

ПОСТРОЕНИЕ МОДЕЛИ МНОГОСВЯЗНОГО ОБЪЕКТА НА ОСНОВЕ СОВМЕСТНОГО ИСПОЛЬЗОВАНИЯ ДАННЫХ И ЭКСПЕРТНЫХ ОЦЕНОК

В.Е. Гвоздев¹, Р.А. Мунасыпов², О.Я. Бежаева³, Д.Р. Ахметова⁴

Уфимский государственный авиационный технический университет, Уфа, Россия

¹wega55@mail.ru, ²rust40@mail.ru, ³obezhaeva@gmail.com, ⁴dinara.akhmetova.92@gmail.com

Аннотация

Рассматривается метод построения статистических моделей многосвязных объектов в виде многомерных регрессионных зависимостей на основе совместного использования измерительных данных и экспертных оценок. Основная идея метода заключается в преобразовании независимо полученных экспертных оценок и измерительных данных к единому виду – закону распределения непрерывной случайной величины. Это делает возможным сформировать корреляционную таблицу на основе решения обратной задачи определения закона распределения функции случайного аргумента. Сформированная корреляционная таблица служит основой построения регрессионной зависимости известными методами регрессионного анализа. Главным результатом работы является впервые предложенный метод построения многомерных регрессионных зависимостей за счёт совместного использования экспертных оценок и измерительных данных.

Ключевые слова: *многосвязный объект, многомерная регрессионная зависимость, закон распределения, случайная величина, экспертная оценка.*

Цитирование: *Гвоздев, В.Е. Построение модели многосвязного объекта на основе совместного использования данных и экспертных оценок / В.Е. Гвоздев, Р.А. Мунасыпов, О.Я. Бежаева, Д.Р. Ахметова // Онтология проектирования. – 2019. – Т.9, №3(33). – С.361-368. – DOI: 10.18287/2223-9537-2019-9-3-361-368.*

Введение

Одной из задач построения дескриптивных статических моделей многомерных многосвязных объектов является задача построения многомерных регрессионных зависимостей. Классический подход к решению этой задачи основан на использовании таблицы совместно наблюдаемых значений независимой и зависимой случайных величин.

На практике формирование таблицы совместно наблюдаемых значений может встречать определённые трудности. Это обусловлено, например, тем, что сбор различных данных осуществляется несколькими организациями, деятельность которых регламентируется разными документами. Либо тем, что происходят потери архивных данных. Либо вследствие того, что измерения параметров состояния сложных систем ранее не осуществлялись, но для их оценки можно использовать мнение экспертов. В качестве примера можно привести задачи, связанные с управлением дефектами в программных компонентах аппаратно-программных комплексов на разных стадиях их жизненного цикла [1].

Значения независимых и зависимых параметров могут задаваться в виде измерительных данных и в виде экспертных оценок разного вида: в виде ожидаемых значений; интервала возможных значений; совокупности ожидаемых значений и интервала возможных значений случайной величины. Разработка метода преобразования измерительных данных и экспертных оценок, задаваемых в разных формах, к виду таблицы совместно наблюдаемых значений, послужит основой построения многомерных регрессионных зависимостей.

1 Методы построения функциональных зависимостей на основе измерительных данных и экспертных оценок

Классическому подходу к построению регрессионных зависимостей можно поставить в соответствие схему

$$(1) \quad A^{(0)}: \{\vec{x}, \vec{y}\}_1^N \rightarrow \vec{y} = \varphi_0(\vec{x}, \vec{\Theta}).$$

Здесь $\{\vec{x}, \vec{y}\}_1^N$ - множество совместно наблюдаемых значений компонент векторов независимой \vec{x} и зависимой \vec{y} величин;

$\varphi_0(\vec{x}, \vec{\Theta})$ – функциональная зависимость, задаваемая в параметрической форме;

N – число пар значений \vec{x} и \vec{y} .

В литературе описаны различные операторы $A^{(0)}$, например, реализующие метод наименьших квадратов [2], либо метод наименьших модулей [3].

При решении практических задач, вследствие сложности организации сбора исходных данных, формирование $\{\vec{x}, \vec{y}\}_1^N$ может встретить серьезные трудности.

В работе [4] описан подход к построению одномерных непараметрических строгих функциональных зависимостей на основе решения обратной задачи построения закона распределения функции случайного аргумента (прямая задача описана в [5]).

Схема решения обратной задачи имеет вид

$$(2) \quad A^{(1)}: \{F(x), F(y)\} \rightarrow y = \varphi_1(x),$$

где $F(x)$, $F(y)$ – оценки одномерных законов распределения независимой и зависимой случайных величин соответственно, определяемые на основе обработки выборочных данных

$$(3) \quad A^{(2)}: \{x\}_1^N \rightarrow x = F(x), A^{(2)}: \{y\}_1^M \rightarrow y = F(y).$$

Особенностью схемы (2) является то, что свойства выборочных данных $\{x\}_1^N$, $\{y\}_1^M$ (объем, точность регистрации) могут быть различными. Ограничением схемы (2) является необходимость обоснования самого факта наличия строгих зависимостей, например, исходя из физического содержания задачи. В работе [5] приведены результаты исследования свойств оценок строгих функциональных зависимостей $\varphi_1(x)$, получаемых посредством (2).

Если объемы выборочных данных $\{x\}_1^N$, $\{y\}_1^M$ являются большими в статистическом смысле, то в качестве $A^{(2)}$ могут использоваться «традиционные» методы, описанные, например, в [2, 5, 6]. Если же объем выборочных данных мал, а тип закона распределения априорно неизвестен, то в качестве $A^{(2)}$ целесообразно использовать методы, основанные на принципе максимизации энтропии [7-9]. Построение оценки закона распределения непрерывной случайной величины сводится к определению параметров μ_j выражения

$$(4) \quad f(z) = \exp(\sum_j \mu_j z^j), j = \overline{1; n}$$

в результате решения системы уравнений вида

$$(5) \quad \int_{a_z}^{b_z} z^k \exp(\sum_j \mu_j z^j) dz = \nu_z^{(k)},$$

где z - непрерывная случайная величина;

$\{a_z, b_z\}$ – границы интервала физически возможных значений случайной величины;

$\nu_z^{(k)}$ - начальный момент k -го порядка, определяемый по выборочным данным.

При решении практических задач целесообразно ограничиться двумя первыми моментами [4]. Широко используемые на практике нормальный, показательный и равномерный законы распределения случайных величин, являются частным случаем выражения (4).

Экспертные оценки представляются в виде оценки ожидаемого значения случайной величины $M[z]$ и/или интервала возможных значений случайной величины Z (возможны следующие варианты представления интервалов: $z \in [a_z, b_z]$; $z \in [a_z, \infty)$) [10].

Если известны лишь границы интервала возможных значений случайной величины, т.е. $z \in [a_z, b_z]$, то

$$(6) \quad A^{(3)}: [a_z, b_z] \rightarrow F^{(3)}(z).$$

Здесь $A^{(3)}$ представляет собой частный случай (5), когда $k=0$. В этом случае оптимальной оценкой $F^{(3)}(z)$ является равномерный закон распределения.

Если известны ожидаемые значения $M[z]$, а интервал возможных значений случайной величины представляется в виде $[a_z, \infty)$, то построение оценки сводится к решению задачи

$$(7) \quad A^{(4)}: \{M[z], [a_z, \infty)\} \rightarrow F^{(4)}(z).$$

Здесь $A^{(4)}$ представляет собой частный случай (5) при $k=1$. Оптимальной оценкой $F^{(4)}(z)$ в этом случае является показательный закон распределения

$$(8) \quad F^{(4)}(z) = 1 - e^{-\lambda z}.$$

Параметр закона распределения определяется соотношением $\lambda = (M[z] - a_z)^{-1}$.

Если известны $M[z]$ и границы интервала $[a_z, b_z]$, то оценка закона распределения ищется в результате решения задачи

$$(9) \quad A^{(5)}: \{M(z), [a_z, b_z]\} \rightarrow F^{(5)}(z).$$

Оценка представляется в виде

$$(10) \quad F^{(5)}(z) = \int_{a_z}^z e^{\mu_0 + \mu_1 \tau} d\tau,$$

причём параметры модели μ_0, μ_1 находятся в результате решения системы уравнений

$$(11) \quad \begin{cases} \int_{a_z}^{b_z} e^{\mu_0 + \mu_1} d\tau = 1 \\ \int_{a_z}^{b_z} \tau e^{\mu_0 + \mu_1 \tau} d\tau = M[z] \end{cases}.$$

В частном случае, если дополнительно известно, что $F^{(5)}(a_z) = 0, F^{(5)}(b_z) = 1, F^{(5)}(z)$ представляет собой треугольное распределение.

Преобразование экспертных оценок к виду законов распределения непрерывных случайных величин делает возможным расширить схему (2) на случай построения одномерных непараметрических регрессионных зависимостей на основе совместного использования измерительных данных и экспертных оценок.

Вместе с тем, методы построения функциональных зависимостей на основе решения обратной задачи построения закона распределения функции случайного аргумента, в том числе на основе совместного использования измерительных данных и экспертных оценок, ориентированы на исследование одномерных зависимостей и не адаптированы к исследованию поведения многомерных систем.

2 Построение многомерных регрессионных зависимостей на основе одномерных непараметрических регрессионных моделей

К числу статических моделей, используемых для описания многомерных безынерционных многосвязных объектов относятся многомерные регрессионные зависимости

$$(12) \quad y_j = \varphi_j(x_{j1}, \dots, x_{jk_j}), j = \overline{1; N},$$

где y_j – компоненты вектора Y зависимой переменной;

$x_{jl}, l_j = \overline{1; k_j}$ – компоненты вектора независимой переменной;

$\varphi_j(\cdot)$ – зависимость (параметрическая либо непараметрическая) значений j -й компоненты вектора зависимой переменной от значений компонент вектора независимой переменной.

Классический подход к построению многомерных регрессионных зависимостей основан на обработке совместно наблюдаемых значений [5, 6], представленной в табличной форме (см. таблицу 1).

Таблица 1 – Головка таблицы совместно наблюдаемых значений независимой и зависимой переменных

Номер измерения	Независимые переменные				Зависимая переменная y_j
	x_{j1}	x_{j2}	...	x_{jl_j}	

Многомерные уравнения регрессии ассоциируются с многомерными многосвязными объектами, модель которого представлена на рисунке 1. Построение соотношения вида (12) в зависимости от формы представления компонент

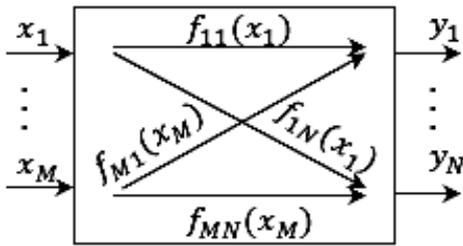


Рисунок 1 – Модель многомерного многосвязного объекта

вектора независимой переменной x_{jl_j} и зависимой переменной y_j (измерительные данные, экспертные оценки) сводится к использованию одного из преобразований (6), (8), (10) для построения одномерных законов распределения случайных величин с последующей оценкой непараметрических функциональных зависимостей на основе (2).

Наличие одномерных функциональных зависимостей $y_j = f_{ij}(x_j)$ позволяет свести формирование таблицы 1 к последовательной

реализации следующих шагов.

Шаг 1. Посредством датчиков равномерно распределенных случайных чисел $x_i \in [0,1]$ методом обратных преобразований [11] генерируются независимые q ($q = \overline{1; Q}$) значения случайных величин $\xi_j^{(q)}$, соответствующие оценкам законов распределения $\hat{F}(x_j)$

$$(13) \quad \xi_{ji}^{(q)} = \hat{F}^{-1}(x_{ji}), i = \overline{1; M}.$$

Шаг 2. По значениям $\xi_{ij}^{(q)}$, используя построенные одномерные параметрические зависимости $f_{ij}(x_j)$ рассчитываются значения y_j по правилам

$$(14) \quad y_j^{(q)} = \sum_i f_{ji}(\xi_{ji}^{(q)}).$$

Шаг 3. Значения $\{\xi_{ji}^{(q)}, y_j^{(q)}\}$ заносятся в виде строки в таблицу 1.

В последующем на основе сформированной таблицы, используя известные методы многомерного регрессионного анализа, строится параметрическая многомерная регрессионная зависимость.

Пример. Допустим, имеется двумерный многосвязный объект. Первому входному параметру ставится в соответствие экспертная оценка вида $x_1 \in [0; 8]$. Второму входному параметру ставится в соответствие выборка случайных величин, которая может быть аппроксимирована нормальным законом распределения $N(4,1)$. Зависимой переменной y_1 ставится в соответствие выборка случайных величин, которая может быть аппроксимирована нормальным законом распределения $N(8,2)$. Зависимой переменной y_2 ставится в соответствие экспертная оценка вида $\{2, [0, \infty)\}$.

Зависимые и независимые переменные связываются соотношениями

$$(15) \quad \begin{cases} y_1 = f_{11}(x_1) + f_{21}(x_2) \\ y_2 = f_{22}(x_2) + f_{12}(x_1) \end{cases}$$

С учётом формы представления экспертной оценки, соответствующей x_1 , в качестве $F_1(x)$ используется равномерный закон распределения $F(x_1) = \frac{1}{8} \int_0^{x_1} dt$. С учётом, что $F(y_1)$

описывается нормальным законом распределения, непараметрическая оценка $f_{11}(x_1)$ в соответствии с (2) приобретает вид, представленный на рисунке 2.

С учётом формы представления экспертной оценки, соответствующей y_2 , в соответствии с (8) в качестве $F(y_2)$ используется показательный закон распределения с параметром $\lambda = 0,5$. Учитывая, что закон распределения $F(x_2)$ является $N(4, 1)$, непараметрическая оценка $f_{22}(x_2)$ в соответствии с (2) приобретает вид, представленный на рисунке 3.

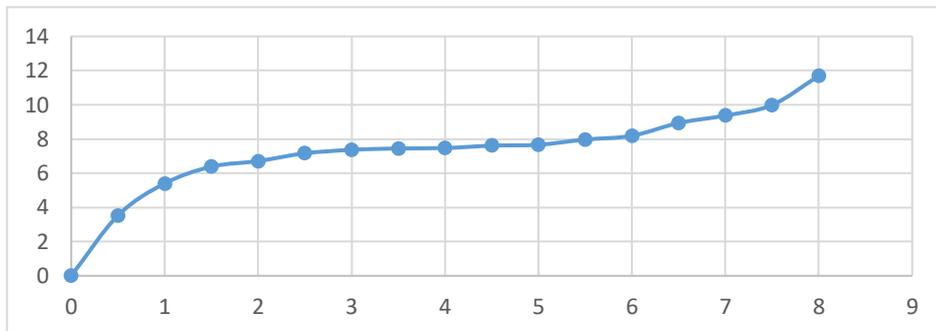


Рисунок 2 – Непараметрическая оценка $f_{11}(x_1)$

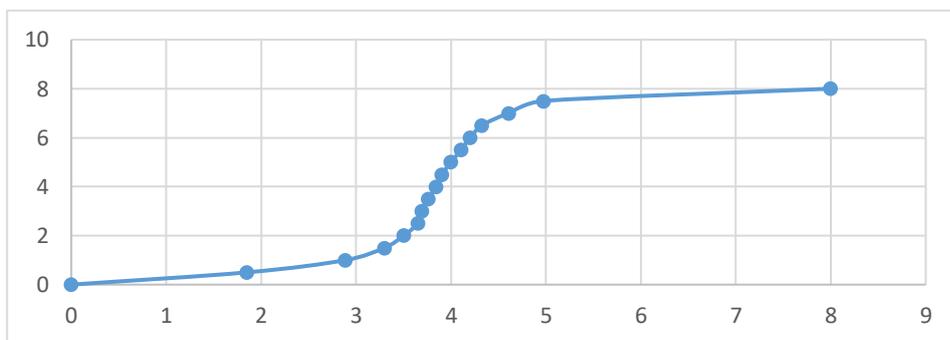


Рисунок 3 – Непараметрическая оценка $f_{22}(x_2)$

С учётом того, что закон распределения $F(x_1)$ является равномерным, а $F(y_2)$ – показательным, использование (2) позволяет получить оценку $f_{12}(x_1)$, представленную на рисунке 4. Учитывая, что $F(y_2)$ имеют одинаковый тип, использование (2) приводит к представлению $f_{21}(x_2)$ в виде линейной зависимости (рисунок 5).

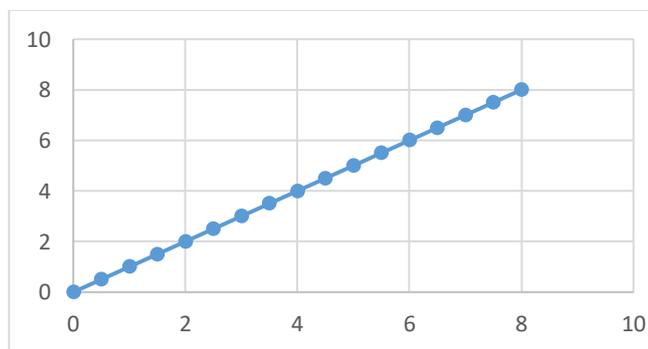


Рисунок 4 – Непараметрическая оценка $f_{12}(x_1)$

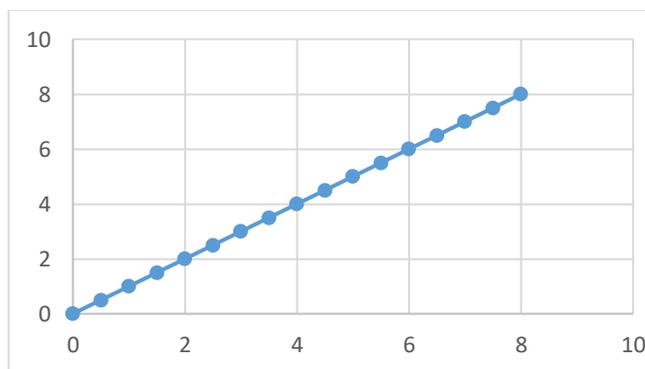


Рисунок 5 – Непараметрическая оценка $f_{21}(x_2)$

Используя описанный трёхшаговый алгоритм, сформированная таблица, содержащая значения независимых x_1 и x_2 и зависимых y_1 и y_2 переменных. Фрагмент её представлен в таблице 2. При построении таблицы значение Q принималось $Q=1000$.

На основе этой таблицы с использованием пакета *Matlab* построены двумерные регрессионные зависимости, представленные на рисунках 6 и 7.

Описанный подход может быть перенесён на случай, когда N в соотношении (12) больше двух.

Таблица 2 – Фрагмент таблицы совместно наблюдаемых значений независимых и зависимых переменных

Номер измерения	x_1	x_2	y_1	y_2
1	0	0	0	0
2	0,5	1,84	3,52	0,5
3	1	2,88	5,40	1
4	1,5	3,29	6,38	1,5

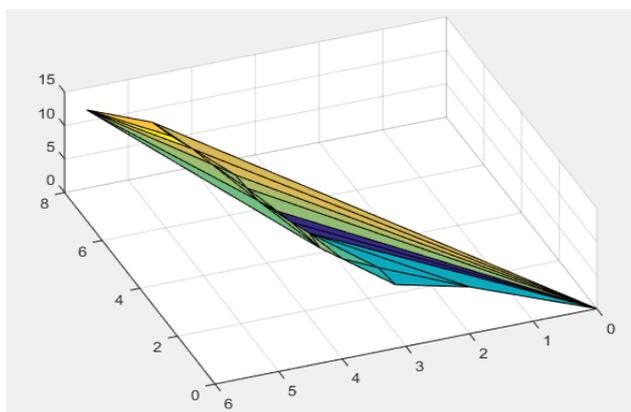


Рисунок 6 – Зависимость $y_1 = \varphi_1(x_1, x_2)$

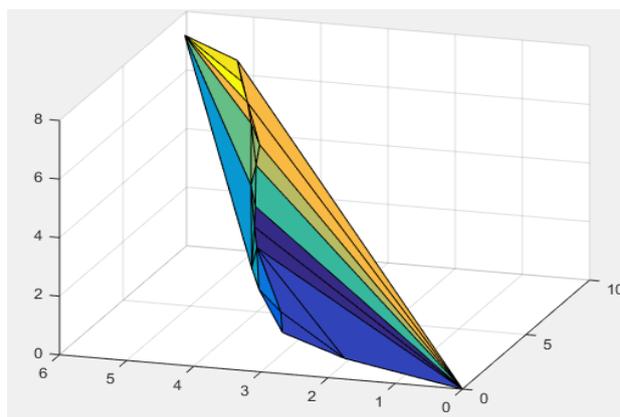


Рисунок 7 – Зависимость $y_2 = \varphi_2(x_1, x_2)$

Заключение

Предложен подход к построению многомерных регрессионных зависимостей для случая, когда исходные данные, соответствующие разным компонентам векторов независимой и зависимой переменных, представляются либо в виде результатов измерений, либо в виде разных по форме экспертных оценок. Основу метода составляет преобразование исходных данных, представляемых в разной форме, к виду закона распределения непрерывной случайной величины. Это делает возможным построение совокупности одномерных монотонных на интервале непараметрических регрессионных зависимостей на основе попарного исследования законов распределения независимых и зависимых случайных величин. Полученные регрессионные зависимости служат основой формирования таблицы совместно наблюдаемых значений независимых и зависимых случайных величин, что делает возможным строить многомерные регрессионные зависимости известными методами. Ограничением подхода является, во-первых, необходимость априорного логического обоснования наличия связи между независимыми и зависимыми переменными. Во-вторых, метод ориентирован на построение лишь монотонных на интервале регрессионных зависимостей.

Благодарности

Работа поддержана грантом 19-08-00177 Методологические, теоретические и модельные основы управления функциональной безопасностью аппаратно-программных комплексов в составе распределенных сложных технических систем.

СПИСОК ИСТОЧНИКОВ

- [1] **Гвоздев, В.Е.** Предупреждение дефектов на ранних стадиях проектирования аппаратно-программных комплексов на основе положений теории интересубъективного управления / В.Е. Гвоздев, Д.В. Блинова, Л.Р. Черняховская // *Онтология проектирования*. – 2016. – Т. 6, №4(22). – С. 452-464. – DOI: 10.18287/2223-9537-2016-6-4-452-464.
- [2] **Линник, Ю.В.** Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. - Изд. 2-е, доп. и испр. - М.: Гос. изд-во физ.-мат. лит., 1962. - 349 с.
- [3] **Мудров, В.И.** Метод наименьших модулей / В.И. Мудров, В.Л. Кушко. - М.: Радио и связь, 1983. - 304 с.
- [4] **Гузаиров, М.Б.** Статистическое исследование территориальных систем: монография / М. Б. Гузаиров и др. - Москва: Машиностроение, 2008. - 187 с.
- [5] **Пугачев, В.С.** Теория вероятностей и математическая статистика. - М.: Физматлит, 2002. - 496 с.
- [6] **Смирнов, Н.В.** Курс теории вероятностей и математической статистики для технических приложений / Н.В. Смирнов, И.В. Дунин-Барковский. - М.: Наука, 1969. - 512 с.
- [7] **Трайбус, М.** Термостатика и термодинамика. – М.: Энергия, 1970. – 504 с.
- [8] **Каган, А.М.** Характеризационные задачи математической статистики / А.М. Каган, Ю.В. Линник, С.Р. Рао – М.: Наука, 1972. – 656 с.
- [9] **Кузин, Л.Т.** Основы кибернетики. Т.1. – М.: Энергия, 1973. – 503 с.
- [10] **Гвоздев, В.Е.** Элементы системной инженерии. Технологии формирования требований к аппаратно-программным комплексам на основе экспертно-статистических методов / Н. К. Криони, В. Е. Гвоздев, Б. Г. Ильясов, О. Я. Бежаева, Д. В. Блинова. - М.: Инновационное машиностроение, 2017. - 295 с.
- [11] **Соболь, И.М.** Численные методы Монте-Карло. - М.: Наука, 1973. – 312 с.

CONSTRUCTION OF A MULTI-CONNECTED OBJECT MODEL OBJECT BASED ON JOINT USE OF DATA AND EXPERT EVALUATIONS

V.E. Gvozdev¹, R.A. Munasyrov², O.Y. Bezhayeva³ D.R. Akhmetova⁴

Ufa State Aviation Technical University, Ufa, Russia

¹wega55@mail.ru, ²rust40@mail.ru, ³obezhaeva@gmail.com, ⁴dinara.akhmetova.92@gmail.com

Abstract

The article discusses a method for constructing statistical models of multi-connected objects in the form of multidimensional regression dependencies based on the joint use of measurement data and expert assessments. The main idea of the method is to convert independently obtained expert estimates and measurement data to a single type - the law of distribution of a continuous random variable. This makes it possible to form a correlation table based on the solution of the inverse problem of determining the distribution law of the function of a random argument. The formed correlation table serves as the basis for the construction of the regression dependence by the known methods of regression analysis. The main result of the work is the first proposed method for constructing multidimensional regressive dependencies by sharing expert estimates and measurement data.

Key words: multi-connected object, multidimensional regression dependence, law of distribution, random variable, expert evaluations.

Citation: Gvozdev VE, Munasyrov RA, Bezhayeva OY, Akhmetova DR. Construction of a multi-connected object model based on joint use of data and expert evaluations [In Russian]. *Ontology of designing*. 2019; 9(3): 361-368. – DOI: 10.18287/2223-9537-2019-9-3-361-368.

Acknowledgments

This work was supported by grant 19-08-00177 Methodological, theoretical, and model foundations for managing the functional safety of hardware and software systems as part of distributed complex technical systems.

References

- [1] *Gvozdev VE, Blinova DV, Chernyakhovskaya LR.* Hardware-software complexes on the basis of the positions of the theory of intersubjective management [In Russian]. *Ontology of Designing.* 2016; 6(4): 452-464. DOI: 10.18287/2223-9537-2016-6-4-452-464.
- [2] *Linnik U.V.* The method of least squares and the basics of the mathematical-statistical theory of processing observations [In Russian]. – Moscow: State Publishing House for Physics and Mathematics; 1962.
- [3] *Mudrov VI, Kushko VL.* Method of the smallest modules [In Russian]. – Moscow: Radio and communication.; 1983.
- [4] *Guzairov MB.* Statistical research of territorial systems: monograph [In Russian]. – Moscow: Engineering.; 2008.
- [5] *Pugachev VS.* Theory of Probability and Mathematical Statistics [In Russian]. – Moscow: Fizmatlit.; 2002.
- [6] *Smirnov NV, Dunin-Barcovskiy IV.* Course of probability theory and mathematical statistics for technical applications [In Russian]. – Moscow: The science.; 1969.
- [7] *Traibus M.* Thermostatistics and thermodynamics [In Russian]. - Moscow: Energy.; 1970.
- [8] *Kagan AM, Linnik UV, Rao SR.* Characterization problems of mathematical statistics [In Russian]. - Moscow.: Science; 1972.
- [9] *Kuzin LT.* Basics of Cybernetics. V.1. [In Russian]. - Moscow: Energy; 1973.
- [10] *Gvozdev VE, Krioni NK, Il'yasov BG, Bezhayeva OY, Blinova DV.* Elements of system engineering. Technologies for the formation of requirements for hardware-software complexes based on expert-statistical methods [In Russian]. – Moscow: Innovative engineering.; 2017.
- [11] *Sobol IM.* Monte Carlo Numerical Methods [In Russian]. – Moscow: The science.; 1973.

Сведения об авторах



Гвоздев Владимир Ефимович, 1956 г. рождения. Окончил Уфимский авиационный институт им. Орджоникидзе в 1973 г., д.т.н. (2000). Заведующий кафедрой технической кибернетики Уфимского государственного авиационного технического университета (УГАТУ). В списке научных трудов более 300 работ в области прикладного статистического анализа, информационной поддержки управления программными системами, информационной поддержки управления состоянием территориальных систем.

Vladimir Efimovich Gvozdev (b. 1956) graduated from the Ufa Aviation Institute (Ufa-city) in 1973, D. Sc. Eng. (2000). He is Head of Department at Ufa State Aviation Technical University (Department of Technical Cybernetics). He is co-author of about 300 scientific articles and

abstracts in the field of applied statistical analysis, information support for managing software systems, information support for managing the state of territorial systems.

Мунасыпов Рустэм Анварович, 1960 г. рождения. Окончил Уфимский авиационный институт им. Орджоникидзе в 1977 г., д.т.н. (2004). Профессор кафедры автоматизации технологических процессов УГАТУ. В списке научных трудов более 240 работ в области адаптивных и интеллектуальных систем управления сложными техническими объектами.

Rustem Anvarovich Munasypov (b. 1960) graduated from the Ufa Aviation Institute (Ufa-city) in 1977, D.Sc.Eng. (2004). He is Professor at Ufa State Aviation Technical University (Department of technological process automation). He is co-author of about 240 scientific articles and abstracts in the field of adaptive and intelligent control systems of complex technical objects.



Бежаева Оксана Яковлевна, 1977 г. рождения. Окончила УГАТУ в 2000 г., к.т.н. (2004). Доцент кафедры технической кибернетики УГАТУ. В списке научных трудов более 100 работ в области разработки моделей и программного обеспечения сложных производственных и социально-экономических систем.

Oksana Yakovlevna Bezhayeva (b. 1977) graduated from the Ufa State Aviation Technical University in 2000, PhD. (2004). She is Assistant professor at Ufa State Aviation Technical University (Department of Technical Cybernetics). She is co-author of about 100 scientific articles and abstracts in the field of development of models and software for complex production and socio-economic systems.



Ахметова Динара Раилевна, 1992 г. рождения. Окончила УГАТУ в 2014 г. Аспирант кафедры технической кибернетики УГАТУ. В списке научных трудов 20 работ в области программно-аппаратных комплексов технических систем.

Dinara Railevna Akhmetova (b. 1992) graduated from the Ufa State Aviation Technical University in 2014. She is Graduate student at Ufa State Aviation Technical University (Department of Technical Cybernetics). She is co-author of 20 scientific articles and abstracts in the field of software and hardware systems of technical systems.