Материалы региональной конференции (с участием иностранных ученых) “Вероятностные идеи в науке и философии”. – Новосибирск: Ин-т философии и права СО РАН / НГУ. 2003. – С. 113-115.

Построение приближений предельного распределения статистики, используемой при проверке гипотез в линейном регрессионном анализе, при законах распределения ошибок, отличающихся от нормального[1]

Лемешко Б.Ю., Трушина Е.А.
НГТУ, Новосибирск. E-mail: headrd@fpm.ami.nstu.ru

В классическом регрессионном анализе аппарат проверки статистических гипотез основан на предположении о нормальности закона ошибок наблюдений. Нарушение данного предположения по-разному отражается на распределениях статистик используемых критериев проверки гипотез. Предельные распределения статистик критериев могут зависеть не только от закона распределения ошибок и применяемого метода оценивания параметров, но и от размерности исследуемой системы.

В данной работе исследуются предельные распределения статистик критериев, применяемых для проверки гипотез в линейном регрессионном анализе, при нарушении предположений о нормальности ошибок, в частности для моделей с ошибками, подчиняющимися следующим симметричным законам: логистическому с плотностью $f(x)=\frac{\pi}{\theta_0\sqrt{3}}\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\Big/\left[1+\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\right]^2$, Коши с плотностью $f(x)=\frac{\theta_0}{\pi[\theta_0^2+(x-\theta_1)^2]}$ и экспоненциальному семейству (ЭС) с плотностью $f(x)=\frac{\lambda}{2\sqrt{2}\theta_1\Gamma(1/\lambda)}\exp\left(-\left(\frac{|x-\theta_0|}{\sqrt{2}\theta_1}\right)^{\lambda}\right)$ при различных значениях параметра формы λ.

Рассматривается классическая модель линейной регрессии

$$Y = X\theta + e\ , \tag{1}$$

где Y – вектор наблюдений размерности $(n\times 1)$, X – детерминированная матрица независимых переменных размерности ($n\times m$), называемая регрессором системы, θ - вектор оцениваемых параметров системы размерности $(m\times 1)$, e – вектор случайных отклонений размерности ($n\times 1$).

Простая гипотеза относительно параметров модели имеет вид:

$$H:\hat{\theta}=r\ , \tag{2}$$

где r - заданный вектор $(k\times 1)$, $\hat{\theta}$ - оценка вектора параметров θ, определяемая, в данной работе, методом максимального правдоподобия (ММП). Для проверки гипотез используется статистика вида:

$$Q=\frac{(r-\hat{\theta})^T(X^TX)(r-\hat{\theta})}{ms^2} \tag{3}$$

где s^2 - оценка выборочной дисперсии отклонений регрессии.

В [1] показано, что в случае нормальности ошибок наблюдений, статистика (3) подчиняется закону распределения Фишера $F(m,n-m)$ со степенями свободы *m* и *n-m*. Отметим, что распределение Фишера является частным случаем бета-распределения II рода:

$$F(m,n-m)=Be_{II}\left(0,\frac{n-m}{m},\frac{m}{2},\frac{n-m}{2}\right), \tag{4}$$

где плотность бета-распределения задается выражением [2]

$$f(x)=\frac{\Gamma(\alpha+\beta)\left(\frac{x-a}{\sigma}\right)^{\alpha-1}}{\sigma\Gamma(\alpha)\Gamma(\beta)\left(1+\frac{x-a}{\sigma}\right)^{\alpha+\beta}}=Be_{II}(a,\sigma,\alpha,\beta)\cdot$$

[1] Работа выполнена при поддержке Минобразования РФ (проект № Т02-3.3-3356)

В [3] показано, что хорошей аппроксимацией предельных распределений для статистики критерия отношения правдоподобия является Бета-распределение II рода. Для различных значений числа наблюдений n и количества неизвестных параметров линейной регрессии m были найдены значения параметров Бета-распределения II рода, аппроксимирующие предельные распределения статистики.

Следующим логическим шагом после получения сетки значений параметров предельного распределения при различных значениях n и m является выявление вида зависимости параметров предельного распределения от размерности системы линейной регрессии.

На основании построенных моделей предельных распределений статистик в данной работе проведен анализ характера зависимостей параметров формы этих распределений как функций от значений n и m :

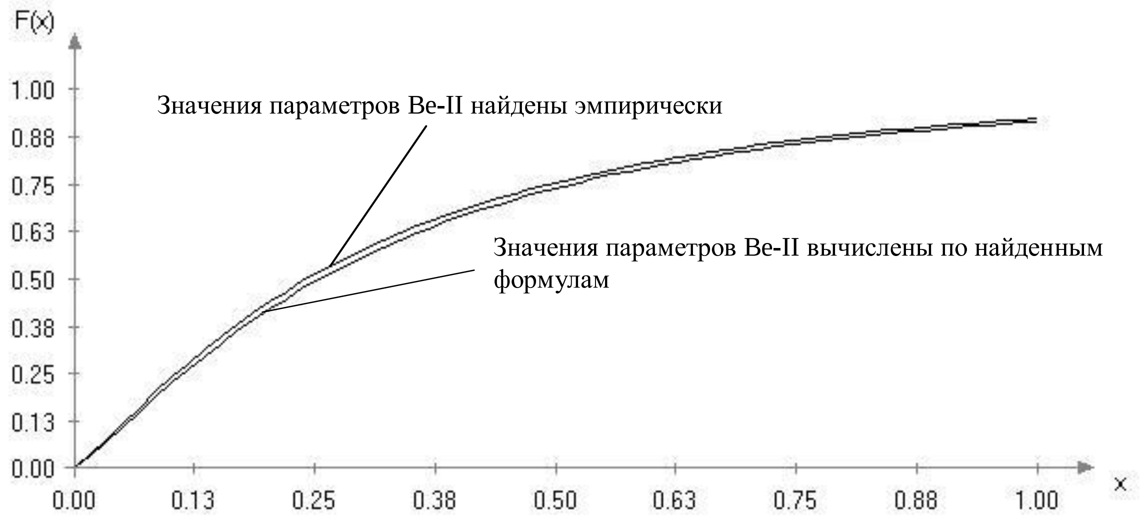
$$\alpha = Z_1(n,m) \qquad \beta = Z_2(n,m), \tag{5}$$

где α и β - первый и второй параметры формы Бета-распределения II рода соответственно, а Z_1, Z_2 - искомые функциональные зависимости. В данном случае для простоты вычислений выражения (5) рассматривались в виде линейных функций двух переменных:

$$Z(n,m) = a + bn + cm + dnm. \tag{6}$$

Вычислены коэффициенты соответствующих билинейных функций для экспоненциального семейства законов распределений с параметрами формы $\lambda = 0.5,\ \lambda = 1,\ \lambda = 3,\ \lambda = 5,\ \lambda = 10$.

На рисунке представлены модели предельных распределений статистики (3) для модели линейной регрессии с $n = 15,\ m = 4$ и ошибками наблюдений, подчиненными экспоненциальному семейству распределений с параметром формы $\lambda = 1.0$. Параметры моделей распределений в первом случае получены эмпирически [3], во втором случае вычислены в соответствии с полученными приближениями.



Результаты показывают, что полученные приближения достаточно хорошо описывают зависимость от n и m параметров формы полученных ранее Бета-распределений II рода и используемых в качестве предельных распределений статистик критерия отношения правдоподобия.

1. Демиденко Е.З. Линейная и нелинейная регрессия. – М.: Финансы и статистика, 1981. – 302 с.
2. Губарев В.В. Вероятностные модели: Справочник в 2-х ч. // Новосиб. электротехн. ин-т. - Новосибирск, 1992. – 422 с.
3. Лемешко Б.Ю., Трушина Е.А. Исследование вопросов проверки статистически гипотез в линейном регрессионном анализе // Тезисы докладов региональной НТК «Наука. Техника. Инновации».- Новосибирск.2001 – Т.2. –С.36-37.