varphi_0, A/x(s)).$$

Доказательство. Действительно, пусть выполняются условие 2 теоремы 1 и условие а). Пусть существуют оптимальные модели

$$A^0(x(\tau)) = M(y(t)/x(\tau), \tau \in T_t) = a_0,$$

$$B^0(x(\tau)) = b_0 \neq a_0.$$

Пусть $p_{y/x}(u) = F'_{y/x}(u + a_0)$ – условная плотность распределения.

В силу условий теоремы

$$(4.8) \quad p(u) = p(-u), \quad p(u_2) > p(u_1) \quad \text{при } u_2 > u_1 > 0.$$

Тогда имеем:

$$(4.9) \quad \begin{aligned} J(\varphi_0, A^0/x(\tau)) &= \int_{-\infty}^{\infty} \varphi_0(y - a_0) p(y - a_0) dy = \\ &= \int_{-\infty}^{\infty} \varphi_0(u) p(u) du, \end{aligned}$$

$$(4.10) \quad \begin{aligned} J(\varphi_0, B^0/x(\tau)) &= \int_{-\infty}^{\infty} \varphi_0(y - b_0) p(y - a_0) dy = \\ &= \int_{-\infty}^{\infty} \varphi_0(u + 2c) p(u) du, \end{aligned}$$

где

$$(4.11) \quad c = (a_0 - b_0)/2.$$

Откуда

$$(4.12) \quad \begin{aligned} J(\varphi_0, A^0/x(\tau)) - J(\varphi_0, B^0/x(\tau)) &= \\ &= \int_0^{\infty} [p(u - c) - p(u + c)] [\varphi_0(u + c) - \varphi_0(u - c)] du = \\ &= 0. \end{aligned}$$

¹⁷ Здесь, по-видимому, просто опечатка – знак неравенства должен быть обратным.

Поскольку последний интеграл равен нулю, то подынтегральное выражение

$$(4.13) \quad [p(u-c) - p(u+c)] [\varphi_0(u+c) - \varphi_0(u-c)] = 0$$

почти при всех u . В силу условия (4.5) первый сомножитель строго больше нуля, следовательно

$$\varphi_0(u+c) = \varphi_0(u-c) \text{ почти всюду.}$$

Последнее невозможно, так как $\varphi_0(u)$ в силу условия б) является строго возрастающей функцией. Следовательно, оптимальная модель A^0 единственна» [4, стр. 32, 33].

Однако следует заметить, что в (4.9), (4.10), (4.12), (4.13) непонятным образом исчез знак модуля при аргументе u , по сравнению с описанием критерия (4.3): всюду должно быть $\varphi_0(|u|)$. Кроме того, совершенно не обязательно, что в (4.11) значения c являются только положительными. Следовательно, доказательство теоремы 3 ошибочно.

Замечание. Ранее теорема 3 (и ее доказательство) были представлены в работе [10], однако при этом в формуле (4.12) нижний предел интегрирования равнялся $-\infty$:

$$(4.14) \quad J(\varphi_0, A^0/x(\tau)) - J(\varphi_0, B^0/x(\tau)) = \\ = \int_{-\infty}^{\infty} [p(u-c) - p(u+c)] [\varphi_0(u+c) - \varphi_0(u-c)] du = \\ = 0 \text{ [10].}$$

Ни в книге [4], ни в работе [10] формула (4.12) (соответственно, (4.14)) не выводится. Исходя из условий теоремы 3 и формулы (4.14) в свое время обращалось внимание редакции журнала «Проблемы управления» на ошибку в доказательстве теоремы 3, поскольку функция

$$f(u) = [p(u-c) - p(u+c)] [\varphi_0(u+c) - \varphi_0(u-c)]$$

принципе **не может быть не быть** знакопеременной **на всей** числовой оси [11]. Но это замечание осталось (и остается) без внимания.

В книге [4] за теоремой 3 следует теорема о единственности оптимальной «в среднем» модели, но ее доказательство опира-

ется на недоказанную, как показано выше, теорему 3 и, следовательно, также не может рассматриваться как достоверное.

5. Заключение

В настоящей работе подробно рассмотрены и проанализированы некоторые положения, методы и теоретические результаты, относящиеся к проблеме идентификации стохастических систем, изложенных в книге [4]. Причиной такого специального анализа явилось то обстоятельство, что справедливость и научная ценность многих из представленных в книге [4] результатов не может не вызывать серьезных сомнений, хотя книга [4] подготовлена при поддержке Российского фонда фундаментальных исследований (инициативный проект 05-08-50336а). Рассмотренные вопросы связаны с применением теоретико-информационных методов и методов теории оптимальных систем. Показано, что данные результаты либо не имеют содержательного значения, либо вообще представляют собой заблуждения, тем более опасные и вредоносные, что являются изложенными в учебном пособии. Таким образом, принимая во внимание все сказанное, следует сделать вывод, который следует из «работы» [12].

Литература

1. Пащенко Ф.Ф. Определение и моделирование закономерностей по экспериментальным данным // В кн. «Системные законы и закономерности в электродинамике, природе и обществе» М.: Наука, 2001. Гл. 7. С. 411-521. ISBN 5-02-013088-5
2. Чернышев К.Р. Эссе о некоторых заблуждениях в идентификации систем // Труды II Международной конференции «Идентификация систем и задачи управления» SICPRO '03. Москва, 29-31 января 2003 г. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2003. С. 2660-2698.
3. Пащенко Ф.Ф. Введение в состоятельные методы моделирования систем / Учеб. пособие: В 2-х ч. Ч. 1. Математические

- основы моделирования систем. М.: Финансы и статистика, 2006. 328 с. ISBN 5-279-02922-X
4. Пашенко Ф.Ф. Введение в состоятельные методы моделирования систем / Учеб. пособие: В 2-х ч. Ч. 2. Идентификация нелинейных систем. М.: Финансы и статистика, 2007. 288 с. ISBN 5-279-03042-2
 5. Чернышев К.Р. Рецензия на книгу Ф.Ф. Пашенко «Введение в состоятельные методы моделирования систем. Часть 1. Математические основы моделирования систем». Москва: Финансы и статистика, 2006. 328 с. // Интернет-конференция «Проблемы управления».
http://ubs.mtas.ru/forum/index.php?PAGE_NAME=read&FID=3&TID=468.
 6. Дургарян И.С., Пашенко Ф.Ф. Информационные методы идентификации / Препринт. М.: Институт проблем управления им. В.А. Трапезникова РАН, 1999. 52 с.
 7. Durgaryan I.S., Pashchenko F.F. An information theory approach to identification // Preprints of the 6th IFAC Symposium on Identification and System Parameter Estimation. Washington, D.C., 1982. Vol. 1.
 8. Андреев Н.И. Корреляционная теория статистически оптимальных систем. М.: Наука, 1966.
 9. Пугачев В.С. Теория случайных функций. М.: Физматгиз, 1962.
 10. Дургарян И.С., Пашенко Ф.Ф. Сложные статистические критерии и модели, оптимальные на классе критериев // Проблемы управления. 2003. № 2. С. 27-34.
 11. Чернышев К.Р. Об одной статье по идентификации стохастических систем. Неопубликованное письмо в редакцию журнала «Проблемы управления». 2005, 2010.
<http://cs.sicpro.org/CheLet.pdf>.
 12. Андерсен Ханс Кристиан. Новое платье короля. М.: Азбука-классика, 2010. 32 с.