

## Влияние методов оценивания параметров на распределения статистик дисперсионного анализа при нарушении предположений о нормальности ошибок наблюдений<sup>1</sup>

Лемешко Б.Ю., Пономаренко В.М.  
НГТУ, Новосибирск. E-mail: headrd@fpm.ami.nstu.ru

В данной работе, являющейся продолжением [1], исследуется влияние метода оценивания параметров модели наблюдений на поведение предельного распределения статистики отношения правдоподобия при отклонении закона распределения ошибок от нормального.

Рассматриваемая модель отклика имеет вид:

$$Y = X\theta + e, \quad (1)$$

где  $Y$  – вектор наблюдений размерности  $(n \times 1)$ ,  $X$  – матрица планирования размерности  $(n \times m)$ ,  $r = \text{rg}(X)$  – ранг  $X$ ,  $\theta$  – вектор оцениваемых параметров размерности  $(m \times 1)$ ,  $e$  – вектор ошибок наблюдений размерности  $(n \times 1)$ . В самом общем виде линейную гипотезу относительно параметров можно представить следующим образом

$$H : K\hat{\theta} = b, \quad (2)$$

где  $K$  – известная матрица размерности  $(k \times m)$ ,  $\text{rg}(K) = k \leq m$ ;  $b$  – заданный вектор размерности  $(k \times 1)$ ,  $\hat{\theta}$  – оценка вектора параметров  $\theta$ , вычисленная одним из рассматриваемых методов: либо методом наименьших квадратов (МНК), либо методом максимального правдоподобия (ММП). Для проверки гипотез вида (2) в дисперсионном анализе используется статистика вида:

$$Q = \frac{(b - K\hat{\theta})^T (KGK^T)^{-1} (b - K\hat{\theta})}{(Y - X\hat{\theta})^T (Y - X\hat{\theta})} \cdot \frac{n - r}{k}, \quad (3)$$

называемая статистикой отношения правдоподобия, где  $G$  – матрица, обобщенно-обратная к матрице  $X^T X$ . В случае нормального распределения ошибок наблюдений статистика (3) подчиняется  $F$ -распределению Фишера со степенями свободы  $k$  и  $n - r$ . Распределение Фишера  $F(k, n - r)$  является частным случаем Бета-распределения II рода (Бе-II):

$$F(k, n - r) = \text{Be}_{II}(0, (n - r)/k; k/2, (n - r)/2). \quad (4)$$

Применительно к показателям используемой модели (1), статистика (3) будет иметь распределение со степенями свободы  $\text{rg}(K)$  и  $n - \text{rg}(X)$ .

Исследования предельных распределений статистики (3) проводились для следующих распределений ошибок наблюдений: нормального, Коши, двустороннего экспоненциального с параметром формы  $\lambda$ , принимающим значения

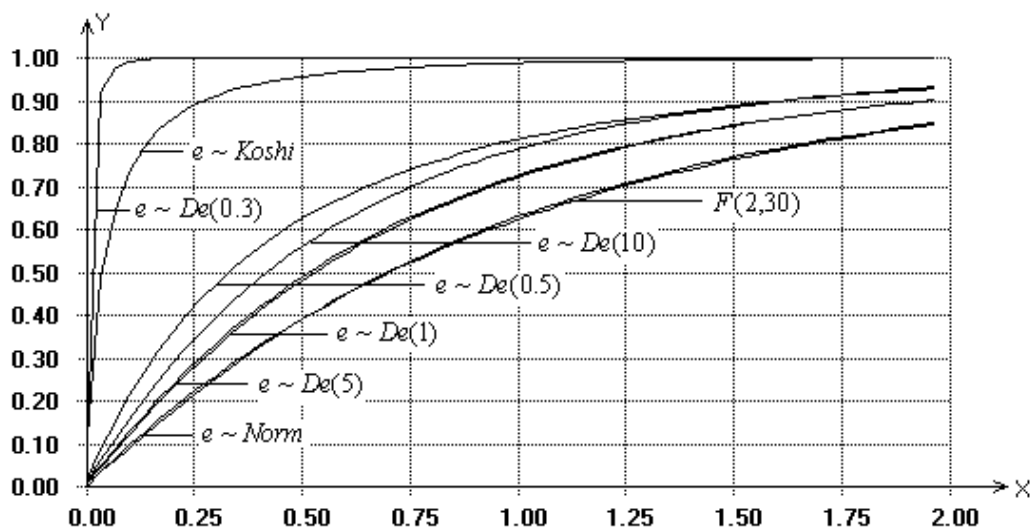
<sup>1</sup> Работа выполнена при финансовой поддержке Минобразования РФ (проект № ТО2-3.3-3356)

0.3, 0.4, 0.5, 1, 3, 5, 10, обозначаемого далее как  $De(\lambda)$ .

Исследования проводились с помощью методики компьютерного моделирования. Построенные в результате моделирования выборки исследуемых статистик в дальнейшем анализируются на предмет принадлежности к некоторому классу теоретических распределений с оцениванием неизвестных параметров предполагаемого закона. Параметры предполагаемого закона, используемого в качестве модели предельного распределения статистики, усреднялись по ряду экспериментов. В большинстве рассмотренных случаев наиболее подходящим классом теоретических распределений для описания эмпирических распределений статистик оказывался класс распределений  $Be-II$ . В некоторых случаях это был класс распределений  $Sb$ -Джонсона,  $Be-III$ , Вейбулла, Максимального значения.

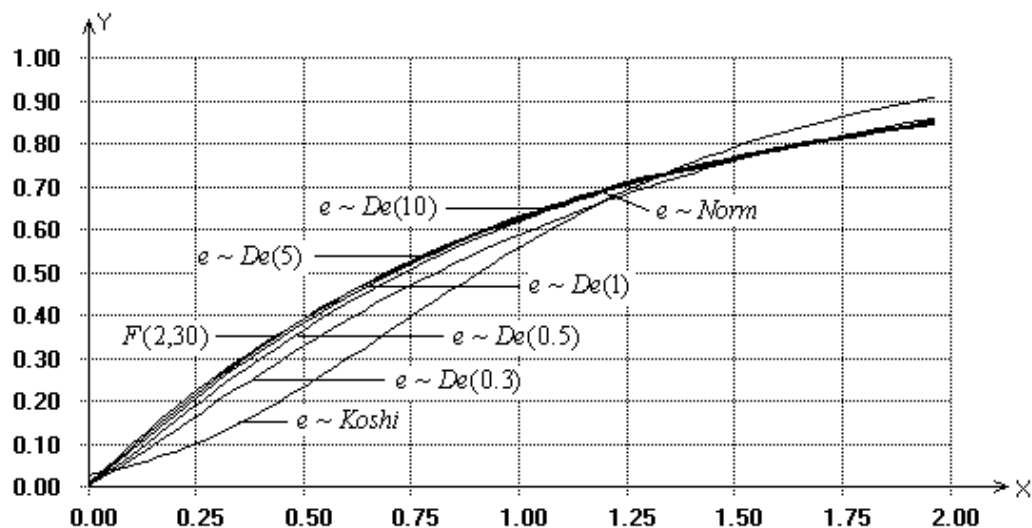
Исследования проводились на нескольких наборах сбалансированных моделей. Каждый набор состоял из десяти моделей, позволяющих построить одно приближение предельного распределения статистики. Рассматривался ряд наборов однофакторных моделей [2-3], два набора двухфакторных моделей с показателями  $n=36$ ,  $m=8$ ,  $r=6$ ,  $k=2$  и по одному набору двухфакторных моделей с показателями  $n=120$ ,  $m=8$ ,  $r=6$ ,  $k=2$  и  $n=1200$ ,  $m=8$ ,  $r=6$ ,  $k=2$ . В полном соответствии с классическими результатами в случае нормально распределенных ошибок наблюдений при  $n=36$  эмпирические распределения статистики (3) хорошо согласуется с распределением  $F(2,30)$ , при  $n=120$  – с распределением  $F(2,124)$ , при  $n=1200$  – с  $F(2,1194)$ .

При проведении исследований изучалось два вопроса. Во-первых, насколько сильно и каким образом при различных законах распределения ошибок приближения предельного распределения статистики (3) отклоняются от соответствующего  $F$ -распределения (от классических результатов). Во-вторых, какое влияние на распределение статистики (3) оказывает выбранный метод оценивания параметров. Все приводимые результаты имеют место для оговоренных выше двухфакторных моделей и согласуются с результатами, полученными для однофакторных моделей.



При оценивании параметров модели по методу максимального правдоподобия сохраняется наблюдавшаяся в случае однофакторных моделей тенден-

ция: при увеличении  $\lambda$  от 2 до 10 распределения статистики удаляются, “поднимаясь” от  $F$ -распределения Фишера. При уменьшении  $\lambda$  от 2 до 0.3 (закон распределения ошибок с “тяжелыми” хвостами) распределения статистики также удаляются, “поднимаясь” от  $F$ -распределения. Сказанное проиллюстрировано на приведенном рисунке, где отображены полученные приближения предельных распределений статистики при различных законах распределения ошибок (на рисунке имеющие сноску  $e \sim \dots$ ) и  $F$ -распределение в случае одного из набора моделей с показателями  $n = 36$ ,  $m = 8$ ,  $r = 6$ ,  $k = 2$ .



При оценивании параметров модели по МНК наблюдается совершенно другое поведение распределения статистики отношения правдоподобия. Во-первых, видимое отклонение получаемого приближения от  $F$ -распределения Фишера наблюдается только в случае распределения ошибок с тяжелыми хвостами (по Коши и  $De(0.3)$ ). В остальных случаях полученные приближения практически полностью совпадают с  $F$ -распределением, как это можно видеть на втором рисунке.

Полученные результаты позволяют сделать вывод о том, что выбор метода оценивания параметров очень существенно влияет на характер поведения распределения статистики при отклонении закона распределения ошибок от нормального, причем этот характер влияния сохраняется при использовании различных моделей.

1. Лемешко Б.Ю., Пономаренко В.М. Исследование влияния законов распределения ошибок на распределения статистик дисперсионного анализа // Мат. межд. НТК "Информатика и проблемы телекоммуникаций". – Новосибирск, 2002. – С. 122-123.
2. Лемешко Б.Ю., Пономаренко В.М., Трушина Е.А. К проверке статистических гипотез в регрессионном и дисперсионном анализе при нарушении предположений нормальности // Мат. VI всеросс. НТК “Информационные технологии в науке, проектировании и производстве”. – Н. Новгород, 2002. – С. 1-5.
3. Лемешко Б.Ю., Пономаренко В.М., Распределения статистик дисперсионного анализа при отклонении закона распределения ошибок наблюдений от нормального // Мат. VI межд. конф. “Актуальные проблемы электронного приборостроения”. – Новосибирск, 2002. – С. 36-38.