Вычислительные технологии, 2020, том 25, № 3, с. 130–141. © ИВТ СО РАН, 2020 ISSN 1560-7534
Computational Technologies, 2020, vol. 25, no. 3, pp. 130–141. © ICT SB RAS, 2020 eISSN 2313-691X

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

DOI:10.25743/ICT.2020.25.3.014

Синтез оптимальных планов эксперимента в условиях GL-распределения ошибок для моделей с ошибками в объясняющих переменных

В. С. Тимофеев†, Е. А. Хайленко
Новосибирский государственный технический университет, Новосибирск, Россия
†Контактный автор: Тимофеев Владимир С., e-mail: v.timofeev@corp.nstu.ru

Поступила 10 марта 2020 г., доработана 24 марта 2020 г., принята в печать 10 апреля 2020 г.

Рассмотрена задача планирования эксперимента в условиях появления ошибок в объясняющих переменных. Сформулировано и доказано утверждение о способе вычисления элементов информационной матрицы Фишера с использованием обобщенного лямбда-распределения, доказано следствие о способе вычисления функции эффективности плана эксперимента. Сравнение результатов вычисления функции эффективности с использованием выведенного в следствии соотношения и с помощью известного соотношения для нормального распределения ошибок показало, что результаты совпадают. Построены оптимальные планы эксперимента для различных распределений случайных компонент.

Ключевые слова: регрессионные зависимости, модель с ошибками в объясняющих переменных, обобщенное лямбда-распределение, план эксперимента, информационная матрица Фишера.

Цитирование: Тимофеев В.С., Хайленко Е.А. Синтез оптимальных планов эксперимента в условиях GL-распределения ошибок для моделей с ошибками в объясняющих переменных. Вычислительные технологии. 2020; 25(3):130–141.

Введение

При современном уровне развития науки и техники многие исследования требуют постановки сложных и дорогостоящих экспериментов. При проведении экспериментов исследователь пытается извлечь наибольший объем информации об изучаемых процессах при наименьших затратах, поэтому возникает задача построения оптимальных планов эксперимента.

Классические алгоритмы построения оптимальных планов эксперимента [1] основаны на предположении о нормальности распределения ошибок наблюдений и позволяют учитывать лишь неоднородность дисперсий на области планирования. Однако на практике это предположение может не выполняться, также на области планирования может иметь место неоднородность распределения ошибок. В работе [2] предложен подход к построению оптимальных планов с использованием обобщенного лямбда-распределения (GL-распределения), которое описывает целый класс распределений, таких как нормальное, экспоненциальное, Вейбулла, гамма-, бета- и др. [3, 4]. Такой подход позволяет учитывать неоднородность как дисперсий, так и формы распределения ошибок наблюдений на всей области планирования. Следует заметить, что перечисленные выше

способы синтеза оптимальных планов эксперимента подходят лишь для классических регрессионных моделей и не учитывают наличия ошибок в объясняющих переменных. В работе [5] предложен способ вычисления информационной матрицы Фишера для моделей с ошибками в объясняющих переменных [6, 7] в случае нормального распределения случайных компонент, но, как отмечалось выше, на практике нормальность распределения ошибок может не выполняться. Таким образом, возникает задача вычисления элементов информационной матрицы при GL-распределении случайных ошибок, решение которой позволит предложить обобщенный алгоритм построения оптимальных планов эксперимента в условиях наличия ошибок в объясняющих переменных.

1. Постановка задачи

Будем рассматривать линейную по параметрам модель

$$y = X\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{1}$$

где $X = \begin{pmatrix} h_1(x_{11}) & h_2(x_{12}) & ... & h_m(x_{1m}) \\ h_1(x_{21}) & h_2(x_{22}) & ... & h_m(x_{2m}) \\ ... & ... & ... & ... \\ h_1(x_{n1}) & h_2(x_{n2}) & ... & h_m(x_{nm}) \end{pmatrix}$ — матрица плана эксперимента, имеющая полный столбцовый ранг; $\mathbf{h}(x) = (h_1(x), h_2(x), ..., h_m(x))^T$ — вектор действительных функций; $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)^T$ — вектор неизвестных параметров; m — число параметров; n — количество наблюдений; $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$ — вектор ошибок в отклике y; $x_{i1}, x_{i2}, ..., x_{im}$, $i = 1, ..., n$, — неизвестные истинные значения входных факторов, наблюдаемые величины которых равны

$$\tilde{x}_{ij} = x_{ij} + \delta_{ij}, \quad i = 1, ..., n, \quad j = 1, ..., m. \tag{2}$$

Для $\boldsymbol{\varepsilon}$ и $\boldsymbol{\delta}_j = (\delta_{1j}, \delta_{2j}, ..., \delta_{nj})^T$, $j = 1, ..., m$, выполняются условия

$$E(\boldsymbol{\varepsilon}) = E(\boldsymbol{\delta}_j) = 0, \quad D(\varepsilon_i) = \sigma_\varepsilon^2, \quad D(\delta_{ij}) = \sigma_{\delta_j}^2,$$
$$cov(\varepsilon_i, \varepsilon_k) = cov(\delta_{ij}, \delta_{kl}) = 0, \quad i \neq k, cov(\varepsilon_i, \delta_{kj}) = 0, \quad i, k = \overline{1, n}, \quad j, l = \overline{1, m}. \tag{3}$$

Следовательно, ковариационная матрица ошибок является диагональной вида

$$\Sigma_{\varepsilon\delta} \begin{pmatrix} \sigma_\varepsilon^2 & 0 & ... & 0 \\ 0 & \sigma_{\delta_1}^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sigma_{\delta_m}^2 \end{pmatrix}.$$

Задача состоит в построении оптимального нормированного плана эксперимента

$$\xi = \begin{pmatrix} x_1 & x_2 & ... & x_s \\ p_1 & p_2 & ... & p_s \end{pmatrix},$$

где $\sum\limits_{i=1}^{s} p_i = 1$, $p_i = \dfrac{n_i}{n}$, s — количество точек в спектре плана и n_i — количество повторных наблюдений в каждой точке спектра плана в условиях наличия ошибок в объясняющих переменных.

2. Вычисление элементов информационной матрицы Фишера при GL-распределении ошибок

Для классических регрессионных моделей, в которых случайные ошибки присутствуют только в отклике, информационная матрица Фишера может быть вычислена с использованием соотношения [1]

$$M(\xi) = \sum_{i=1}^{s} \frac{p_i}{\sigma_\varepsilon^2(x_i)} \mathbf{h}(x_i)\mathbf{h}^T(x_i) = \sum_{i=1}^{s} p_i \lambda(x_i)\mathbf{h}(x_i)\mathbf{h}^T(x_i), \tag{4}$$

где $\lambda(x) = \dfrac{1}{\sigma_\varepsilon^2(x)}$ — функция эффективности, $\sigma_\varepsilon^2(x)$ — дисперсия отклика в точке спектра плана.

Такой способ вычисления элементов информационной матрицы позволяет учитывать лишь нормально распределенные ошибки в отклике. При наличии ошибок в объясняющих переменных в работе [5] предложен следующий способ вычисления элементов информационной матрицы для модели (1), (2) с предположениями (3):

$$M(\xi) = \sum_{i=1}^{s} \frac{p_i}{\sigma_1^2(x_i)} \mathbf{h}(x_i)\mathbf{h}^T(x_i), \tag{5}$$

где обратная величина к функции эффективности $\lambda(x_i)$ вычисляется следующим образом:

$$\sigma_1^2(x_i) = \left(1, \frac{\partial \sum\limits_{i=j}^{m} \theta_j \mathbf{h}(x_i)}{\partial x_i} \right) \Sigma_{\varepsilon\delta} \left(1, \frac{\partial \sum\limits_{j=1}^{m} \theta_j \mathbf{h}(x_i)}{\partial x_i} \right)^T. \tag{6}$$

Данный способ вычисления элементов информационной матрицы применим в случае нормального распределения ошибок в отклике и в объясняющих переменных, что на практике может не выполняться. Поэтому в данной работе для описания распределения случайных компонент предлагается использовать универсальное GL-распределение [3, 4]. Функция GL-распределения зависит от четырех параметров $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ и определяется с точки зрения квантилей распределения [3]. Будем предполагать, что случайные ошибки $\boldsymbol{\varepsilon}$ имеют GL-распределение с параметрами $(\lambda_1^\varepsilon, \lambda_2^\varepsilon, \lambda_3^\varepsilon, \lambda_4^\varepsilon)$, а случайные ошибки $\boldsymbol{\delta}_j$, $j = 1, ..., m$, — с параметрами $(\lambda_1^{\delta_j}, \lambda_2^{\delta_j}, \lambda_3^{\delta_j}, \lambda_4^{\delta_j})$.

Функция распределения i-й ошибки может быть выписана следующим образом [4]:

$$z_i = y_i - \sum_{j=1}^{m} \theta_j h_j(\tilde{x}_{ij} - \delta_j) = Q(u_i, \lambda_1^\varepsilon, \lambda_2^\varepsilon, \lambda_3^\varepsilon, \lambda_4^\varepsilon) =$$

$$= \lambda_1^\varepsilon + \frac{1}{\lambda_2^\varepsilon} \left(\frac{u^{\lambda_3^\varepsilon}}{\lambda_3^\varepsilon} - \frac{(1-u)^{\lambda_4^\varepsilon}}{\lambda_4^\varepsilon} \right), \quad 0 \le u \le 1, \quad \boldsymbol{\varepsilon} = Q(u, \lambda_1^\varepsilon, \lambda_2^\varepsilon, \lambda_3^\varepsilon, \lambda_4^\varepsilon).$$

Соответствующая ей функция плотности имеет вид [4]

$$g(z_i) = \frac{\lambda_2^\varepsilon}{u^{\lambda_3^\varepsilon - 1} + (1-u)^{\lambda_4^\varepsilon - 1}}.$$

Учитывая тот факт, что ошибки в объясняющих переменных не могут быть выражены в явном виде, функция распределения записывается как

$$z_i = y_i - \sum_{j=1}^{m} \theta_j h_j(\tilde{x}_{ij} - \boldsymbol{\delta}_j),$$

$$Q(u_i, \lambda_1^{\delta_j}, \lambda_2^{\delta_j}, \lambda_3^{\delta_j}, \lambda_4^{\delta_j}) = \lambda_1^{\delta_j} + \frac{1}{\lambda_2^{\delta_j}} \left(\frac{u^{\lambda_3^{\delta_j}}}{\lambda_3^{\delta_j}} - \frac{(1-u)^{\lambda_4^{\delta_j}}}{\lambda_4^{\delta_j}} \right),$$

$$0 \le u \le 1, \quad \delta_j = Q(u_i, \lambda_1^{\delta_j}, \lambda_2^{\delta_j}, \lambda_3^{\delta_j}, \lambda_4^{\delta_j}), \quad i = 1, ..., n, \quad j = 1, ..., m.$$

Обозначим через $f_{\delta_j}(z_i)$ соответствующую функцию плотности распределения.

Для вычисления элементов информационной матрицы было сформулировано и доказано следующее утверждение.

Утверждение. *Для регрессионной модели (1), (2), где на ошибки* $\boldsymbol{\varepsilon}$ *и* $\boldsymbol{\delta}_j$, $j = 1, ..., m$, *имеющие GL-распределение, наложены ограничения (3), элементы информационной матрицы вычисляются по формуле*

$$M_{lk} = -\sum_{i=1}^{n} h_l(x_{il}) h_k(x_{ik}) \int_0^1 g_\varepsilon''(z_i) g_\varepsilon(z_i) du -$$

$$-\sum_{i=1}^{n} \frac{h_l(x_{il}) h_k(x_{ik})}{\theta_l \theta_k \dfrac{\partial h_l(x_{il})}{\partial x_i} \dfrac{\partial h_k(x_{ik})}{\partial x_i}} \int_0^1 f_{\delta_j}''(z_i) f_{\delta_j}(z_i) du. \tag{7}$$

Доказательство. Как известно из [8], элементы информационной матрицы вычисляются по формуле

$$M_{lk} = -E\left(\frac{\partial^2 \ln L(\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \theta_l \partial \theta_k} \right), \quad l = 1, ..., m, \quad k = 1, ..., m. \tag{8}$$

Для доказательства (7) необходимо записать выражение для вычисления логарифма функции правдоподобия для моделей с ошибками в объясняющих переменных. Поскольку $\boldsymbol{\varepsilon}$ и $\boldsymbol{\delta}_j$, $j = 1, ..., m$, являются независимыми, функция правдоподобия имеет вид [6]

$$L(\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\theta}) = \Pi_{i=1}^{n} \Pi_{j=1}^{m} f_{\delta_j}(z_i) g(z_i). \tag{9}$$

После логарифмирования соотношения (9) получим

$$\ln L(\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\theta}) = \ln \left(\Pi_{i=1}^{n} \Pi_{j=1}^{m} f_{\delta_j}(z_i) g(z_i) \right) =$$

$$= \sum_{i=1}^{n} \left(\ln \Pi_{j=1}^{m} f_{\delta_j}(z_i) g(z_i) \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \ln f_{\delta_j}(z_i) + \ln g(z_i) \right).$$

Найдем соотношения для вычисления первой и второй частных производных логарифма функции правдоподобия. Так как функции плотностей распределения и выражение для вычисления отклонений наблюдаемых значений объясняющих переменных

от истинных заданы неявно, первая частная производная функции правдоподобия по параметрам θ_l может быть вычислена следующим образом [9]:

$$\frac{\partial \ln L(\boldsymbol{\varepsilon},\boldsymbol{\delta},\boldsymbol{\theta})}{\partial\theta_l}=\sum_{i=1}^{n}\frac{1}{g(z_i)}\frac{\partial g(z_i)}{\partial u_i}\frac{\partial u_i}{\partial Q(u_i,\lambda_1^{\varepsilon},\lambda_2^{\varepsilon},\lambda_3^{\varepsilon},\lambda_4^{\varepsilon})}\frac{\partial Q(u_i,\lambda_1^{\varepsilon},\lambda_2^{\varepsilon},\lambda_3^{\varepsilon},\lambda_4^{\varepsilon})}{\partial\theta_l}+$$

$$+\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{f_{\delta_j}(z_i)}\frac{\partial f_{\delta_j}(z_i)}{\partial u_i}\frac{\partial u_i}{\partial Q(u_i,\lambda_1^{\delta_j},\lambda_2^{\delta_j},\lambda_3^{\delta_j},\lambda_4^{\delta_j})}\frac{\partial Q(u_i,\lambda_1^{\delta_j},\lambda_2^{\delta_j},\lambda_3^{\delta_j},\lambda_4^{\delta_j})}{\partial\theta_l}=$$

$$=-\sum_{i=1}^{n}\frac{h_l(x_{il})}{g(z_i)}g^{'}(z_i)\frac{\partial u_i}{\partial Q(u_i,\lambda_1^{\varepsilon},\lambda_2^{\varepsilon},\lambda_3^{\varepsilon},\lambda_4^{\varepsilon})}-$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{f^{'}_{\delta_j}(z_i)}{f_{\delta_j}(z_i)}\frac{\partial u_i}{\partial Q(u_i,\lambda_1^{\delta_j},\lambda_2^{\delta_j},\lambda_3^{\delta_j},\lambda_4^{\delta_j})}\frac{h_l(x_{il})}{(-\theta_l)\dfrac{\partial h_l(x_{il})}{\partial x_i}(-1)}=$$

$$=-\sum_{i=1}^{n}\frac{h_l(x_{il})}{g(z_i)}g(z_i)g^{'}(z_i)-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{h_l(x_{il})}{f_{\delta_j}(z_i)\theta_l\dfrac{\partial h_l(x_{il})}{\partial x_i}}f_{\delta_j}(z_i)f^{'}_{\delta_j}(z_i)=$$

$$=-\sum_{i=1}^{n}h_l(x_{il})g^{'}(z_i)-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{h_l(x_{il})f^{'}_{\delta_j}(z_i)}{\theta_l\dfrac{\partial h_l(x_{il})}{\partial x_i}},\quad l=1,...,m.$$

Вторая частная производная функции правдоподобия по параметрам имеет вид

$$\frac{\partial^2 L(\boldsymbol{\varepsilon},\boldsymbol{\delta},\boldsymbol{\theta})}{\partial\theta_l\partial\theta_k}=-\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta_k}h_l(x_{il})g^{'}(z_i)\right)-\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{\partial}{\partial\theta_k}\frac{h_l(x_{il})f^{'}_{\delta_j}(z_i)}{\theta_l\dfrac{\partial h_l(x_{il})}{\partial x_i}}\right)=$$

$$=-\sum_{i=1}^{n}h_l(x_{il})\frac{\partial g^{'}(z_i)}{\partial\theta_k}-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{h_l(x_{il})}{\theta_l\dfrac{\partial h_l(x_{il})}{\partial x_i}}\frac{\partial f^{'}_{\delta_j}(z_i)}{\partial\theta_k}=$$

$$=-\sum_{i=1}^{n}h_l(x_{il})\frac{\partial g^{'}(z_i)}{\partial u}\frac{\partial u}{\partial Q(u_i,\lambda_1^{\varepsilon},\lambda_2^{\varepsilon},\lambda_3^{\varepsilon},\lambda_4^{\varepsilon})}\frac{\partial Q(u_i,\lambda_1^{\varepsilon},\lambda_2^{\varepsilon},\lambda_3^{\varepsilon},\lambda_4^{\varepsilon})}{\partial\theta_k}-$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{h_l(x_{il})}{\theta_l\dfrac{\partial h_l(x_{il})}{\partial x_i}}\frac{\partial f^{'}_{\delta_j}(z_i)}{\partial u}\frac{\partial u}{\partial Q(u_i,\lambda_1^{\delta_j},\lambda_2^{\delta_j},\lambda_3^{\delta_j},\lambda_4^{\delta_j})}\frac{\partial Q(u_i,\lambda_1^{\delta_j},\lambda_2^{\delta_j},\lambda_3^{\delta_j},\lambda_4^{\delta_j})}{\partial\theta_k}=$$

$$=\sum_{i=1}^{n}h_l(x_{il})h_k(x_{ik})g^{''}(z_i)g(z_i)+\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{h_l(x_{il})h_k(x_{ik})f^{''}_{\delta_j}(z_i)f_{\delta_j}(z_i)}{\theta_l\theta_k\dfrac{\partial h_l(x_{il})}{\partial x_i}\dfrac{\partial h_k(x_{ik})}{\partial x_i}},\quad l,k=\overline{1,m}.\qquad(10)$$

В соотношение (8) подставим (10) и после некоторых преобразований получим

$$M_{lk} = -E\left(\frac{\partial^2 L(\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\theta})}{\partial\theta_l \partial\theta_k}\right) =$$

$$= -E\left(\sum_{i=1}^{n} h_l(x_{il})h_k(x_{ik})g^{''}(z_i)g(z_i) + \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{h_l(x_{il})h_k(x_{ik})f^{''}_{\delta_j}(z_i)f_{\delta_j}(z_i)}{\theta_l\theta_k \dfrac{\partial h_l(x_{il})}{\partial x_i}\dfrac{\partial h_k(x_{ik})}{\partial x_i}}\right).$$

Известно [8], что для нахождения математического ожидания необходимо вычислить следующий интеграл:

$$-E\left(\sum_{i=1}^{n} h_l(x_{il})h_k(x_{ik})g^{''}(z_i)g(z_i) + \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{h_l(x_{il})h_k(x_{ik})f^{''}_{\delta_j}(z_i)f_{\delta_j}(z_i)}{\theta_l\theta_k \dfrac{\partial h_l(x_{il})}{\partial x_i}\dfrac{\partial h_k(x_{ik})}{\partial x_i}}\right) =$$

$$= -\int_{-\infty}^{\infty}\left(\sum_{i=1}^{n} h_l(x_{il})h_k(x_{ik})g^{''}(z_i)g(z_i)\right)\prod_{r=1}^{m} f_{\delta_r}(z_i)g(z_i)dz_i -$$

$$-\int_{-\infty}^{\infty}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{h_l(x_{il})h_k(x_{ik})f^{''}_{\delta_j}(z_i)f_{\delta_j}(z_i)}{\theta_l\theta_k \dfrac{\partial h_l(x_{il})}{\partial x_i}\dfrac{\partial h_k(x_{ik})}{\partial x_i}}\right)\prod_{r=1}^{m} f_{\delta_r}(z_i)g(z_i)dz_i =$$

$$= -\int_{0}^{1}\left(\sum_{i=1}^{n} h_l(x_{il})h_k(x_{ik})g^{''}(z_i)g(z_i)\right)du - \int_{0}^{1}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{h_l(x_{il})h_k(x_{ik})f^{''}_{\delta_j}(z_i)f_{\delta_j}(z_i)}{\theta_l\theta_k \dfrac{\partial h_l(x_{il})}{\partial x_i}\dfrac{\partial h_k(x_{ik})}{\partial x_i}}\right)du,$$

$$k, l = \overline{1, m}.$$

Выполнив некоторые преобразования, получим, что элементы информационной матрицы могут быть вычислены следующим образом:

$$M_{lk} = -\sum_{i=1}^{n} h_l(x_{il})h_k(x_{ik})\int_{0}^{1} g^{''}(z_i)g(z_i)du - \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{h_l(x_{il})h_k(x_{ik})}{\theta_l\theta_k \dfrac{\partial h_l(x_{il})}{\partial x_i}\dfrac{\partial h_k(x_{ik})}{\partial x_i}}\int_{0}^{1} f^{''}_{\delta_j}(z_i)f_{\delta_j}(z_i)du,$$

$$k, l = \overline{1, m}.$$

Утверждение доказано. □

Следствие. *Для регрессионной модели (1), (2), где на ошибки* $\boldsymbol{\varepsilon}$ *и* $\boldsymbol{\delta}_j$, $j = 1, ..., m$, *имеющие GL-распределение, наложены ограничения (3) и измерения проводятся в соответствии с нормированным планом* $\xi = \begin{pmatrix} x_1 & x_2 & ... & x_s \\ p_1 & p_2 & ... & p_s \end{pmatrix}$, $\sum\limits_{i=1}^{s} p_i = 1$, $p_i = \dfrac{n_i}{n}$, s — *количество точек в спектре плана и* n_i — *количество повторных наблюдений в каждой*

точке спектра плана, соотношение для вычисления функции эффективности имеет вид

$$\frac{1}{\lambda(x_i)} = \left(1, \frac{\partial h_1(x_i)}{\partial x_i}, ..., \frac{\partial h_m(x_i)}{\partial x_i}\right) \hat{\Sigma}_{\varepsilon\delta} \left(1, \frac{\partial h_1(x_i)}{\partial x_i}, ..., \frac{\partial h_m(x_i)}{\partial x_i}\right)^T, \tag{11}$$

где

$$\hat{\Sigma}_{\varepsilon\delta} = \begin{pmatrix} \int\limits_0^1 g''(z)g(z)du & 0 & ... & 0 \\ 0 & \int\limits_0^1 f''_{\delta_1}(z)f_{\delta_1}(z)du & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \int\limits_0^1 f''_{\delta_m}(z)f_{\delta_m}(z)du \end{pmatrix}.$$

Доказательство. Сопоставим выражения (5) и (7). Нетрудно заметить, что

$$\int\limits_0^1 g''_{\varepsilon}(z_i)g_{\varepsilon}(z_i)du \quad \text{и} \quad \int\limits_0^1 f''_{\delta_j}(z_i)f_{\delta_j}(z_i)du$$

есть обратная величина дисперсий случайных ошибок $\boldsymbol{\varepsilon}$ и $\boldsymbol{\delta}_j$, $j = 1, ..., m$, соответственно, а знаменатель в (7) играет роль производных из выражения (6).

Следствие доказано. □

3. Результаты исследований

В данной работе приведены результаты исследований синтеза оптимальных планов эксперимента с использованием доказанных утверждения и следствия. В качестве истинной зависимости использовалась следующая модель:

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \varepsilon_i,$$

где x_{i1}, x_{i2}, $i = \overline{1,n}$, — неизвестные истинные значения входных факторов x_1 и x_2, наблюдаемые значения которых

$$\tilde{x}_{ij} = x_{ij} + \delta_{ij}, \quad i = 1, ..., n, \quad j = 1, ..., m;$$

причем $x_1 \in [-1, 1]$; $x_2 \in [0, 1]$, количество регрессоров $m = 3$; истинные значения вектора параметров регрессионной модели $\boldsymbol{\theta}^{\text{ист}} = (25, 25, 25)^T$; случайные ошибки ε_i и δ_{ij}, $i = 1, ..., n$, $j = 1, .., m$, — независимые и имеют GL-распределение с параметрами $(\lambda_1^{\varepsilon}, \lambda_2^{\varepsilon}, \lambda_3^{\varepsilon}, \lambda_4^{\varepsilon})$ и $(\lambda_1^{\delta_j}, \lambda_2^{\delta_j}, \lambda_3^{\delta_j}, \lambda_4^{\delta_j})$ соответственно. Задача состоит в построении оптимального нормированного плана эксперимента:

$$\xi = \begin{pmatrix} x_1 & x_2 & ... & x_s \\ p_1 & p_2 & ... & p_s \end{pmatrix}.$$

Здесь s — количество точек в спектре плана.

Т а б л и ц а 1. Вычисление функции эффективности
Table 1. Calculation results of the efficiency function

Нормальное распределение	GL-распределение	$\lambda(x)$ по (6)	$\lambda(x)$ по (11)
$\boldsymbol{\varepsilon} \sim N(0,1)$ $\boldsymbol{\delta}_1 \sim N(0,1)$ $\boldsymbol{\delta}_2 \sim N(0,1)$	$\boldsymbol{\varepsilon} \sim GLD(0, 1.408, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 1.408, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 1.408, 0.161, 0.161)$	$7.994\text{E}-4$	$8.021\text{E}-4$
$\boldsymbol{\varepsilon} \sim N(0,0.25)$ $\boldsymbol{\delta}_1 \sim N(0,0.25)$ $\boldsymbol{\delta}_2 \sim N(0,0.25)$	$\boldsymbol{\varepsilon} \sim GLD(0, 2.816, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 2.816, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 2.816, 0.161, 0.161)$	$3.197\text{E}-3$	$3.205\text{E}-3$
$\boldsymbol{\varepsilon} \sim N(0,2)$ $\boldsymbol{\delta}_1 \sim N(0,2)$ $\boldsymbol{\delta}_2 \sim N(0,2)$	$\boldsymbol{\varepsilon} \sim GLD(0, 0.996, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 0.996, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 0.996, 0.161, 0.161)$	$3.997\text{E}-4$	$4.014\text{E}-4$
$\boldsymbol{\varepsilon} \sim N(0,1)$ $\boldsymbol{\delta}_1 \sim N(0,0.25)$ $\boldsymbol{\delta}_2 \sim N(0,0.25)$	$\boldsymbol{\varepsilon} \sim GLD(0, 1.408, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 2.816, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 2.816, 0.161, 0.161)$	$3.190\text{E}-3$	$3.201\text{E}-3$
$\boldsymbol{\varepsilon} \sim N(0,1)$ $\boldsymbol{\delta}_1 \sim N(0,0.25)$ $\boldsymbol{\delta}_2 \sim N(0,2)$	$\boldsymbol{\varepsilon} \sim GLD(0, 1.408, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 2.816, 0.161, 0.161)$ $\boldsymbol{\delta}_1 \sim GLD(0, 0.996, 0.161, 0.161)$	$7.106\text{E}-4$	$7.121\text{E}-4$

Как отмечалось ранее, нормальное распределение является частным случаем GL-распределения, поэтому вычисленные с помощью соотношений (6) и (11) значения функции эффективности должны совпадать.

Рассмотрим нормированный план вида

$$\xi = \begin{pmatrix} (-1;0) & (1;0) & (1;1) & (-1;1) \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix},$$

количество точек в спектре плана $s = 4$. Ошибки во всей области планирования имеют нормальное распределение с одинаковыми дисперсиями.

В результате вычисления функций эффективности с помощью соотношений (6) и (11) получены значения, которые представлены в табл. 1. В первом столбце приведены параметры нормального распределения ошибок наблюдений, во втором — соответствующие ему значения параметров GL-распределения, в третьем столбце — значения функции эффективности, вычисленные с использованием соотношения (6), в четвертом — значения, вычисленные предложенным способом (11). Из табл. 1 видно, что предложенный метод вычисления функции эффективности (см. (11)) дает результаты, схожие с методом для нормального распределения ошибок, где функция эффективности вычисляется с помощью соотношения (6). Данный результат демонстрирует справедливость утверждения и следствия.

Далее проведено построение оптимальных планов с использованием предложенного в работе [1] алгоритма планирования, где информационная матрица имеет вид (4), функция эффективности — (11). При построении оптимального плана выбран критерий D-оптимальности плана. Напомним [1], что план называется D-оптимальным, если он минимизирует определитель дисперсионной матрицы (максимизирует определитель информационной матрицы). Этот критерий оптимальности минимизирует обобщенную дисперсию всех оценок регрессионной модели:

$$\xi^* = Arg \min_{\xi} |M^{-1}(\xi)|, \quad \left(\xi^* = Arg \max_{\xi} |M(\xi)| \right).$$

Согласно теореме эквивалентности [1] условие D-оптимальности плана можно записать следующим образом:

$$\lambda(x)d(x,\xi^*) = m, \quad x \in \xi^*, \tag{12}$$

где $d(x,\xi^*) = \mathbf{h}^T(x)M(\xi^*)\mathbf{h}(x)$ — дисперсия оценки функции отклика в точке.

Рассмотрены два варианта:

1) ошибки на всей области планирования имеют однородное распределение, т. е. функция эффективности постоянна на всей области планирования;

2) ошибки на всей области планирования имеют неоднородное распределение, т. е. элементы информационной матрицы вычисляются с использованием соотношения (4).

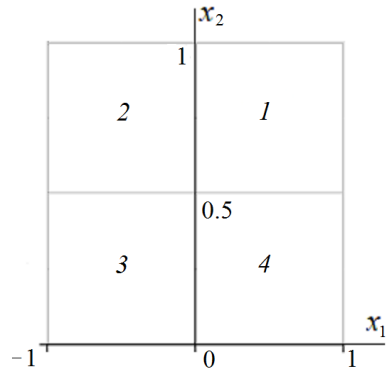
Далее представлены результаты исследования четырех случаев распределения ошибок $\boldsymbol{\varepsilon}$, $\boldsymbol{\delta}_1$ и $\boldsymbol{\delta}_2$ на области планирования:

1. Одинаковое нормальное распределение ошибок на всей области планирования $\boldsymbol{\varepsilon} \sim N(0;1)$, $\boldsymbol{\delta}_i \sim N(0;0.25)$, $j = 1,2$.

2. Во второй и четвертой четвертях области планирования ошибки $\boldsymbol{\varepsilon} \sim N(0;1)$, $\boldsymbol{\delta}_i \sim N(0;0.25)$, $i = 1,2$; в первой и третьей четвертях — $\boldsymbol{\varepsilon} \sim N(0;1)$, $\boldsymbol{\delta}_i \sim N(0;2)$, $i = 1,2$, разбиение области планирования по четвертям представлено на рисунке.

Рассмотрим ситуации, когда ошибки наблюдений имеют отличное от нормального распределение. В качестве исследуемых взяты следующие GL-распределения: несимметричное с левой асимметрией $GLD_1(0;1;0.002;0.5)$, несимметричное с правой асимметрией $GLD_2(0;1;0.5;0.002)$ и симметричное $GLD_3(0;1;0.5;0.5)$.

3. Во второй и четвертой четвертях области планирования ошибки $\boldsymbol{\varepsilon}$, $\boldsymbol{\delta}_1$ и $\boldsymbol{\delta}_2$ имеют одинаковое асимметричное распределение GLD_1, в первой и третьей четвертях — GLD_2.



Распределение случайных ошибок по подобластям области планирования / Distribution of random errors by sub-areas of the planning area

Т а б л и ц а 2. Вычисление функций эффективности на области планирования
Table 2. Calculation results of the efficiency functions in the planning area

Случай	Значение функции эффективности на каждой четверти			
	1	2	3	4
1	3.201E−3	3.201E−3	3.201E−3	3.201E−3
2	4.02E−4	3.201E−3	4.02E−4	3.201E−3
3	2.624E−3	2.624E−3	2.624E−3	2.624E−3
4	4.973E−3	2.624E−3	4.973E−3	2.624E−3

Т а б л и ц а 3. Построение оптимальных планов экспериментов
Table 3. Optimal experimental designs

Случай	Оптимальный план	(12)
1	$\begin{pmatrix} (-1;0) & (1;0) & (1;1) & (-1;1) \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$	3.0020
2	$\xi = \begin{pmatrix} (-1;1) & (1;1) & (-1;0.5) & (0.8;0.5) \\ 0.263 & 0.263 & 0.237 & 0.237 \end{pmatrix}$	3.0017
3	$\xi = \begin{pmatrix} (-1;0) & (1;0) & (1;1) & (-1;1) \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$	3.0018
4	$\xi = \begin{pmatrix} (-1;0) & (-1;1) & (1;0) & (1;1) \\ 0.286 & 0.213 & 0.286 & 0.215 \end{pmatrix}$	3.0016

4. Во второй четверти области планирования ошибки $\boldsymbol{\varepsilon}$, $\boldsymbol{\delta}_1$ и $\boldsymbol{\delta}_2$ имеют одинаковое асимметричное распределение GLD_1, в четвертой четверти — GLD_2, в первой и третьей четвертях — GLD_3.

В табл. 2 представлены результаты вычисления функции эффективности на каждой четверти области планирования эксперимента, в табл. 3 — результаты синтеза планов эксперимента. Как видно из табл. 3, построенные планы удовлетворяют условию оптимальности. Как следует из табл. 2 и 3, при постоянной функции эффективности на всей области планирования оптимальным является равновесный четырехточечный план. При появлении различий в значениях функции эффективности на области планирования оптимальный план перестает быть равновесным.

Заключение

Рассмотрена задача синтеза оптимальных планов эксперимента в условиях появления ошибок в объясняющих переменных. Сформулировано и доказано утверждение о способе вычисления элементов информационной матрицы Фишера в условиях GL-распределения случайных ошибок для таких моделей, доказано следствие о способе вычисления функции эффективности плана эксперимента. Проведено сравнение значений функций эффективности планов эксперимента, вычисленных с использованием соотношения из работы [5] и предложенного способа. Установлено, что при нормальном распределении случайных компонент полученные значения функций эффективности совпадают. Приведены результаты построения оптимальных планов при различных условиях вычислительных экспериментов.

Список литературы

[1] **Федоров В.В.** Теория оптимального планирования эксперимента. М.: Наука; 1971: 312.

[2] **Тимофеев В.С., Хайленко Е.А.** Оптимальное планирование эксперимента для регрессионных моделей с обобщенным лямбда-распределением ошибок. Научный вестник НГТУ. 2011; 1(42):27–37.

[3] **Karian Z.A., Dudewicz E.J.** Fitting statistical distributions: the Generalized Lambda Distribution and Generalized Bootstrap methods. New York: CRC Press LLC; 2000: 435.

[4] **Lakhany A., Mausser H.** Estimation the parameters of the Generalized Lambda Distribution. ALGO Research Quarterly. 2000; 3(3):27–58.

[5] **Konstantionu M., Dette H.** Locally optimal designs for errors-in-variables models. Biometrika. 2015; 102(4):951–958.

[6] **Грешилов А.А., Стакун В.А., Стакун А.А.** Математические методы построения прогнозов. М.: Радио и связь; 1997: 112.

[7] **Fuller W.A.** Measurement error models. New York: John Wiley and Sons; 1987: 440.

[8] **Ивченко Г.И., Медведев Ю.И.** Математическая статистика. Учеб. пособие для втузов. М.: Высшая школа; 1984: 248.

[9] **Кудрявцев Л.Д.** Курс математического анализа. В 3 томах. Т. 1. М.: Дрофа; 2003: 704.

Вычислительные технологии, 2020, том 25, № 3, с. 130–141. © ИВТ СО РАН, 2020 ISSN 1560-7534
Computational Technologies, 2020, vol. 25, no. 3, pp. 130–141. © ICT SB RAS, 2020 eISSN 2313-691X

INFORMATION TECHNOLOGIES

DOI:10.25743/ICT.2020.25.3.014

Synthesis of optimal experiment designs under the conditions of Generalized Lambda-distribution of errors for models with errors in variables

Timofeev Vladimir S.*, Khailenko Ekaterina A.
Novosibirsk State Technical University, 630073, Novosibirsk, Russia
*Corresponding author: Timofeev Vladimir S., e-mail: v.timofeev@corp.nstu.ru
Received March 10, 2020, revised March 24, 2020, accepted April 10, 2020

Abstract

The problem of experimental design under conditions of errors in the explanatory variables is considered. The proposition of the method for calculating the Fisher information matrix elements using the Generalized Lambda-distribution is formulated and proved, the consequence of the method for calculating the efficiency function of the experimental design is proved. This method of calculating the Fisher information matrix takes into account the heterogeneity of the errors in random distribution throughout the planning area. In this paper, studies of the synthesis of optimal experimental designs using proven proposition and consequence under various conditions of computational experiments are presented. The results of calculating the efficiency function using the obtained relation and using the known relation for the normal distribution of errors are compared, it is found that the results coincide. Optimal experimental designs are constructed for various distributions of random components. The results of the synthesis of optimal experimental design showed that when function of efficiency is constant throughout the planning area then the optimal experimental design is equilibrium plan. When there are differences in the values of the efficiency function in the planning area, the optimal plan ceases to be equilibrium.

Keywords: regression dependencies, models with errors in explanatory variables, generalized lambda distribution, experimental design, Fisher information matrix.

Citation: Timofeev V.S., Khailenko E.A. Synthesis of optimal experiment designs under the conditions of Generalized Lambda-distribution of errors for models with errors in variables. Computational Technologies. 2020; 25(3):130–141. (In Russ.)

References

1. Fedorov V.V. Teoriya optimal'nogo planirovaniya eksperimenta [Theory of optimal experimental design]. Moscow: Nauka; 1971: 312. (In Russ.)
2. Timofeev V.S., Khailenko E.A. Optimal designing an experiment for regression models with generalized lambdaistributed errors. Science Bulletin of the NSTU. 2011; 1(42):27–38. (In Russ.)
3. Karian Z.A., Dudewicz E.J. Fitting statistical distributions: the Generalized Lambda Distribution and Generalized Bootstrap methods. New York: CRC Press LLC; 2000: 435.
4. Lakhany A., Mausser H. Estimation the parameters of the Generalized Lambda Distribution. ALGO Research Quarterly. 2000; 3(3):27–58.
5. Konstantionu M., Dette H. Locally optimal designs for errors-in-variables models. Biometrika. 2015; 102(4):951–958.
6. Greshilov A.A., Stakun V.A., Stakun A.A. Matematicheskie metody postroeniya prognozov [Mathematical methods for building forecasts]. Moscow: Radio i Svyaz'; 1997: 112. (In Russ.)
7. Fuller W.A. Measurement error models. New York: John Wiley and Sons; 1987: 440.
8. Ivchenko G.I., Medvedev Yu.I. Matematicheskaya statistika [Mathematical statistics]. Uchebnoe posobie dlya vtuzov. Moscow: Vysshaya Shkola; 1984: 248. (In Russ.)
9. Kudryavtsev L.D. Kurs matematicheskogo analiza. V 3 tomakh. T. 1 [Calculus. Vol. 1]. Moscow: Drofa; 2003: 704. (In Russ.)