

УДК 519.2

## ОЦЕНИВАНИЕ ЛИНЕЙНЫХ ВЗАИМОСВЯЗЕЙ ПАРАМЕТРОВ ТЕХНИЧЕСКИХ ОБЪЕКТОВ ПРИ ОТСУТСТВИИ КОРРЕЛЯЦИОННОЙ ТАБЛИЦЫ ЭМПИРИЧЕСКИХ ДАННЫХ

В. Е. Гвоздев<sup>1</sup>, А. С. Субхангулова<sup>2</sup>, О. Я. Бежаева<sup>3</sup>

<sup>1</sup>wega55@mail.ru, <sup>2</sup>aliyasr21@gmail.com, <sup>3</sup>obezhaeva@gmail.com

ФГБОУ ВПО «Уфимский государственный авиационный технический университет» (УГАТУ)

*Поступила в редакцию 27 июня 2014 г.*

**Аннотация.** Рассматривается задача построения линейной одномерной функциональной зависимости при разных законах распределения случайных величин, объемах и погрешности регистрации независимой и зависимой переменных; предлагается подход к оцениванию параметров линейной зависимости, при отсутствии корреляционной таблицы эмпирических данных. Устанавливается связь предлагаемого подхода с методом наименьших квадратов.

**Ключевые слова:** случайная величина; закон распределения случайного аргумента; линейные регрессионные зависимости; коэффициент корреляции.

### ВВЕДЕНИЕ

Задача анализа связей между параметрами, характеризующими состояния объектов, относится к основным задачам информационной поддержки управления сложными объектами. Примерами задачи анализа связей являются: формирование матриц взаимосвязи при поэтапном преобразовании ожиданий пользователей в характеристики изделия, процесса производства изделий и систем контроля за ходом процесса в рамках QFD (Quality Function Deployment) – методологии [1]; анализ влияния характеристик потока заявок на значения характеристик качества систем массового обслуживания (СМО) [2]; влияние условий испытаний и эксплуатации на характеристики надежности технических систем [3]; когнитивное моделирование слабо структурированных систем [4, 5] и др.

Необходимость оценивания характеристик связи, с одной стороны, и объективные ограничения на возможность получения измерительных данных, с другой, привели к разработке разных подходов к анализу связей: на основе экспертных оценок и использования лингвистических переменных [1]; на основе исследования непрямых влияний, на базе аппарата знаково-ориентированных графов и нечетких когнитивных карт; на основе знаковых корреляционных

функций, парных, множественных и парциальных коэффициентов корреляции [6]; на основе коэффициентов ассоциативности [7] и др.

Одной из наиболее известных задач оценивания взаимосвязей случайных величин является задача оценивания линейных регрессионных зависимостей (одномерных и многомерных) на основе выборочных данных. В литературе приведено описание большого числа инженерных задач, в основе решения которых лежит построение линейных функциональных зависимостей по экспериментальным данным. В качестве примеров задач укажем следующие: задачу определения характеристик линейных случайных функций при анализе параметрической надежности изделий [3]; задачу исследования зависимости характеристик надежности от условий работы объекта [3]; построение линейных дескриптивных моделей по результатам полного факторного эксперимента [8]. Однако, несмотря на большое число проведенных исследований, задача анализа линейных зависимостей на основе экспериментальных данных далека от окончательного решения.

В статье рассматривается метод построения линейной одномерной регрессионной зависимости при независимой регистрации значений случайных параметров объекта, т. е. при разных объемах и погрешности регистрации независимой и зависимой переменных.

## ПОДХОДЫ К ПОСТРОЕНИЮ ЛИНЕЙНЫХ ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЕЙ

Наиболее известным подходом к построению линейных функциональных зависимостей является подход, основанный на анализе корреляционной таблицы наблюдаемых значений независимой  $X$  и зависимой  $Y$  переменных.

В одномерном случае формальная задача имеет вид:

$$A: \{x, y\}_1^N \rightarrow \{a, b\}, \quad (1)$$

где  $\{x, y\}_1^N$  – таблица, содержащая совокупность из  $N$  пар чисел  $(x, y)$ , полученных в одинаковых условиях при наблюдениях за одним и тем же объектом, причем измерения выполняются одним и тем же способом;  $\{a, b\}$  – параметры линейной зависимости  $y = a + bx$ ;  $A$  – оператор, позволяющий получать значения  $\{a, b\}$  на основе исходной таблицы  $\{x, y\}_1^N$ .

В литературе описаны разные способы оценивания  $\{a, b\}$ : метод наименьших квадратов [3, 9, 10]; метод наименьших модулей [9]. Среди этих методов наибольшее распространение получил метод наименьших квадратов (хотя в литературе, например в [10], описано много примеров его некорректного использования).

Другим подходом к оцениванию параметров линейной зависимости является подход, основанный на анализе закона распределения функции случайного аргумента [11, 12]. В основе этого подхода лежит соотношение:

$$F_2(y) = \int_{y < \varphi(x)} f_1(x) dx, \quad (2)$$

где  $F_2(y)$  – функция распределения случайного аргумента;  $f_1(x)$  – плотность распределения случайного аргумента.

В [12] рассмотрен случай построения линейной зависимости при нормальных законах распределения  $N(x, M[x], \sigma_x)$ ;  $N(y, M[y], \sigma_y)$ . В [13] приведены результаты исследования модели (2), соответствующие случаю, когда  $y = \varphi(x)$  является строгой функциональной зависимостью, а  $F_2(y)$  и  $f_1(x)$  относятся к классу унимодальных непрерывных законов распределения случайных величин, у которых  $0 \leq X < \infty$ ;  $0 \leq Y < \infty$ .

## ВЛИЯНИЕ ТИПА ЗАКОНА РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОГО АРГУМЕНТА НА ХАРАКТЕР ПОЛЯ РАССЕЙЯНИЯ

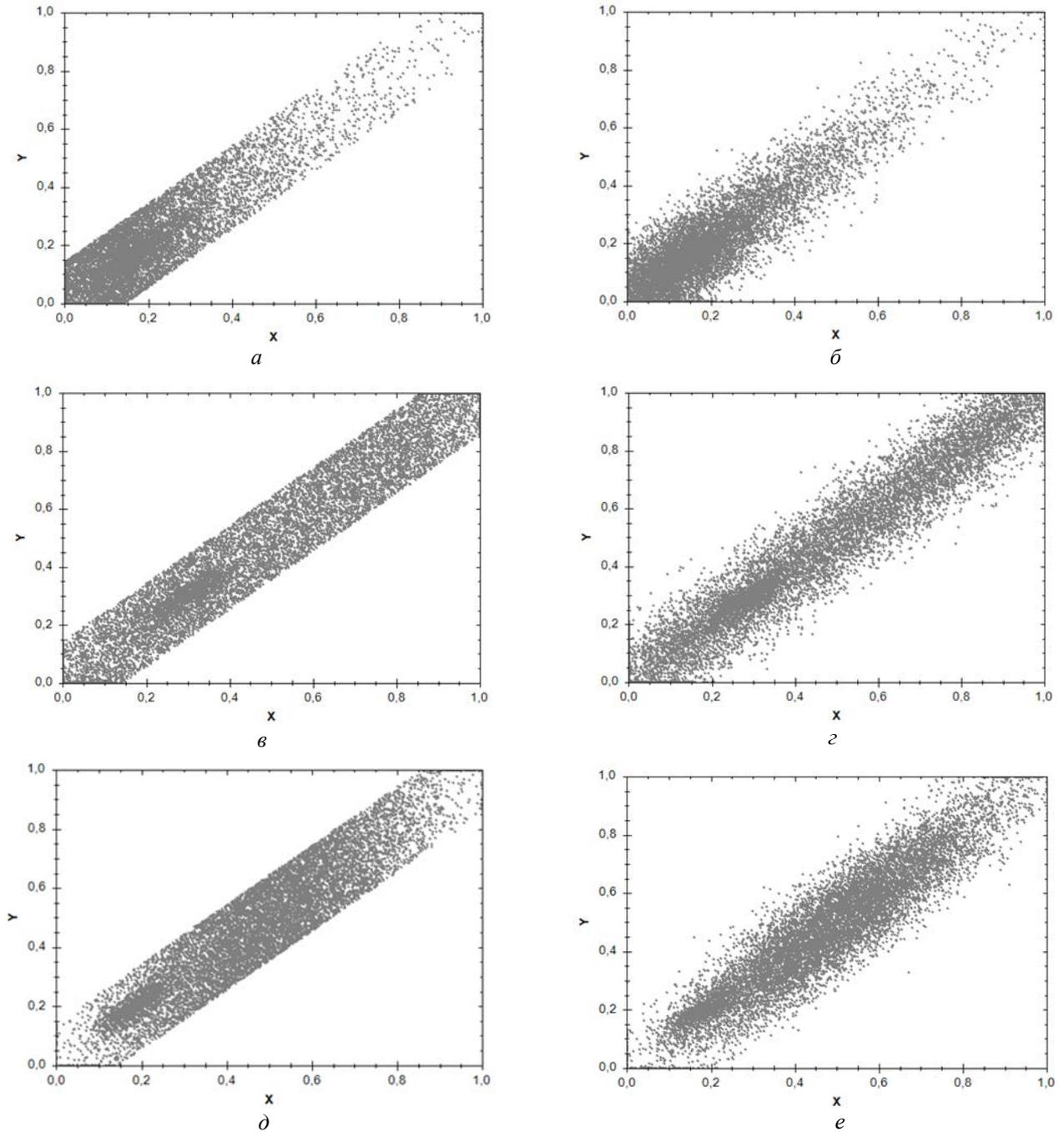
В известной литературе, посвященной построению регрессионных зависимостей на основе корреляционной таблицы, отмечается, что

основу анализа составляет так называемое «поле рассеяния» [2]. Оценим влияние типа  $f_1(x)$  и погрешности регистрации выборочных данных на характер поля рассеяния.

На рис. 1 в качестве примера показаны поля рассеяния в случае, когда  $f_1(x)$  соответствовало показательному, нормальному и равномерному законам распределения при объемах регистрируемых значений  $X$  и  $Y$   $N = 10$ . В случае равномерного закона распределения  $X \in [0; 1]$ ; в случае показательного и нормального законов параметры распределения подбирались таким образом, чтобы выполнялось условие  $\int_0^1 f_1(x) dx > 0.98$ . Это условие соблюдается, когда  $\sigma_x = 1/6$ .

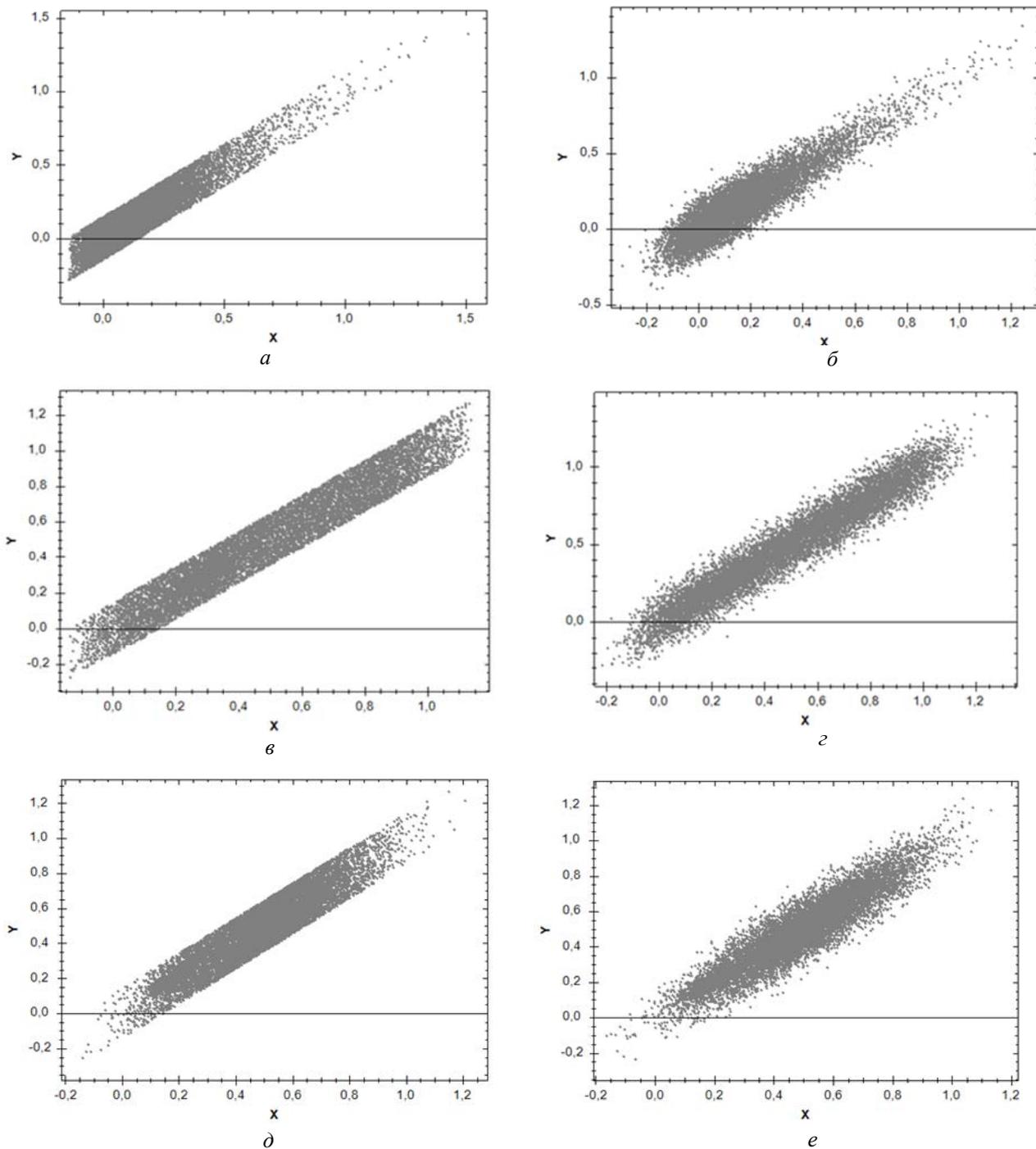
В ходе исследований рассматривался случай, когда зависимая переменная была связана с независимой соотношением  $y = x$ . На независимые выборочные значения  $x_i$  ( $i = \overline{1; N}$ ) накладывалась помеха  $\varepsilon_i^{(x)}$  ( $i = \overline{1; N}$ ), а на  $y_i$  ( $i = \overline{1; M}$ ) накладывалась помеха  $\varepsilon_j^{(y)}$  ( $j = \overline{1; M}$ ), так что при зависимости вида  $y = x$  фактически рассматривалась  $x_i^* = x_i + \varepsilon_i$ ;  $y_i^* = x_i + \varepsilon_i^{(x)} + \varepsilon_j^{(y)}$ . В качестве законов распределения ошибки  $\Psi(\varepsilon)$  выступали равномерный либо нормальный законы распределения с нулевым математическим ожиданием  $M[\varepsilon] = 0$ . Параметр масштаба  $\sigma_\varepsilon$  законов  $\Psi(\varepsilon)$  определялся из соотношения  $\sigma_\varepsilon = \alpha \sigma_x$ ,  $\alpha \in \{0; 0.1; 0.5; 0.9\}$ . Приведенные рисунки соответствуют случаю  $\alpha = 0.5$ . При построении полей рассеяния следовали правилу: если значения  $x_i^*$ ,  $y_i^*$  оказывались меньше нуля, их значения принимались нулевыми; если значения  $x_i^*$ ,  $y_i^*$  оказывались больше единицы, они принимались равными единице, учитывая, что по условию задачи  $X \in [0; 1]$ ;  $Y \in [0; 1]$ . Для сравнения на рис. 2 приведены поля рассеяния, соответствующие случаю, когда на области распределения  $\Omega_X$  и  $\Omega_Y$  не накладывалось каких-либо ограничений.

Из полученных результатов можно заключить, что в сопоставимых условиях вид поля рассеяния при одних и тех же значениях зависит от типа  $f_1(x)$ , типа  $\Psi(\varepsilon)$  и характеристики масштаба  $\sigma_\varepsilon$ , а также от того, накладываются ли ограничения на области определения  $\Omega_X$  и  $\Omega_Y$ . При проведении эксперимента для обеспечения независимости выборочных данных при генерации  $x_i^*$  использовался алгоритм «с перемешиванием», укрупненная схема которого представлена на рис. 3. В ходе исследований число повторений эксперимента составляло 1000 раз.



**Рис. 1.** Изменения формы поля рассеяния:

- a* – при показательном законе распределения аргумента  $F_1(x)$  и равномерном законе распределения ошибки  $\Psi(\varepsilon)$ ;  
*б* – при показательном законе распределения аргумента  $F_1(x)$  и нормальном законе распределения ошибки  $\Psi(\varepsilon)$ ;  
*в* – при равномерном законе распределения аргумента  $F_1(x)$  и равномерном законе распределения ошибки  $\Psi(\varepsilon)$ ;  
*г* – при равномерном законе распределения аргумента  $F_1(x)$  и нормальном законе распределения ошибки  $\Psi(\varepsilon)$ ;  
*д* – при нормальном законе распределения аргумента  $F_1(x)$  и равномерном законе распределения ошибки  $\Psi(\varepsilon)$ ;  
*е* – при нормальном законе распределения аргумента  $F_1(x)$  и нормальном законе распределения ошибки  $\Psi(\varepsilon)$



**Рис. 2.** Изменения формы поля рассеяния при отсутствии ограничений на области распределения  $\Omega_x$  и  $\Omega_y$ :

- $a$  – для показательного закона распределения аргумента  $F_1(x)$  и равномерного закона распределения ошибки  $\Psi(\epsilon)$ ;
- $б$  – для показательного закона распределения аргумента  $F_1(x)$  и нормального закона распределения ошибки  $\Psi(\epsilon)$ ;
- $в$  – для равномерного закона распределения аргумента  $F_1(x)$  и равномерного закона распределения ошибки  $\Psi(\epsilon)$ ;
- $з$  – для равномерного закона распределения аргумента  $F_1(x)$  и нормального закона распределения ошибки  $\Psi(\epsilon)$ ;
- $д$  – для нормального закона распределения аргумента  $F_1(x)$  и равномерного закона распределения ошибки  $\Psi(\epsilon)$ ;
- $е$  – для нормального закона распределения аргумента  $F_1(x)$  и нормального закона распределения ошибки  $\Psi(\epsilon)$



Рис. 3. Алгоритм генерации случайных чисел «с перемешиванием»

### ОЦЕНИВАНИЕ ПАРАМЕТРОВ ЛИНЕЙНОЙ ЗАВИСИМОСТИ ПРИ ОТСУТСТВИИ КОРРЕЛЯЦИОННОЙ ТАБЛИЦЫ ЭМПИРИЧЕСКИХ ДАННЫХ

При решении практических задач приходится сталкиваться с ситуацией, когда объемы и погрешности измерений  $X$  и  $Y$  различны. Это, например, возможно, когда за получение измерительных данных отвечают разные организации, действующие в соответствии со своими правилами. В качестве примера упомянем [3], где описываются проблемы, связанные со сбором данных по результатам подконтрольной эксплуатации.

Будем считать, что зависимая  $y$  и независимая  $x$  величины связаны соотношением:

$$y = bx. \quad (3)$$

Это допущение не снижает общности дальнейших рассуждений, т. к. от зависимости вида  $y = a + bx$  можно перейти к зависимости  $\hat{y} = bx$ , где  $\hat{y} = y - a$ .

Известно, что математическое ожидание непрерывной случайной величины определяется на основе соотношения:

$$M[x] = \int_{\Omega_X} xf(x)dx, \quad (4)$$

где  $\Omega_X$  – область определения случайной величины. Преобразуем его к виду

$$M[x] = \sum_{i=1}^N \int_{\Delta_i} xf(x)dx. \quad (5)$$

Будем подбирать  $\Delta_i$  таким образом, чтобы

$$\int_{\Delta_i} xf(x)dx = p_i = \frac{1}{N}, \quad (i = \overline{1; N}), \quad (6)$$

где  $N$  – наперед заданное число.

Если в качестве представителя  $\Delta_i$  взять любую точку внутри  $\Delta_i$ , т. е.  $x_i \in \Delta_i$ , (6) преобразуется к известному соотношению для оценивания центра группирования непрерывной случайной величины по выборочным данным:

$$M[x] = \frac{\sum_{i=1}^N x_i}{N}. \quad (7)$$

Рассуждая аналогичным образом, можно установить, что при  $y = bx$  справедливы соотношения:

$$M[\hat{y}] = bM[x]; \quad (8)$$

$$\sigma_y = b\sigma_x. \quad (9)$$

Здесь  $\sigma_x$ ,  $\sigma_y$  – характеристики масштаба в форме среднеквадратического отклонения непрерывных случайных величин  $X$  и  $Y$ .

Установим теперь значение коэффициента корреляции  $r_{x\hat{y}}$  в случае, когда между зависимой и независимой величинами существует линейная связь.

Согласно [14] расчет коэффициента корреляции осуществляется на основе соотношения:

$$r_{x\hat{y}} = \frac{\text{cov}(x\hat{y})}{\sigma_x\sigma_y} = M \left[ \frac{x-M[x]}{\sigma_x} \cdot \frac{\hat{y}-M[\hat{y}]}{\sigma_y} \right].$$

В случае  $y = bx$

$$\begin{aligned} r_{x\hat{y}} &= M \left[ \frac{x-M[x]}{\sigma_x} \cdot \frac{\hat{y}-M[\hat{y}]}{\sigma_y} \right] = \\ &= M \left[ \frac{(x-M[x])^2}{\sigma_x^2} \right] = 1. \end{aligned} \quad (10)$$

Вышеприведенные выкладки имели целью подчеркнуть то обстоятельство, что коэффициент корреляции является, на наш взгляд, характеристикой взаимосвязи не только в случае нормальных законов распределения случайных величин  $N(x, M[x], \sigma_x)$ ;  $N(y, M[y], \sigma_y)$  (что подчеркивается, например, в [2], [15], [16]), но и в случае иных непрерывных законов распределения случайных величин в случае, когда  $\hat{y} = bx$ .

Учитывая, что  $M[\hat{y}]$  и  $M[y]$  связаны соотношением:

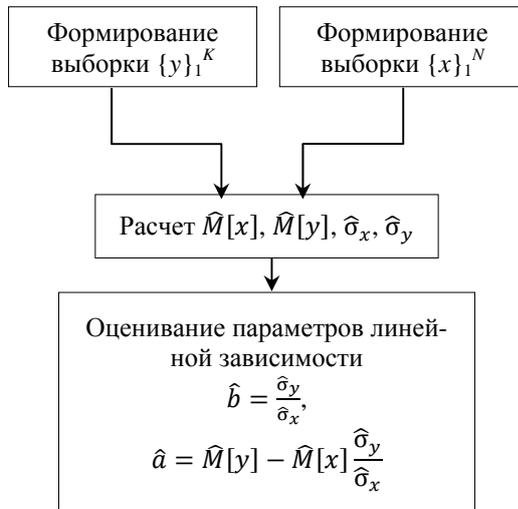
$$M[\hat{y}] = M[y] - a,$$

параметр  $a$  зависимости  $y = a + bx$  может быть определен из соотношения:

$$a = M[y] - M[x] \frac{\sigma_y}{\sigma_x}. \quad (11)$$

Соотношения (10) и (11) позволяют разработать простую вычислительную процедуру оцени-

вания параметров линейной регрессионной зависимости (рис. 4) в случае, когда  $f_1(x)$  и  $f_2(x)$  представляют собою однотипные плотности распределения непрерывных случайных величин. Отметим, что эти соотношения справедливы и в случае, когда объемы выборочных данных различны и, более того, выборки  $\{x\}_1^N$  и  $\{y\}_1^K$  формируются независимо друг от друга (при этом предполагается, что из содержательного анализа известно наличие линейной зависимости  $y = a + bx$ . Например, известно, что механизмы деградации свойств изделий в штатном и форсированном режимах испытаний одни и те же, однако, скорости процессов деградации различны).



**Рис. 4.** Схема оценивания параметров линейной регрессионной зависимости по выборочным данным

**ИССЛЕДОВАНИЕ КАЧЕСТВА ОЦЕНИВАНИЯ ПАРАМЕТРОВ ЛИНЕЙНОЙ ЗАВИСИМОСТИ ПО ВЫБОРОЧНЫМ ДАННЫМ ПРИ РАЗНЫХ СПОСОБАХ ФОРМИРОВАНИЯ ВЫБОРОК**

На рис. 5 приведена укрупненная схема алгоритма формирования независимых выборок  $\{x\}_1^N$  и  $\{y\}_1^K$  разного объема.

В качестве характеристик качества оценивания параметра модели использовались метрики: средняя погрешность оценивания:

$$M_{\hat{b}} = \frac{\sum_{j=1}^{\eta} (\hat{b}^{(j)} - b)}{\eta}; \quad (12)$$

средняя абсолютная погрешность оценивания:

$$M_{\hat{b}}^{(A)} = \frac{\sum_{j=1}^{\eta} |\hat{b}^{(j)} - b|}{\eta}; \quad (13)$$

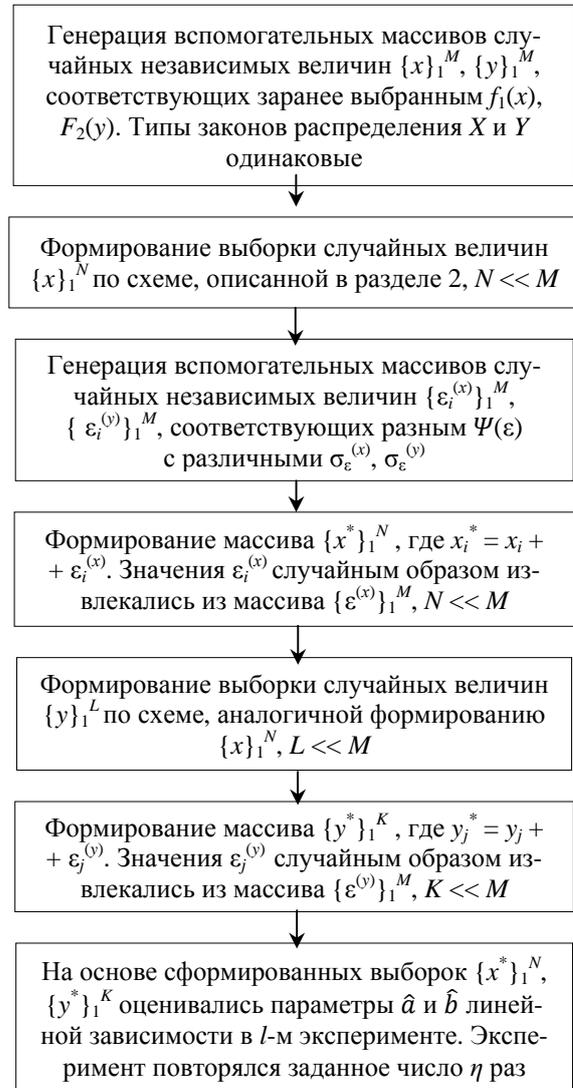
максимальная абсолютная погрешность оценивания:

$$\Delta_{max} = \max_{j \in \{1, L\}} \{|\hat{b}^{(j)} - b|\}; \quad (14)$$

устойчивость оценки:

$$\sigma_{\hat{b}} = \sqrt{\frac{\sum_{j=1}^{\eta} (\hat{b}^{(j)} - b)^2}{\eta}}. \quad (15)$$

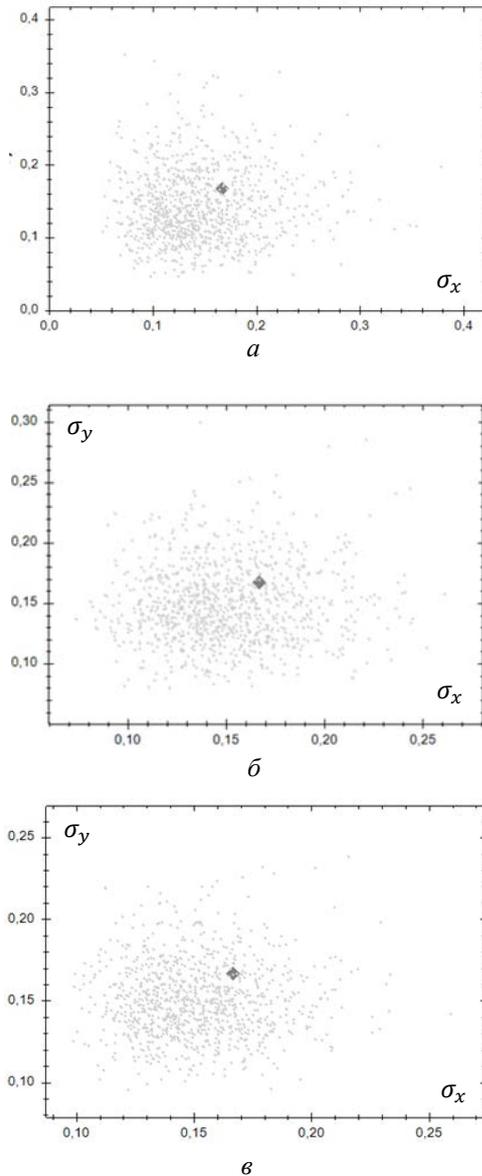
Здесь  $\eta$  – число повторений вычислительного эксперимента. В ходе исследований  $\eta$  принималось равным 1000.



**Рис. 5.** Схема формирования независимых выборок разного объема  $\{x^*\}_1^N, \{y^*\}_1^K$

На рис. 6 в качестве примера приведены поля рассеяния значений среднеквадратических отклонений  $\sigma_x, \sigma_y$ , рассчитанные по выборкам разного объема при показательном законе распределения  $f_1(x)$ , равномерном распределении ошибки  $\Psi(\epsilon)$ ,  $\sigma_{\epsilon} = 0.1\sigma_x$ . Рис. 6, а соответствует случаю  $N = K = 20$ ; рис. 6, б –  $N = K = 50$ ; рис. 6, в – случаю, когда  $N = K = 100$ . На рисунках выделена точка, соответствующая значениям среднеквадратических отклонений при отсутствии ошибок в регистрации выборочных

данных  $\sigma_x = \sigma_y = 1/6$ . В ходе исследований  $\eta$  принималось равным 1000.



**Рис. 6.** Поля рассеяния значений  $\sigma_x, \sigma_y$  при показательном законе распределения  $f_1(x)$  и равномерном распределении ошибки  $\Psi(\varepsilon)$

На рис. 7 в качестве примера приведены статистические характеристики качества оценивания параметров линейной зависимости при независимо формируемых выборках  $\{x\}_1^N$  и  $\{y\}_1^L$  в случае показательного распределения аргумента  $f_1(x)$ , равномерного распределения ошибки  $\Psi(\varepsilon)$ . На рис. 7 пунктирные линии соответствуют равным значениям характеристик качества оценивания параметров модели.

Полученные результаты позволяют заключить, что в рамках предлагаемого метода возможна компенсация недостатка данных за счет

повышения устойчивости измерений. Рис. 7, а демонстрирует зависимость характеристик качества, определяемых соотношениями (12)–(15), от объемов выборок и погрешности регистрации исходных данных. На рис. 7, б и в показаны плотности распределения  $\varphi(\hat{a}), \varphi(\hat{b})$  параметров  $\hat{a}$  и  $\hat{b}$  соответственно при параметре масштаба ошибки  $\sigma_\varepsilon = 0.9\sigma_x$ . На рис. 7 цифра 1 соответствует случаю, когда объем выборки  $\{x\}_1^N$  составлял  $N = 100$ ; выборки  $\{y\}_1^K$  составлял  $K = 10$ . Цифра 2 соответствует случаю, когда объем выборки  $\{x\}_1^N$  составлял  $N = 100$ ; объем выборки  $\{y\}_1^K$  составлял  $K = 20$ . Цифра 3 – случай, когда объем выборки  $\{x\}_1^N$  составлял  $N = 100$ ; выборки  $\{y\}_1^K$  составлял  $K = 50$ . Исследования, выполненные на основе методик, описанных в [15], по критериям  $\omega^2, \chi^2, F$ -Колмогорова позволяют заключить, что  $\varphi(\hat{a}), \varphi(\hat{b})$  можно аппроксимировать нормальными законами распределения. В ходе эксперимента теоретические значения  $a$  и  $b$  принимали значения «0» и «1» соответственно.

#### СРАВНЕНИЕ МЕТОДОВ ОЦЕНИВАНИЯ ПАРАМЕТРОВ ЛИНЕЙНОЙ РЕГРЕССИОННОЙ ЗАВИСИМОСТИ

Приведены результаты сравнительного анализа подходов к оцениванию параметров одномерной регрессионной зависимости на основе метода наименьших квадратов (МНК) и метода, основанного на решении задачи, обратной по отношению к анализу закона распределения функции случайного аргумента (ФСА):

$$A^{(\text{ФСА})}: f_1(x), F_2(y) \rightarrow \{a, b\}. \quad (16)$$

Подробно постановка обратной задачи для случая строгих функциональных зависимостей  $y = \varphi(x)$  приводится в [13].

В основе распределения параметров линейной регрессионной зависимости  $\{a, b\}$  методом наименьших квадратов лежит преобразование вида

$$A^{(\text{МНК})}: \{x, y\}_1^N \rightarrow \{a, b\}, \quad (17)$$

где  $\{x, y\}_1^N$  – корреляционная таблица наблюдаемых значений независимой  $X$  и зависимой  $Y$  случайных величин. Предполагается, что элементы множества удовлетворяют требованиям области адекватности МНК [10].

В случае одномерной линейной регрессионной зависимости  $A^{(\text{МНК})}$  сводится к решению системы линейных алгебраических уравнений [2]:

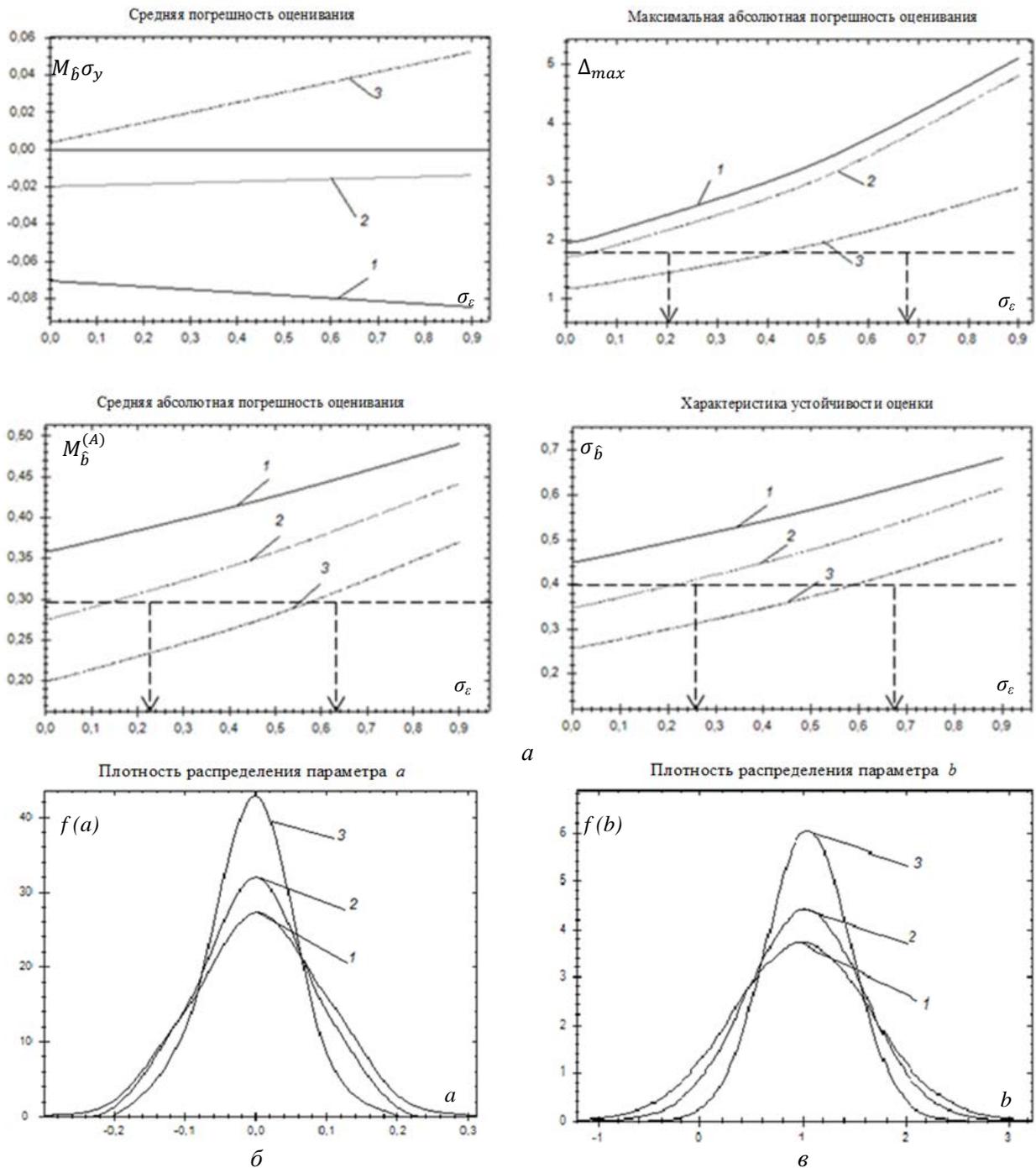


Рис. 7. Характеристики качества оценивания параметров линейной зависимости при независимом формировании выборок  $\{x\}_1^N, \{y\}_1^K$

$$\begin{cases} N \cdot a + b \sum_{i=1}^N x_i = \sum_{i=1}^N y_i; \\ a \sum_{i=1}^N x_i + b \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i, \end{cases} \quad (18)$$

откуда для  $a$ :

$$a = \frac{\sum_{i=1}^N y_i}{N} - b \frac{\sum_{i=1}^N x_i}{N} = M[y] - bM[x], \quad (19)$$

где  $M[x], M[y]$  – оценки первых начальных моментов  $X$  и  $Y$  соответственно:

$$M[x] = \frac{\sum_{i=1}^N x_i}{N}; \quad M[y] = \frac{\sum_{i=1}^N y_i}{N}; \quad (20)$$

для  $b$ :

$$b = \frac{\frac{\sum_{i=1}^N x_i y_i}{N} - \frac{\sum_{i=1}^N x_i}{N} \cdot \frac{\sum_{i=1}^N y_i}{N}}{\frac{\sum_{i=1}^N x_i^2}{N} - \left(\frac{\sum_{i=1}^N x_i}{N}\right)^2} = \frac{\frac{\sum_{i=1}^N x_i y_i}{N} - M[x]M[y]}{D_x}, \quad (21)$$

где  $D_x$  – оценка дисперсии.

Действительно [2]:

$$D_x = \frac{\sum_{i=1}^N (x_i - M[x])^2}{N}$$

$$\begin{aligned}
&= \frac{\sum_{i=1}^N x_i^2}{N} - 2M[x] \cdot \frac{\sum_{i=1}^N x_i}{N} + (M[x])^2 = \\
&= \frac{\sum_{i=1}^N x_i^2}{N} - 2(M[x])^2 + (M[x])^2 = \\
&= \frac{\sum_{i=1}^N x_i^2}{N} - (M[x])^2. \quad (22)
\end{aligned}$$

Определение параметров линейной модели на основе анализа закона распределения функции случайного аргумента основано на использовании преобразований вида:

$$A^{(\text{ФСА})}: \{x\}_1^N, \{y\}_1^N \rightarrow \{a, b\}, \quad (23)$$

где  $\{x\}_1^N, \{y\}_1^N$  – независимо наблюдаемые выборочные значения величин  $X$  и  $Y$  одного и того же объема  $N$ .

Как было показано ранее (11), параметр  $a$  связан зависимостью того же вида, что и (19). Параметр  $b$  определяется соотношением (9), а именно:

$$b = \frac{\sigma_y}{\sigma_x}.$$

Преобразуем это соотношение очевидным образом:

$$b = \frac{\sigma_y}{\sigma_x} = 1 \cdot \frac{\sigma_y}{\sigma_x}. \quad (24)$$

Теперь, с учетом того, что для случайных величин, связанных линейной зависимостью, коэффициент корреляции равен единице (10), преобразуем (24) к виду:

$$\begin{aligned}
b &= 1 \cdot \frac{\sigma_y}{\sigma_x} = \\
&= \frac{\sum_{i=1}^N (x_i - M[x])(y_i - M[y])}{N \cdot \sigma_x \cdot \sigma_y} \cdot \frac{\sigma_y}{\sigma_x} \quad (25)
\end{aligned}$$

с учетом известного правила расчета коэффициента корреляции [2].

Из (24) следует, что:

$$\begin{aligned}
b &= \frac{\frac{\sum_{i=1}^N x_i y_i}{N} - 2M[x] \cdot M[y] + M[x] \cdot M[y]}{\sigma_x^2} = \\
&= \frac{\frac{\sum_{i=1}^N x_i y_i}{N} - M[x] \cdot M[y]}{D_x}. \quad (26)
\end{aligned}$$

Выражение (26) совпадает с (21), из чего можно заключить, что оценки параметров линейной модели, получаемые в результате решения обратной задачи определения закона распределения функции случайного аргумента, совпадают с оценками МНК. Учитывая то, что МНК позволяет решать задачу лишь при наличии совместно наблюдаемых значений  $\{x, y\}_1^N$ , а соотношения (9) и (11) позволяют решать ту же задачу, во-первых, при независимых наблюдениях  $X$  и  $Y$ ; во-вторых, при разных объемах выборок

$\{x\}_1^N, \{y\}_1^L$ , можно заключить, что в случае оценивания параметров линейной регрессионной зависимости МНК является частным случаем решения задачи, обратной по отношению к известной задаче определения закона распределения функции случайного аргумента.

### ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК НАДЕЖНОСТИ ПО РЕЗУЛЬТАТАМ ФОРСИРОВАННЫХ ИСПЫТАНИЙ

В качестве примера приложения полученных результатов рассмотрим задачу оценивания характеристик надежности технических изделий по результатам форсированных испытаний [16]. Теоретической базой пересчета данных об отказах в режиме форсированных испытаний на нормальные условия испытаний является так называемая «базовая закономерность». В основе построения базовой закономерности лежит следующее положение [16]: каждому моменту времени  $t_j$  отказа изделия в нормальном режиме  $\Gamma_0$  соответствует единственное значение времени  $\tau_j$  отказа тождественного изделия в форсированном режиме  $\Gamma_1$ . Иными словами, моменты времени отказов изделий в нормальном  $\Gamma_0$  и форсированном  $\Gamma_1$  режимах связаны функциональной зависимостью

$$t(\vec{v}, \Gamma_0) = f(\tau(\vec{v}, \Gamma_1)), \quad (27)$$

где  $\vec{v} = \{v_1, \dots, v_n\}$  – вектор внутренних параметров изделия, определяющих технические характеристики объекта в нормальном

$$X(t, \Gamma_0) = \Psi_1[v_1(t, \Gamma_0), \dots, v_n(t, \Gamma_0)]$$

и в форсированном

$$X(\tau, \Gamma_1) = \Psi_2[v_1(\tau, \Gamma_1), \dots, v_n(\tau, \Gamma_1)]$$

режимах. Как отмечается в [16], (27) есть формальная запись условия независимости расходования ресурса от предыстории. Изделие, отказавшее за время  $t$  в режиме  $\Gamma_0$ , откажет в режиме  $\Gamma_1$  за время  $\tau$ , определяемое из условия равных вероятностей:

$$\frac{q(t, \Gamma_0)}{q(\tau, \Gamma_1)} = 1, \quad (28)$$

где  $q(t, \Gamma_0), q(\tau, \Gamma_1)$  – вероятности работы до отказа в нормальном и форсированном режимах соответственно.

В контексте статьи (28) есть запись обратной задачи определения закона распределения функции случайного аргумента [17].

Учитывая, что тип закона распределения наработки до отказа является косвенной характеристикой механизма деградации изделия [18], можно утверждать, что если в нормальном и

форсированном режиме механизмы деградации одни и те же (но в форсированном режиме выше скорость деградации), то типы законов распределения  $F_1^{(\Gamma_0)}(t) = F_2^{(\Gamma_1)}(\tau)$  являются одинаковыми. Иными словами, зависимость (27) приобретает вид

$$t = k\tau.$$

Такая формула представления линейной зависимости обосновывается тем, что

$$F_1^{(\Gamma_0)}(0) = F_2^{(\Gamma_1)}(0) = 0.$$

Коэффициент ускорения  $k$ , согласно [18], определяется из соотношения

$$k = \frac{\sigma_\tau^{(\Gamma_1)}}{\sigma_t^{(\Gamma_0)}}. \quad (29)$$

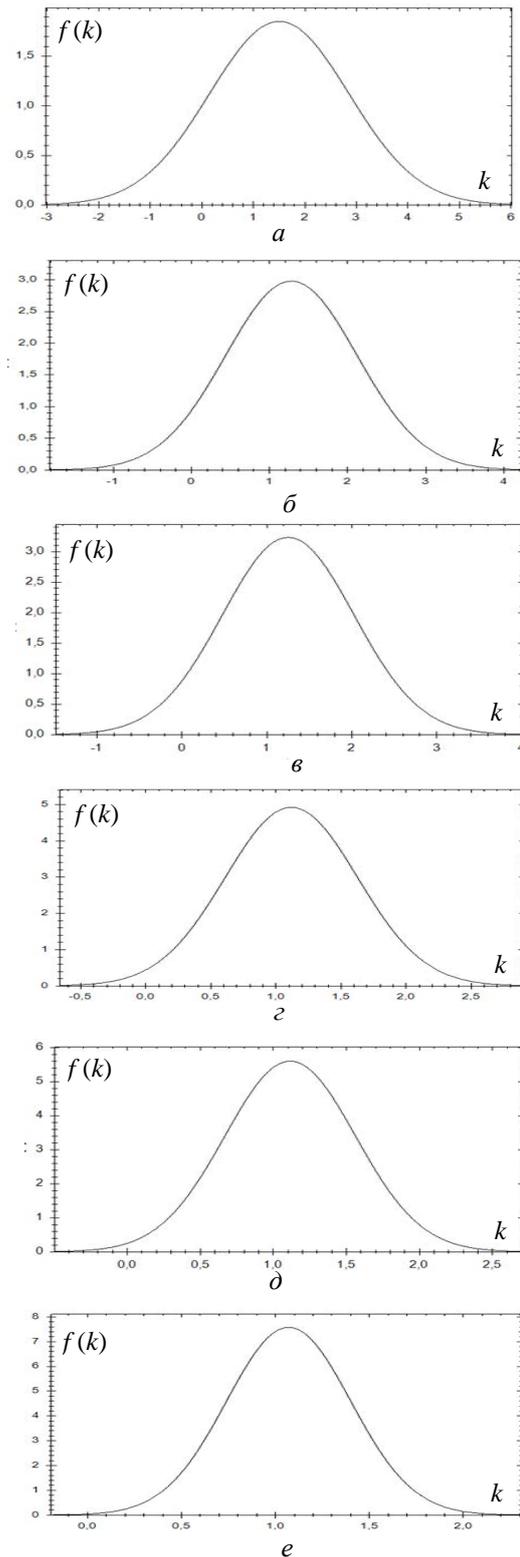
Заметим, что в случае показательных законов распределения  $F_1^{(\Gamma_0)}(t)$ ,  $F_2^{(\Gamma_1)}(\tau)$ , (29) преобразуется к известному виду [16]:

$$k = \frac{T_\tau^{(\Gamma_1)}}{T_t^{(\Gamma_0)}}, \quad (30)$$

где  $\{T_\tau^{(\Gamma_1)}, T_t^{(\Gamma_0)}\}$  – значения математических ожиданий наработки до отказа в режимах  $\Gamma_1, \Gamma_0$  соответственно. Корректность преобразования (29) к виду (30) обосновывается тем, что для показательного закона распределения  $T_\tau^{(\Gamma_1)} = \sigma_\tau^{(\Gamma_1)}$ ;  $T_t^{(\Gamma_0)} = \sigma_t^{(\Gamma_0)}$ .

Вместе с тем в [2] подчеркивается, что погрешность оценивания значений моментов по выборке конечного объема увеличивается с ростом порядка момента. Кроме того, необходимо учитывать то обстоятельство, что при проведении испытаний в режимах  $\Gamma_0$  и  $\Gamma_1$  приходится иметь дело с выборками конечного объема (и, как правило, объем этот мал), следует пользоваться не точечной, а интервальной оценкой коэффициента  $k$ . Иными словами, значение коэффициента ускорения  $k$  является величиной случайной.

На рис. 8 приведены оценки законов распределения коэффициента ускорения, построенные по результатам обработки одних и тех же выборок с использованием соотношений (29) и (30).



**Рис. 8.** Плотность распределения коэффициента ускорения при разных способах обработки выборок:  
 а – на основе (29) при  $N = 5, K = 10$ ;  
 б – на основе (30) при  $N = 5, K = 10$ ;  
 в – на основе (29) при  $N = 10, K = 15$ ;  
 г – на основе (30) при  $N = 10, K = 15$ ;  
 д – на основе (29) при  $N = 20, K = 30$ ;  
 е – на основе (30) при  $N = 20, K = 30$

В ходе испытаний число повторений эксперимента бралось равным 1000. Законы распределения погрешностей  $\Psi(\varepsilon^{(x)})$ ,  $\Psi(\varepsilon^{(y)})$  – равномерные; масштабы помех  $\sigma_{\varepsilon^{(x)}} = 0.5\sigma_x$ ;  $\sigma_{\varepsilon^{(y)}} = 0.5\sigma_y$ . Законы распределения  $f_1(x) = 6e^{-6x}$ ;  $f_2(y) = 6e^{-6y}$ .

Из полученных результатов можно заключить, что при показательных законах распределения  $F_1^{(\Gamma_0)}(t)$ ,  $F_2^{(\Gamma_1)}(\tau)$  использование соотношения (30) является более предпочтительным. Это обусловлено тем, что при ограниченном числе данных погрешность оценивания значений моментов законов распределения непрерывных случайных величин тем больше, чем выше порядок момента [2].

### ЗАКЛЮЧЕНИЕ

Предложен метод построения линейной регрессионной зависимости, основанный на решении обратной задачи построения закона распределения функции случайного аргумента. Предлагаемый метод отличается от известных тем, что позволяет строить линейные регрессионные зависимости при отсутствии корреляционной таблицы наблюдаемых значений независимой и зависимой случайных величин, в том числе при разных объемах выборочных данных. Показано, что результаты, получаемые посредством предлагаемого метода, совпадают с результатами, получаемыми известным методом наименьших квадратов в случае наличия корреляционной таблицы наблюдаемых значений независимой и зависимой случайных величин. Установлено, что влияние учета границ области возможных значений тем больше, чем более асимметричным является закон распределения случайного аргумента, а также чем больше масштаб ошибки измерений. Получены количественные значения оценок параметров линейной регрессионной модели в зависимости от объемов исходных данных, погрешности регистрации данных, закона распределения функции случайного аргумента. Показано, что при заданном значении характеристик качества оценивания параметров регрессионной модели низкую погрешность регистрации данных можно компенсировать увеличением объема выборочных данных. И, наоборот, при ограниченных возможностях получения выборочных данных требуемое качество оценивания параметров регрессионной модели можно обеспечить за счет уменьшения ошибки в регистрации выборочных данных.

### СПИСОК ЛИТЕРАТУРЫ

1. **ГОСТ Р ИСО 9000-2008.** Системы менеджмента качества. Основные положения и словарь. М.: Стандартинформ, 2009. 35 с. [[ Quality management systems. Fundamentals and vocabulary, (in Russian), Federal standard R ISO 9000-2008, Moscow, Standatrinform, 2009. ]]
2. **Вентцель Е. С.** Теория вероятностей: учебник для студ. вузов, 9-е изд., стер. М.: Академия, 2003. 576 с. [[ E. S. Wentzel, *Probability theory*, (in Russian). Moscow: Academiya, 2003. ]]
3. **Дружинин Г. В.** Надежность автоматизированных систем. М.: Энергия, 1977. 536 с. [[ G. V. Druzhinin, *Reliability of automated systems*, (in Russian). Moscow: Energiya, 1977. ]]
4. **Кузнецов О. П.** Когнитивное моделирование слабо структурированных ситуаций // Поспеловские чтения. Искусственный интеллект – проблемы и перспективы. 2006. № 7. С. 86–100. [[ O. P. Kuznecov, "Cognitive modeling of poorly structured situations," (in Russian), in *Pospelovskiy chteniya*, no. 7, pp. 86-100, 2006. ]]
5. **Pelaez C., Bowles J.** Using fuzzy cognitive maps as a system model for failure models and effect analysis // Information Sciences. 1996. Vol. 88. P. 177–199. [[ C. Pelaez, J. Bowles, "Using fuzzy cognitive maps as a system model for failure models and effect analysis," in *Information Sciences*, 1996. vol. 88, pp. 177-199. ]]
6. **Мирский Г. Л.** Аппаратное определение характеристик случайных процессов. М.: Энергия, 1972. 456 с. [[ G. L. Mirskiy, *Hardware characterization of random processes*, (in Russian). Moscow: Energiya, 1972. ]]
7. **Кульба В. В., Миронов П. Б., Назаретов В. М.** Анализ устойчивости социально-экономических систем с использованием знаковых графов // Автоматика и телемеханика. 1993. № 7. С. 130–137. [[ V. V. Kul'ba, P. B. Mironov, V. M. Nazaretov, "Stability analysis of socio-economic systems using iconic graphs," in *Avtomatica i telemehanika*, 1993, pp. 130-137. ]]
8. **Адлер Ю. П., Маркова Е. В., Грановский Ю. В.** Планирование эксперимента при поиске оптимальных условий. М.: Наука, 1976. 280 с. [[ U. P. Adler, E. V. Markova, U. V. Granovsky, *Planning experiment in finding the optimal conditions*, (in Russian). Moscow: Nauka, 1976. ]]
9. **Мудров В. И., Кушко В. Л.** Метод наименьших модулей. М.: Знание, 1971. 64 с. [[ V. I. Mudrov, V. L. Kushko, *Least method modules*, (in Russian). Moscow: Znaniye, 1971. ]]
10. **Демиденко Е. З.** Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981. 304 с. [[ E. Z. Demidenko, *Linear and nonlinear regression*, (in Russian). Moscow: Finansi i statistika, 1981. ]]
11. **Пугачев В. С.** Теория вероятностей и математическая статистика. М.: Физматлит, 2002. 496 с. [[ V. S. Pugachev, *Probability theory and mathematical statistics*, (in Russian). Moscow: Fizmatlit, 2002. ]]
12. **Пугачев В. С., Казаков И. Е, Евланов Л. Г.** Основы статической теории автоматических систем. М.: Машиностроение, 1974. 400 с. [[ V. S. Pugachev, I. E. Kazakov, L. G. Evlanov, *The static theory fundamentals of automatic systems*, (in Russian). Moscow: Mashinostroeniye, 1974. ]]
13. **Гвоздев В. Е., Колоденкова А. Е.** Непараметрическое оценивание функциональных зависимостей по эмпирическим данным // Мехатроника, автоматизация, управление. 2005. № 8. С. 12–18. [[ V. E. Gvozdev, A. E. Kolodenkova, "Nonparametric estimation of functional

dependencies using experimental data," (in Russian), in *Mehatronika, avtomatizaciya, upravleniye*, 2005. № 8. P. 12–18. ]]

14. **Корн Г. А., Корн Т. М.** Справочник по математике для научных работников инженеров. М.: Наука, 1974. 832 с. [[ G. A. Korn, T. M. Korn, *Mathematical handbook for scientists engineers*, (in Russian). Moscow: Nauka, 1974. ]]

15. **Смирнов Н. В., Дунин-Барковский И. В.** Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1989. 512 с. [[ N. V. Smirnov, I. V. Dunin-Barkovsky, *Course in the theory of probability and mathematical statistics for technical applications*, (in Russian). Moscow: Nauka, 1989. ]]

16. **Перротте А. И., Карташов Г. Д., Цветаев К. Н.** Основы ускоренных испытаний радиоэлементов на надежность. М.: Советское радио, 1968. 224 с. [[ A. I. Perrote, G. D. Kartashov, K. N. Cvetaev, *Fundamentals of accelerated tests on the radioactive elements reliability*, (in Russian). Moscow: Sovetskoye radio, 1968. ]]

17. **Гузаиров М. Б., Гвоздев В. Е., Ильясов Б. Г., Колоденкова А. Е.** Статистическое исследование территориальных систем. М.: Машиностроение, 2008. 186 с. [[ M. B. Guzaïrov, V. E. Gvozdev, B. G. Il'yasov, A. E. Kolodenkova, *Statistical study of territorial systems*, (in Russian). Moscow: Mashinostroeniye, 2008. ]]

18. **Комаров Ю. Л.** Надежность радиоэлектронной аппаратуры [Электронный ресурс]. URL: <http://www.universalinternetlibrary.ru/book/36302/ogl.shtml> (дата обращения: 03.05.2014). [[ U. L. Komarov. (2014, May 03). *Reliability of electronic equipment* [Online], (in Russian). Available: <http://www.universalinternetlibrary.ru/book/36302/ogl.shtml> ]]

#### ОБ АВТОРАХ

**ГВОЗДЕВ Владимир Ефимович**, зав. каф. автоматиз. проектирования инф. систем. Дипл. инж. эл. техн. (УАИ, 1978). Д-р техн. наук по АСУ (УГАТУ, 2000). Иссл. в обл. геоинформ. и мат. моделирования состояния территориальных систем, упр. прогр. проектами, анализа надежности техн. объектов.

**СУБХАНГУЛОВА Алия Салаватовна**, м-т каф. автоматиз. проектирования инф. систем. Б-р техн. и технол. по информатике и выч. техн. (УГАТУ, 2013).

**БЕЖАЕВА Оксана Яковлевна**, доц. каф. автоматиз. проектирования инф. систем. Дипл. экон. по инф. системам в экономике (УГАТУ, 2000). Канд. техн. наук (УГАТУ, 2004). Иссл. в обл. упр. в чрезв. ситуациях, мат. моделир. и упр. прогр. проектами.

#### METADATA

**Title:** Linear correlations estimation of technical objects parameters without correlation table of empirical data.

**Authors:** V. E. Gvozdev<sup>1</sup>, A. S. Subhangulova<sup>2</sup>, O. Y. Bezhaeva<sup>3</sup>

**Affiliation:**

Ufa State Aviation Technical University (UGATU), Russia.

**Email:** <sup>1</sup>wega55@mail.ru, <sup>2</sup>aliyasr21@gmail.com,

<sup>3</sup>obezhaeva@gmail.ru

**Language:** Russian.

**Source:** Vestnik UGATU, vol. 19, no. 4 (70), pp. 106-117, 2015. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

**Abstract:** The problem of constructing a linear one-dimensional functional dependence at different argument distribution laws, different initial data volumes, and various registration accuracy of measured parameters is discussed. An approach to the parameters estimation in the absence of a linear dependence of the observed values of the table together is described. The proposed approach relation with the method of least squares is established.

**Key words:** occasional value; occasional value distribution laws; linear regression dependences; correlation coefficient.

**About authors:**

**GVOZDEV, Vladimir Efimovich**, Prof., Dept. of Computer Science and Robotics. Dipl. Engineer in electronics technology (UGATU, 1978), Dr. of Tech. Sci. (UGATU, 2000).

**SUBHANGULOVA, Aliya Salavatovna**, Undergraduate. Student, Dept. of Computer Science and Robotics. Bachelor of Technics & Technology (UGATU, 2013).

**BEZHAEVA, Oksana Yakovlevna**, Docent., Dept. of Computer Science and Robotics. Dipl. Economist (2000). Cand. of Tech. Sci. (UGATU, 2004).