УДК 519.86 (075.8)
ББК 22.18я73

РЕЦЕНЗИЯ НА КНИГУ Ф.Ф. ПАЩЕНКО «ВВЕДЕНИЕ В СОСТОЯТЕЛЬНЫЕ МЕТОДЫ МОДЕЛИРОВАНИЯ СИСТЕМ. ЧАСТЬ 1. МАТЕМАТИЧЕСКИЕ ОСНОВЫ МОДЕЛИРОВАНИЯ СИСТЕМ». МОСКВА: ФИНАНСЫ И СТАТИСТИКА, 2006. 328 С.

К.Р. Чернышев
(*Учреждение Российской академии наук
Институт проблем управления 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, 2006 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. Главное внимание уделено тем или иным аспектам применения такой состоятельной, по терминологии А.Н. Колмогорова, мер зависимости случайных величин (процессов), как максимальная корреляция. 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, меры зависимости, максимальная корреляция

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 переменных.

Именно на преодоление этого недостатка направлено использование в задачах идентификации и моделирования более сложных, нелинейных, мер зависимости, примером чему и является книга [1], которая, как заявлено в ее аннотации, представляет собой первое в России учебное пособие по состоятельным методам моделирования и идентификации систем. В нем рассматриваются состоятельные методы моделирования стохастических систем, основы математических методов моделирования и понятий теории систем. Впервые дается систематическое изложение состоятельных мер зависимости (по А.Н. Колмогорову) между случайными величинами и функциями и их применение для моделирования нелинейных систем. Исследованы уравнения идентификации систем, обобщающие известные методы типа Винера-Хопфа, регрессионные подходы и применения состоятельных методов к решению разнообразных задач моделирования.

При этом книга [1] допущена «УМО в области прикладной математики и управления качеством в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлению 230400 «Прикладная математика» специальности 230401 «Прикладная математика», а ее материал, хотя и не содержит никаких новых научных результатов, подготовлен при поддержке Российского фонда фундаментальных исследований (проект 05-08-50336а) и Программы № 15 фундаментальных исследований Отделения энергетики, машиностроения, механики и процессов управления Российской академии наук.

Тем не менее, в настоящей работе показывается, что книга [1] содержит целый ряд положений, которые либо лишены содержательного смысла, либо вообще представляют собой заблуждения, что, безусловно, недопустимо ни в научной, ни, тем более, учебной литературе.

2. Некоторые примеры и контрпримеры

2.1. СТОХАСТИЧЕСКАЯ СВЯЗЬ НЕ ЭКВИВАЛЕНТНА ЗАВИСИМОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

Пример 1. В рамках «проблемы моделирования и анализа статистических связей (А.Н. Колмогоров, О.В. Сарманов, А. Реньи, Г. Гебелайн, Г. Бокс, Н.С. Райбман и др.), о которых обычно умалчивается в исследовательских работах и учебных пособиях, но которые имеют большое значение при моделировании различных систем» [1], стр. 70, в разделе 1.8 книги [2] приводится следующий пример.

Пусть ξ и ν – пара независимых одинаково распределенных случайных величин. Никаких других условий на эти случайные величины в разделе 1.8 книги [1, стр. 70] не накладывается, что очевидно подразумевает и то, что их распределение может считаться *произвольным*. Далее вводится пара случайных величин X и Y, определяемых соотношениями

(2.1) $$X = \xi + \nu,$$

(2.2) $$Y = \xi - \nu.$$

Тогда случайные величины X и Y не коррелированны, хотя очевидно, что между ними существует стохастическая связь, которая имеет вид:

(2.3) $$Y = X - 2\nu.$$

Но некорректность данного примера из книги [1, стр. 70], вводящая в заблуждение, в первую очередь, студентов, как раз и заключается в том, что *стохастическая связь между случайными величинами не обязательно обусловливает зависимость между ними*.

Действительно, соотношения (2.1), (2.2) записываются в матричной форме следующим образом

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \xi \\ \nu \end{bmatrix}. \tag{2.4}$$

При этом, хорошо известно выражение для плотности распределения результата Z_f детерминированного взаимно однозначного преобразования f случайного вектора Z_0 через его плотность распределения, $Z_f = f(Z_0)$:

$$p_{z_f}(Z_f) = p_{Z_0}\left(f^{-1}(Z_f)\right)\left|\frac{D\left(f^{-1}(Z_f)\right)}{D(Z_f)}\right|, \tag{2.5}$$

где $\dfrac{D\left(f^{-1}(Z_f)\right)}{D(Z_f)}$ – якобиан обратного преобразования $Z_0 = f^{-1}(Z_f)$. $p_{Z_f}(Z_f)$, $p_{Z_0}(Z_0)$ – плотности распределения случайных векторов Z_f и Z_0 соответственно. В свою очередь, обратное к (2.4) преобразование имеет вид

$$\begin{bmatrix} \xi \\ \nu \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \end{bmatrix}. \tag{2.6}$$

Тогда из формул (2.4), (2.5), (2.6) вытекает следующее выражение для совместной плотности распределения $p_{XY}(X,Y)$ случайных величин *X* и *Y*, определяемых выражениями (2.1) и (2.2),

$$p_{XY}(X,Y) = p_\xi\left(\frac{X+Y}{2}\right) \cdot p_\nu\left(\frac{X-Y}{2}\right) \cdot \frac{1}{2}, \tag{2.7}$$

где $p_\xi(\xi)$, $p_v(v)$ – маргинальные плотности распределения случайных величин *u* и *v*, и из соотношения (2.7) немедленно следует, что в случае, когда плотность распределения случайных величин ξ и *v* является нормальной со средним *m* и среднеквадратическим отклонением *σ*,

$$p_\xi(\xi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\xi-m)^2}{2\sigma^2}}, \tag{2.8}$$

$$p_v(v) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(v-m)^2}{2\sigma^2}},$$

то из формулы (2.7) следует

$$p_{XY}(X,Y) = \frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}\sigma}\cdot\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\frac{1}{2}X^2-2Xm+\frac{1}{2}Y^2+2m^2}{2\sigma^2}} =$$

$$= \left\{\frac{1}{\sqrt{2\pi}\left(\sqrt{2}\sigma\right)} e^{-\frac{(X-2m)^2}{2\cdot\left(\sqrt{2}\sigma\right)^2}}\right\}\cdot\left\{\frac{1}{\sqrt{2\pi}\left(\sqrt{2}\sigma\right)} e^{-\frac{Y^2}{2\cdot\left(\sqrt{2}\sigma\right)^2}}\right\} =$$

$$= p_X(X)p_Y(Y),$$

где $p_X(X)$, $p_Y(Y)$ – маргинальные плотности распределения случайных величин *X* и *Y*. Таким образом, в условиях (2.1), (2.2), (2.8) *случайные величины X и Y являются независимыми, хотя между ними существует стохастическая связь*, определяемая соотношением (2.3). При этом следует подчеркнуть, что связь между *X* и *Y* является именно стохастической, а не детерминированной. На эту интересную особенность, отмеченную в работе [2], когда при стохастической связи между случайными величинами может сохраняться их независимость, следовало бы специально обратить внимание в книге [1], однако вместо этого она была преднамеренно (в силу знакомства автора книги [1] с работой [2]) скрыта.

2.2. «СОСТОЯТЕЛЬНЫЙ» МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Пример 2. В разделе 1.8 книги [2] в еще одном примеру обращения в ноль коэффициентов корреляции при зависимости случайных величин рассматривается «стохастический объект вида

(2.9) $$y = a_1 x_1^2 + a_2 x_2^2 ,$$

где x_1, x_2 – случайные входные сигналы с нормальными законами распределения $N_1(0,1), N_2(0,1)$.

Модель ищется в виде $y = a_1 x_1 + a_2 x_2$.

Оценки метода наименьших квадратов для этой модели имеют вид

(2.10) $$A = \left(X^T X\right)^{-1} X^T Y ,$$

где $\left(X^T X\right) = \begin{bmatrix} K_{x_1 x_1} & K_{x_1 x_2} \\ K_{x_2 x_1} & K_{x_2 x_2} \end{bmatrix}$; $X^T Y = \begin{pmatrix} K_{y x_1} \\ K_{y x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Следовательно вектор оценок неизвестных параметров равен $(a_1, a_2)^T = (0,0)$, $y = 0$ для любых значений x_1, x_2. Отсюда следует, что входные факторы не определяют выходного сигнала системы, что не соответствует действительности. Аналогичные рассуждения справедливы, если объект имеет вид

$$y = a_1 x_1^2 + a_2 x_2^2 + \xi ,$$

где ξ – случайный шум с параметрами $\sim N\xi(0,1)$ – статистически независимый с x_1 и x_2.

Таким образом, метод наименьших квадратов, широко использующийся при решении различных задач обработки экспериментальных данных в естествознании, экономике, социологии и других науках, может дать в корне неверные результаты, характеризующие исследуемое явление или систему в случае неправильного выбора структуры модели или некорректной постановки задачи идентификации и выбора решения» (стр. 72, 73).

Здесь необходимо обратить внимание на то замечание, что «входные факторы не определяют выходного сигнала системы, что не соответствует действительности» не верно, поскольку речь идет о конструируемой исследователем модели и соответственно, о выходном сигнале модели, а не системы (объекта). Более того, если исходить из постановки задачи идентификации по среднеквадратическому критерию, решение которой и приводит к соотношению (2.10), то именно нулевые значения коэффициентов в указанной линейной модели и дают наименьше критерия идентификации. Другими словами, решение задачи идентификации достигнуто: построенная *модель*

$$y_M^0 \equiv 0$$

является линейной, для *любой другой линейной модели* $\widetilde{y}_M$ имеет место неравенство

$$\mathbf{M}\left(a_{\cdot 1}x_1^2 + a_{\cdot 2}x_2^2 - \widetilde{y}_M\right)^2 > \mathbf{M}\left(a_{\cdot 1}x_1^2 + a_{\cdot 2}x_2^2 - y_M^0\right)^2.$$

Таким образом, метод наименьших квадратов не дает «в корне неверные результаты», а как раз наоборот является эффективным индикативным инструментом, указывающим исследователю на необходимость изменения исходной постановки задачи идентификации, если получаемое решение его не устраивает по другим, не формализованным в первоначальной постановке задачи, характеристикам.

В книге же [1] подход к «решению» «проблемы», поставленной в рассмотренном примере из раздела 1.8, изложен в разделе 4.3.1 под названием «Состоятельный метод наименьших квадратов» – вольность обращения с общепринятой терминологией не может не вызывать удивления: «Использование аппарата обобщенных корреляционных функций и обобщенных корреляций приводит к созданию состоятельного метода наименьших квадратов.

Рассмотрим статическую систему, описываемую нелинейным уравнением

$$Y = F(\bar{x}), \tag{2.11}$$

где $\bar{x}$ – n-мерный вектор. Модель системы ищем в классе полулинейных систем

(2.12) $$By = ACx,$$

где B и C – некоторые нелинейные преобразователи, A – вектор неизвестных параметров» (стр. 272).

Помимо неконкретности принятых обозначений, следует также отметить и крайнюю неудачность термина «полулинейные системы»[1].

В (2.12) преобразования B и C определяются «как решение следующей задачи

(2.13) $$(B, C) = \arg \sup_{\{B\},\{C\}} \frac{\mathbf{cov}(By, Cx)}{\sqrt{\mathbf{D}(By)\mathbf{D}(Cx)}},$$

$$\mathbf{D}(By) = \mathbf{D}(Cx) = 1, \; \mathbf{M}(By) = \mathbf{M}(Cx) = 0.$$

Тогда вектор неизвестных параметров A, соответствующий минимуму среднеквадратической ошибки, при вычисленных преобразователях B и C удовлетворяет уравнению

(2.14) $$A = \left([Cx]^T [Cx]\right)^{-1} [Cx]^T By,$$

где матрица $\left([Cx]^T [Cx]\right)$ имеет вид

(2.15) $$\left([Cx]^T [Cx]\right) = \begin{pmatrix} 1 & r^{\Phi}_{x_1 x_2} & \cdots & r^{\Phi}_{x_1 x_n} \\ r^{\Phi}_{x_2 x_1} & 1 & \cdots & r^{\Phi}_{x_2 x_n} \\ \cdots & \cdots & \cdots & \cdots \\ r^{\Phi}_{x_n x_1} & r^{\Phi}_{x_n x_2} & \cdots & 1 \end{pmatrix},$$

а матрица $[Cx]^T By$ представляет собой вектор-столбец нормированных максимальных коэффициентов корреляции

(2.16) $$[Cx]^T By = \left(r^{\Phi \max}_{yx_1}, \quad r^{\Phi \max}_{yx_2}, \quad \ldots, \quad r^{\Phi \max}_{yx_n} \right)^T.$$

В общем случае, когда преобразования B и C ищутся из условия

(2.17) $$(B, C) = \arg \frac{\mathbf{cov}(By, Cx)}{\sqrt{\mathbf{D}(By)\mathbf{D}(Cx)}},$$

[1] Ядыкин И.Б. Выступление на дискуссии 1 октября 2001 г.

матрица $[Cx]^T By$ принимает вид:

$$[Cx]^T By = \begin{pmatrix} r_{yx_1}^{\Phi}, & r_{yx_2}^{\Phi}, & \ldots, & r_{yx_n}^{\Phi} \end{pmatrix}^T, \tag{2.18}$$

где $r_{yx_i}^{\Phi}$ – нормированный функциональный коэффициент корреляции между выходной – y и i-й входной – x_i переменными моделируемой системы.

Учитывая, что функциональные корреляционные функции и коэффициенты всегда отличны от нуля, когда величины x_i и у статистически зависимы, предложенный метод представляет собой состоятельный вариант метода наименьших квадратов. Проблема, изложенная в примере 2, решена» (стр. 272-273).

Оставляя за рамками рассмотрения «общий случай» (2.17), (2.18), следует в первую очередь уточнить принятые в формулах (2.13)-(2.16) обозначения. Если исходить из модели (2.12), то Cx – вектор-столбец. Если исходить из выражения (2.13), то Cx – скаляр. Таким образом, и, принимая во внимание обозначения (2.15), (2.16), можно заключить, что выражение (2.13) очевидно представляет собой некую упрощенную запись следующего критерия

$$(B, C_i) = \arg \sup_{\{B\},\{C_i\}} \frac{\mathbf{cov}(By, C_i x_i)}{\sqrt{\mathbf{D}(By)\mathbf{D}(C_i x_i)}},$$

$$\mathbf{D}(By) = \mathbf{D}(C_i x_i) = 1, \qquad \mathbf{M}(By) = \mathbf{M}(C_i x_i) = 1,$$

$i = 1, \ldots, n$.

Соответственно «нормированный функциональный коэффициент корреляции между выходной – y и i-й входной – x_i переменными моделируемой системы» определяется выражением

$$r_{yx_i}^{\Phi} = \frac{\mathbf{cov}(By, C_i x_i)}{\sqrt{\mathbf{D}(B_i y)\mathbf{D}(C_i x)}}.$$

А теперь необходимо отметить, что изложенный в разделе 4.3.1 книги [1] метод вообще не имеет отношения к проблеме, изложенной в примере 2, поскольку в этом примере рассматривается задача *параметрической* идентификации при *заданной* структуре модели (линейная модель). Таким образом, в примере 2 и в методе из раздела 4.3.1 постановки задачи идентификации принципиально различны, поскольку принципиально различны как модели, подлежащие идентификации (ср. выражения вида $y = Ax$ для примера 2 и выражение $By = ACx$ (2.12), так и критерии идентификации (в примере 2 оптимальному выбору подлежит только вектор неизвестных параметров, в «состоятельном методе наименьших квадратов – два нелинейных преобразования и вектор параметров»).

Далее, использование максимальной корреляционной функции исключает рассмотрение случаев, когда на «выход» исследуемой системы действует независимая от «входа» аддитивная помеха. В случае такой постановки задачи идентификации, получаемые в конечном итоге преобразования «входа» и «выхода» будут непосредственно зависеть от помехи. Следовательно, получаемое решение задачи идентификации нельзя будет признать сколь-нибудь приемлемым не только с практической, но и с теоретической, методологической, точек зрения. Так, даже в случае идентификации линейных систем в задаче параметрической идентификации основное внимание в условиях воздействия помех уделяется устранению их влияния на оценки параметров идентифицируемой системы: соответствующие методы опираются на использование информации о структуре собственно системы, о структуре помех, в конечном итоге – на применение метода инструментальных переменных. Тем более не допустимо пренебрегать влиянием помех в нелинейном случае, а такое влияние неизбежно сохраняется при нелинейном преобразовании «выхода» системы. В противном случае можно было бы вообще рассматривать вместо «входа» и «выхода» системы их смеси с ненаблюдаемыми помехами произвольной структуры и определять входо-выходное соотношение на основе таких наблюдений, которые в принципе уже не соответствовали бы ни истинному входному, ни истинному выходному сигналам. Таким образом, принципиальная особенность использования максимальной корреляции в задачах идентификации состоит в рассмотрении систем без помех и для объектов, подобных (2.14), не применима: *не выполняется* условие

$$\left(r_{yx_1}^{\Phi\max},\quad r_{yx_2}^{\Phi\max},\quad \ldots,\quad r_{yx_n}^{\Phi\max} \right) =$$

$$= \left(r_{\tilde{y}x_1}^{\Phi\max},\quad r_{\tilde{y}x_2}^{\Phi\max},\quad \ldots,\quad r_{\tilde{y}x_n}^{\Phi\max} \right),$$

поскольку в общем случае

$$\arg \sup_{\{B\},\{C_i\}} \frac{\mathbf{cov}\left(B(\tilde{y}+\xi), C_i x_i\right)}{\sqrt{\mathbf{D}\left(B(\tilde{y}+\xi)\right)\mathbf{D}\left(C_i x_i\right)}} \neq \arg \sup_{\{B\},\{C_i\}} \frac{\mathbf{cov}\left(B\tilde{y}, C_i x_i\right)}{\sqrt{\mathbf{D}\left(B\tilde{y}\right)\mathbf{D}\left(C_i x_i\right)}},$$

$$\mathbf{D}\left(B(\tilde{y}+\xi)\right) = \mathbf{D}\left(C_i x_i\right) = 1 \qquad \mathbf{D}\left(B\tilde{y}\right) = \mathbf{D}\left(C_i x_i\right) = 1$$

$$i = 1, \ldots, n$$

и

$$\sup_{\{B\},\{C_i\}} \frac{\mathbf{cov}\left(B(\tilde{y}+\xi), C_i x_i\right)}{\sqrt{\mathbf{D}\left(B(\tilde{y}+\xi)\right)\mathbf{D}\left(C_i x_i\right)}} \neq \sup_{\{B\},\{C_i\}} \frac{\mathbf{cov}\left(B\tilde{y}, C_i x_i\right)}{\sqrt{\mathbf{D}\left(B\tilde{y}\right)\mathbf{D}\left(C_i x_i\right)}},$$

$$\mathbf{D}\left(B(\tilde{y}+\xi)\right) = \mathbf{D}\left(C_i x_i\right) = 1 \qquad \mathbf{D}\left(B\tilde{y}\right) = \mathbf{D}\left(C_i x_i\right) = 1$$

$$i = 1, \ldots, n.$$

Здесь $\tilde{y}$ – выходной сигнал объекта (2.14) при отсутствии помехи, то есть при $\xi \equiv 0$ п.н. [2].

Благодаря работе [2], в книге [1] «состоятельный метод наименьших квадратов» рассмотрен для «бесшумового» описания объекта. Но, тем не менее, в разделе 4.3.1 книги [1] в качестве замечания делается ничем не подтвержденное утверждение о том, что рассмотренный метод «применим и для систем, в каналах измерений которых действуют шумы

$$Y = F(\bar{x}) + \xi,$$

где ξ – приведенная помеха (случайный шум), статистически независимая от вектора $\bar{x}$ и распределенная по нормальному закону с параметрами $N(0, \sigma_\xi)$. Однако в этом случае необходимо учитывать влияние шумов на оценки обобщенных корреляционных функций и нелинейных операторов *B* и *C*» (стр. 273). При этом заявление автора книги [1] о необходимости «учитывать влияние шумов …» – образец тривиального ответа при отсутствии конструктивных предложений, тем более недопустимого для учебного пособия. Это тривиальное утверждение вызвано, естественно, тем, что сам автор книги [1] не знает как «учитывать влияние шумов …». Поэтому эта проблема предоставлена студентам для самостоятельного решения – раздел «Задачи для самостоятельно решения» главы 4 книги [1, стр. 319] завершается следующей «постановкой»:

«22. Построить алгоритмы оценивания влияния помех на точность метода функциональных преобразований и оценки преобразований A, B, C» [1, стр. 319].

3. «Полулинейность» («обобщенная» статистическая линеаризация)

Статистическая линеаризация входо-выходного отображения исследуемых систем относится к таким задачам нелинейной идентификации, решение которых существенно определяется характеристиками зависимости входных и выходных процессов системы. При этом известные подходы основаны либо на применении обычных корреляционных функций, либо дисперсионных функций. При этом методы дисперсионной линеаризации, вообще говоря, выводят за рамки линейных моделей.

В разделе 4.4 книги [1] рассматривается дальнейшее «обобщение» метода статистической «линеаризации» (кавычки в данном случае использованы постольку, поскольку получаемые в итоге модели не являются линейными (относительно центрированного входного процесса), и, следовательно, задача собственно линеаризации не решается, а заменяется аппроксимацией исходного объекта некоторой нелинейной моделью из заданного класса). А именно, пусть имеется «нелинейный объект

$$Y(t) = F\big(X(s), t\big), \tag{3.1}$$

где $Y(t)$ – выходной случайный сигнал, $X(s)$ – входной случайный сигнал, $F(\cdot, \cdot)$ – нелинейный безынерционный или динамический преобразователь, который может быть оператором Урысона, нелинейным дифференциальным уравнением, нелинейной функцией, содержать δ-функции и их производные, логические операторы и т.д.». (стр. 293).

В качестве модели объекта рассматривается выражение

(3.2) $$By(t) = ACx(s),$$

В (3.2) B – некоторое нелинейное преобразование выходного процесса модели $y(t)$, C – некоторое нелинейное преобразование входного процесса модели $x(s)$, A – некоторое линейное отображение.

В качестве критерия статистической линеаризации в разделе 4.4.2 книги [1] рассматриваются: первый критерий – условие совпадения математических ожиданий «выхода» объекта и «выхода» модели, условие совпадения функциональных автокорреляционных функций «выхода» объекта и «выхода» модели; второй критерий – условие минимума среднеквадратической ошибки.

В разделе 4.4.2 книги [1] для «простоты изложения результатов» предполагается, «что в классе моделей (3.2) существует обратный оператор B^{-1}, а сама модель системы (3.1) ищется в классе моделей

(3.3) $$y(t) = B^{-1}ACx(s).$$

Рассмотрим теперь задачу идентификации нелинейной системы по первому и второму критериям статистической линеаризации.

Первый критерий в данном случае приобретает вид

(3.4) $$\mathbf{M}y(t) = \mathbf{M}y_M(t); \quad R^{\phi}_{yy}(t,s) = R^{\phi}_{y_M y_M}(t,s),$$

где

(3.5а) $$y_M(t) = K_0(t)\mathbf{M}\left(B^{-1}ACx(s)\right) + K_1(t)B^{-1}AC\overset{\circ}{x}(s)$$

или

(3.5б) $$y_M(t) = K_0(t)A\mathbf{M}Cx(s) + K_1(t)AC\overset{\circ}{x}(s).$$

Пусть A –линейный нестационарный интегральный оператор вида

(3.6) $$Ax(t,s) = \int_0^T g(t,s)x(s)ds,$$

где $g(t,s)=0$ при $s,t \notin [0,T]$; $g(t,s)\neq 0$ при $s,t \in [0,T]$» (стр. 298).

(В формуле (3.4) $R_{zz}^{\phi}(t,s)=\mathbf{cov}\big(B(z(t)), Bz(s)\big)$ – функциональная автокорреляционная функция соответствующего случайного процесса (то есть автокорреляционная функция некоторого функционального преобразования данного процесса)).

Тогда из условий (3.4) в книге [1] «получена» следующая система уравнений (стр. 298)

$$\mathbf{M}y(t)=K_0(t)\mathbf{M}B^{-1}\int_0^T g(t,\tau)C[x(\tau)]d\tau, \tag{3.7а}$$

$$R_{yy}^{\phi}(t,s)= \\ =K_1(t)K_1(s)\int_0^T\int_0^T g(t,\tau)g(s,\lambda)R_{xx}^{\phi}(\tau,\lambda)d\tau d\lambda. \tag{3.7б}$$

При использовании второго критерия статистической линеаризации (критерия минимума «средней квадратической ошибки») в книге [1] выписана следующая система уравнений

$$\mathbf{M}\left(\big(By(t)\big)-K_0(t)\int_T g(t,\tau)C[x(\tau)]d\tau\right)=0, \tag{3.8а}$$

$$R_{yx}^{\phi}(t,s)=K_1(t)\int_T g(t,\tau)R_{xx}^{\phi}(s,\tau)d\tau \tag{3.8б}$$ » (стр. 299).

При этом в разделе 4.4.2 книги [1] указывается, что решение данной задачи статистической «линеаризации» состоит из следующих четырех этапов.

На первом определяются операторы B и C в соответствии с критерием

$$(B,C)=\arg\sup_{\{B\},\{C\}}\frac{\mathbf{cov}\big(By(t),Cx(s)\big)}{\sqrt{\mathbf{D}\big(By(t)\big)\mathbf{D}\big(Cx(s)\big)}}, \\ \mathbf{D}\big(By(t)\big)=\mathbf{D}\big(Cx(s)\big)=1. \tag{3.9}$$

«На втором этапе определяется коэффициент усиления по случайной составляющей $K_1(t)$. Например, (3.10) $$K_1(t) = R^{\phi}_{yy}(t,t) \Big/ R^{\phi}_{xx}(t,t)\,.$$

На третьем этапе определяется оператор A или его весовая функция $g(t,s)$, и, наконец, на четвертом этапе из первых выражений систем уравнений (3.7) и (3.8) в зависимости от выбранного критерия статистической линеаризации определяется коэффициент усиления по систематической составляющей – $K_0(t)$» (стр. 300).

Как видно, первые два из данных четырех этапов носят сугубо декларативный характер и никак не связаны с критериями статистической «линеаризации» – еще один крайне неудачный термин, поскольку ни о каких линейных моделях здесь речь не идет (если исходить из принятого термина «полулинейные» модели, то данный поход следовало бы называть «статистическая *полу*линеаризация». В представленной схеме условие (3.9) никак не связано с указанными критериями «линеаризации» как мерами близости объекта и модели, причем при этом не оговаривается тот совершенно обязательный момент, что преобразование B, определяемое условием (3.9) должно быть обратимым. Условие (3.10) вообще заявлено как налагаемое «например», из чего следует, что коэффициент $K_1(t)$ может также выбираться любым другим способом, «например», полагаться равным единице или какой-либо другой константе, либо – какой-нибудь априори заданной функции, что указывает на бессмысленность введения коэффициента $K_1(t)$ в модели (3.5а), (3.5б).

Более того, сами уравнения (3.7), (3.8) ни в книге [1], ни в публикациях, цитируемых в книге [1], по данной теме не выводятся. В то же время правдоподобность данных формул не может не вызывать сомнений, принимая во внимание, по крайней мере, нелинейность оператора B. Более точно, можно лишь утверждать, что справедливо только уравнение (3.7а), получаемое как результат взятия математических ожиданий левой и правой частей выражения (3.5) (при условии

$$\mathbf{M}B^{-1}AC\overset{\circ}{x}(s) = 0\,).$$

Что касается остальных уравнений – (3.7б), (3.8а), (3.8б), – то легко видеть что они *линейны* как по коэффициентам $K_0(t)$, $K_1(t)$, так и по весовой функции $g(t,s)$, в то время как модель (3.5) *линейна* по коэффициентам $K_0(t)$, $K_1(t)$, но *нелинейна* по весовой функции $g(t,s)$ (следует также заметить, что *верное* уравнение (3.7а) *линейно* по коэффициенту $K_0(t)$, но *нелинейно* по весовой функции $g(t,s)$). Данное обстоятельство подтверждает предположение об ошибочности уравнений (3.7б), (3.8а), (3.8б).

Можно попытаться «вывести», например, уравнение (3.7б). Из критерия (3.4) с учетом рассмотренных выше этапов из книги [1] (стр. 298), следует:

$$\mathbf{M}\left(B\overset{\circ}{y}(t)B\overset{\circ}{y}(s)\right)\overset{\Delta}{=}R^{\phi}_{yy}(t,s)=$$

$$=R^{\phi}_{y_M y_M}(t,s)\overset{\Delta}{=}\mathbf{M}\left(B\overset{\circ}{y}_M(t)B\overset{\circ}{y}_M(s)\right)=$$

$$=\mathbf{M}\left(B\left\{K_1(t)B^{-1}\int_T g(t,\tau)C\left[x(\tau)\right]d\tau\right\}B\left\{K_1(s)B^{-1}\int_T g(s,\lambda)C\left[x(\lambda)\right]d\lambda\right\}\right).$$

Следовательно, уравнение (3.7б) могло быть справедливо при условии, что коэффициент $K_1(t)$ коммутирует с *нелинейным* преобразованием *B*, но для такого предположения нет никаких оснований.

Еще большая неопределенность имеется в отношении выяснения справедливости уравнений (3.8а), (3.8б), поскольку в книге [1] «при использовании второго критерия статистической линеаризации (критерия минимума средней квадратической ошибки)» его аналитическое выражение не представлено. Однако с учетом рассмотренной выше четырехэтапной схемы решения задачи «обобщенной» статистической линеаризации, согласно которой преобразования *B* и *C* выбираются из условия (3.9), можно заключить, что такой критерий имеет вид

(3.11)

$$\mathbf{M}\left(y(t) - K_0(t)\mathbf{M}\left(B^{-1}ACx(s)\right) + K_1(t)B^{-1}AC\overset{\circ}{x}(s \right)^2 \to \min_{K_0,K_1,A} .$$

Соответственно, утверждение о том, что из критерия (3.11) вытекают уравнения (3.8) не выглядит даже правдоподобным, не говоря уже о его полноценном математическом обосновании.

Таким образом, все сказанное свидетельствует о несостоятельности (в общеупотребительном смысле этого слова) рассмотренного метода статистической «линеаризации».

5. Тавтология и смысл

5.1. ТАВТОЛОГИЯ

В череде «обобщений», в книге [1] представлены следующие рассуждения (при этом рассматривается входо-выходное описание некоторой системы вида

(5.1) $BY=ACX$,

где Y – «выход», X – «вход» системы):

«Пусть A, B, C – некоторые в общем случае нелинейные операторы. Следует отметить, что такое представление системы или идентифицируемого объекта очень привлекательно, поскольку включает в себя как общее описание стохастических систем через наиболее полные характеристики:

- функции вероятностей входо-выходных соотношений,

(5.2) $$P(Y/X) = P(X,Y)P^{-1}(X),$$

где

$$B : Y \to P(Y/X);\ C : X \to P^{-1}(X);$$

$A : (X,Y) \to P(X,Y)$,

$P(X,Y)$ – вероятность событий (случайных величин X,Y); $P(Y/X)$ – условная плотность, $P(X)$ – одномерная вероятность события (случайной величины);

- информационные характеристики случайных величин…

$$(5.3) \qquad H(Y/X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log(p(y/x))dxdy\,.$$

$$B: Y \to H(Y/X),\ C: X \to \log(p(y/x)),$$

$$A: CX \to -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)CXdxdy\,,$$

где $H(Y/X)$ – условная энтропия, $p(x,y)$ – совместная плотность распределения, $p(y/x)$ – условная плотность распределения случайных величин X и Y,

- так и частные описания систем через корреляционные функции.» (стр. 256-257).

Однако заявленная автором книги [1] «привлекательность такого описания» в представленной интерпретации крайне сомнительна, а, в сущности, процитированная интерпретация является не более чем тавтологией, поскольку совместная плотность распределения $p(x,y)$ *непосредственно* влечет любые другие характеристики – условные и маргинальные плотности, энтропии (условную, взаимную, маргинальную), моменты (взаимные, безусловные, условные) и т.п. Соответственно, возможность представления выражений (5.2), (5.3) через (5.1) не более содержательна, чем, например, возможность при проведении численных расчетов представления 8 в виде: $8 = 2 \times 4$.

5.2. СМЫСЛ

В настоящем разделе будут приведены два примера из книги [1], характеризуемых полным отсутствием содержательного смысла.

Во-первых, в разделе 2.6 книги [1, стр. 148-156] введен целый набор уникальных по своей содержательной пустоте определений (83-106), сводимых к следующему: Пусть для некоторого случайного процесса (функции) имеет место свойство “A”. Случайный процесс (функция) $X(t)$ обладает свойством ϕ-“A”, если существует такое преобразование C, что случайный процесс (функция) $CX(t)$ обладает свойством “A”. Поскольку в дальнейшем эти определения не используются, а даже в качестве отвлеченного примера существование таких преобразований *C* не продемонстрировано (и, тем более, не, доказано), то следует заключить, что *смысл* введения данных определений заключался лишь в увеличении объема книги [1].

Во-вторых, на стр. 77-80 книги [1] представлен материал, который целесообразно (рамках темы текущего раздела данной работы) просто процитировать без каких-либо комментариев, поскольку нижеприводимый фрагмент из книги [1] говорит сам за себя:

«Предположим, что зависимость между двумя величинами выражается функцией следующего вида: $y = f(x; a_1, a_2, \ldots, a_m)$, (например, $y = a_1 x + a_2$), где неизвестны только параметры $a_1, a_2, \ldots, a_m$, а тип функции известен (например, линейный, квадратический и т.д.). Если можем измерить значения *y* только со случайными ошибками наблюдений, т.е. вместо $y_i = f(x_i; a_1, a_2, \ldots, a_m)$ наблюдаем значения Y_i, подверженные ошибкам, то согласно методу наименьших квадратов оценки неизвестных параметров a_i минимизируют сумму квадратов

$$\sum_{i=1}^{N} \left(Y_i - f(x_i; a_1, a_2, \ldots, a_m)\right)^2 .$$

Если $f(x) = e^{ax}$, то оценка параметра *a* соответственно минимизирует сумму

$$\sum_{i=1}^{N} \left(Y_i - e^{ax_i}\right)^2 .$$

В этом случае задача вычисления кривой регрессии *f* обычно упрощается, если вычислить логарифм от обоих членов разности, стоящих в скобках, и минимизировать величину

$$\sum_{i=1}^{N} \left(\ln Y_i - ax_i\right)^2 ,$$

что нетрудно сделать, находя минимум квадратического многочлена. Однако эти два подхода к минимизации дают разные оценки. Какой выход из этой парадоксальной ситуации?

Предположим, что тип функции f можно выбрать различными способами: например f_1 – это многочлен, f_2 – экспоненциальная функция. Кажется естественным предпочесть тот тип, для которого указанная выше сумма квадратов меньше (при оптимальном выборе параметров). Хотя этому принципу часто следуют на практике, обычно он не оправдан (иногда следует установить хотя бы теоретическую возможность такого выбора).

Пусть $y = ax$ есть теоретическая линия регрессии и $Y_i = ax_i + \varepsilon_i$, где ε_i, $(i = 1, 2, \ldots, N)$ являются независимыми нормально распределенными ошибками с математическим ожиданием 0 и дисперсией $D^2(\varepsilon_i) = c$ (c – известная постоянная). Теперь предположим, что наблюдения идеально согласуются с линией регрессии, т.е. $Y_i = a_0 x_i$ для некоторого a_0, и

$$\sum_{i=1}^{N} (Y_i - a_0 x_i)^2 = 0 .$$

Тогда оценкой параметра a по методу наименьших квадратов будет a_0; но, как ни парадоксально, она не является «лучшей» оценкой в смысле максимального правдоподобия.

Действительно, оценка $\hat{a} = a_0$ не подходит, так как в этом случае оценка для $D^2(\varepsilon_i)$ равнялась бы нулю, что противоречит условию $D^2(\varepsilon_i) = c$. Более оправданной будет оценка (максимального правдоподобия) вида

$$\left[\left(\sqrt{1+4c^2} - 1\right) \Big/ \left(2c^2\right)\right] a_0 .$$

Методу наименьших квадратов, несомненно, отвечает первая сумма, однако полезно разобраться не только в букве, но и духе метода наименьших квадратов, сутью которого является минимизация суммарного влияния ошибок. Эта цель может быть достигнута путем минимизации суммы квадратов

$$\sum_{i=1}^{N} \left(h(Y_i) - h\left(f\left(x_i; a_1, a_2, \ldots, a_m\right)\right)\right)^2 ,$$

где $h(x)$ – монотонно возрастающая функция (например, $h(x) = \ln x$). Хороший выбор h «линеаризует» задачу, т.е. делает выражение для $h(f(x_i; a_1, a_2, \dots, a_m))$ линейной функцией от неизвестных параметров a_i (в этом случае легко находятся оптимальные значения для a_i). Если мы хотим определить неизвестные параметры, следуя духу метода наименьших квадратов, то, очевидно, лучше выбрать второй способ. Однако иногда нужно минимизировать исходную сумму, например, когда известно, что результатом ошибок являются финансовые потери, пропорциональные этой сумме, хотя такая возможность совершенно нетипична.

Ответ, казалось бы, прост: если сумма квадратов окажется меньше для f_1, чем для f_2, то следует выбрать f_1. Однако ряде случаев, если взять чуть больше элементов выборки, то сумма квадратов становится меньше при выборе f_2. Математическая статистика старается избегать подобных неустойчивых ситуаций. Существует несколько методов принятия решений, которые применимы в ряде случаев и указывают выбор с заданной надежностью, например 99% (т.е. если функция f_1 отвергнута, то вероятность того, что правильным являлся выбор f_1, равна 1%). …

К сожалению, многие из типичных задач по выбору вида регрессии невозможно решить должным образом. Например, правило Вебера-Фехнера утверждает, что между раздражителем и ощущением существует логарифмическая зависимость, в частности между объемом и интенсивностью звука или между частотой и высотой звука. В настоящее время это правило теоретически и практически рассматривается лишь как первое приближение. Представляется, что ближе к истине является степенная зависимость». [1, стр. 77-80]

6. Заключение

Очевидно, что приведенный анализ фактически дезавуирует полезность книги [1]. При этом, практически все, изложенное в настоящей рецензии, было рассмотрено в свое время работах [2, 3] и целом ряде других, хорошо известных автору книги [1]. Какую цель в этом случае преследовал автор учебного пособия [1], включая проанализированные положения в свою книгу? Снизить уровень подготовки отечественных студентов? Или просто для увеличения объема книги? В последнем случае можно было бы рассмотреть, например, тонкие моменты, связанные со структурой зависимости случайных величин, следуя, возможно, хорошо известной книге [4].

Литература

1. 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.
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. Чернышев К.Р. Вопросы идентификации: состоятельные меры зависимости. М. Институт проблем управления им. В.А. Трапезникова РАН, 2003. 60 с.
4. Mari D.D., Kotz S. Correlation and Dependence. Singapore: World Scientific, 2001. 219 p.