Теорія Ймовір. та Матем. Статист.
Вип. 86, 2012

Theor. Probability and Math. Statist.
No. 86, 2013, Pages 79–91
S 0094-9000(2013)00890-4
Article electronically published on August 20, 2013

LIMIT THEOREMS FOR THE MAXIMAL RESIDUALS IN LINEAR AND NONLINEAR REGRESSION MODELS

UDC 519.21

A. V. IVANOV AND I. K. MATSAK

Abstract. Limit theorems for the maximal residuals in linear and nonlinear regression models are obtained in the paper. An application of the main result for constructing a regression model adequacy test is given.

Classical results of the regression analysis describe various properties of residual sums of squares errors (see, for example, [1, 2]). In other words, the classical results deal with the properties of the sum of squares of deviations between observations and a regression function where the least squares estimator is substituted in place of the unknown parameter.

The current paper deals with the asymptotic behavior of the maximal residual. We prove the convergence of distributions of the normalized appropriately maximal residual to a limit distribution of the normalized maximum as the size of a sample grows to infinity of identically distributed errors of observations. We consider both cases, the linear and nonlinear regression models.

We apply the asymptotic results to construct some regression model adequacy test (see Section 3).

1. Linear regression model

Consider the following linear regression model

$$y_j = \sum_{i=1}^{q} \theta_i x_{ji} + \varepsilon_j, \qquad j = 1, \dots, n, \tag{1}$$

where ε_j are independent identically distributed random variables such that $\mathsf{E}\,\varepsilon_j = 0$ and $\mathsf{E}\,\varepsilon_j^2 = \sigma^2 < \infty$.

Let

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1q} \\ x_{21} & x_{22} & \dots & x_{2q} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix}$$

be the $n \times q$ regression design matrix such that $\det(X^T X) \neq 0$. Then the least squares estimator of unknown parameters $\theta^T = (\theta_1, \dots, \theta_q) \in \mathbf{R}^q$ constructed from observations (1) is defined as a random vector $\hat\theta_n^T = (\hat\theta_{1n}, \dots, \hat\theta_{qn})$ that minimizes the functional

$$Q(\theta) = (Y - X\theta)^T (Y - X\theta) = \sum_{j=1}^{n} \left(y_j - \sum_{i=1}^{q} \theta_i x_{ji} \right)^2, \tag{2}$$

2010 *Mathematics Subject Classification.* Primary 60G70, 62J05.

Key words and phrases. Regression model, maximal error, regression model adequacy test.

where $Y^T = (y_1, \dots, y_n)$. The least squares estimator is evaluated as follows:

$$\hat{\theta}_n = \left(X^T X\right)^{-1} X^T Y. \tag{3}$$

Put

$$\begin{gathered}\hat{y}_j = \sum_{i=1}^{q} \hat{\theta}_{in} x_{ji}, \qquad \hat{\varepsilon}_j = y_j - \hat{y}_j, \qquad j = 1, \dots, n, \\ Z_n = \max_{1\le j\le n} \varepsilon_j, \qquad \hat{Z}_n = \max_{1\le j\le n} \hat{\varepsilon}_j, \\ d_{in}^2 = \sum_{j=1}^{n} x_{ji}^2, \qquad d_n = \operatorname{diag}(d_{in}, i = 1, \dots, q).\end{gathered} \tag{4}$$

The following are two standard conditions in the regression analysis when studying the asymptotic properties of the least squares estimators:

(i)

$$d_{in}^{-1} \max_{1\le j\le n} |x_{ji}| \le k_i n^{-1/2}, \qquad i = 1, \dots, q. \tag{5}$$

For

$$J_n = d_n^{-1} X^T X d_n^{-1} = \left(d_{kn}^{-1} d_{ln}^{-1} \sum_{j=1}^{n} x_{jk} x_{jl} \right)_{k,l=1}^{q}, \tag{6}$$

let $\lambda_{\min}(A)$ denote the minimal eigenvalue of a positive definite matrix A:

(ii)

$$\lambda_{\min}(J_n) \ge \lambda_0 > 0 \quad \text{for } n > n_0. \tag{7}$$

Below we need some classical results of the theory of extreme values of random variables. Some of these results are listed below.

The asymptotic theory of extreme values studies the limit distribution of the random variable Z_n as $n \to \infty$ (recall the definition of Z_n in (4)). It turns out that the limit distribution of Z_n, normalized appropriately, necessarily belongs to one of the three families of distribution functions. It is worth mentioning that assumptions like $\mathsf{E}\,\varepsilon = 0$ or $\mathsf{E}\,\varepsilon^2 = \sigma^2 < \infty$ are useless in the general theory of extreme values.

Let random variables (ε_j) be independent and have the distribution function $F(x)$. Assume that

$$b_n(Z_n - a_n) \xrightarrow{\mathbf{D}} \zeta \tag{8}$$

as $n \to \infty$ for some constants $b_n > 0$ and a_n and that ζ has a nondegenerate distribution function $G(x) = \mathsf{P}(\zeta < x)$.

We say that the distribution function F belongs to the domain of max-attraction of the law G if relation (8) holds. In the latter case, we write $F \in D(G)$.

Theorem A (B. V. Gnedenko [3]). *If a distribution function F belongs to the domain of* max-*attraction of a law G, then G coincides with one of the members of the following three types of extreme value distribution functions:*

$$\begin{aligned} \textit{Type I}: \quad \Phi_\alpha(x) &= \begin{cases} 0, & \textit{for} \quad x \le 0, \\ \exp(-x^{-\alpha}), & \textit{for } \alpha > 0,\ x > 0; \end{cases} \\ \textit{Type II}: \quad \Psi_\alpha(x) &= \begin{cases} \exp(-(-x)^{\alpha}), & \textit{for } \alpha > 0,\ x \le 0, \\ 1, & \textit{for} \quad x > 0; \end{cases} \\ \textit{Type III}: \quad \Lambda(x) &= \exp(-e^{-x}), \quad \textit{for } -\infty < x < \infty. \end{aligned} \tag{9}$$

Conversely, every distribution function Φ_α, Ψ_α, *or* $\Lambda(x)$ *defined by* (9) *may appear as the limit in relation* (8) (*this happens, for example, if* $F(x)$ *equals either* Φ_α, Ψ_α, *or* $\Lambda(x)$).

Let $x_F = \sup\{x\colon F(x) < 1\}$. Below are simple sufficient conditions for a distribution function $F(x)$ to belong to the domains of the max-attraction $D(\Phi_\alpha)$, $D(\Psi_\alpha)$, or $D(\Lambda)$ of the laws Φ_α, Ψ_α, or $\Lambda(x)$, respectively (see [3]–[6]).

Theorem B (B. V. Gnedenko, R. von Mises)**.** *A distribution function F belongs to the domain of the* max-*attraction of a law of Type I if*

$$x_F = \infty \quad \text{and} \quad \forall x > 0 \quad \lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \quad \alpha > 0. \tag{10}$$

A distribution function F belongs to the domain of the max-*attraction of a law of Type II if*

$$x_F < \infty \quad \text{and} \quad \forall x > 0 \quad \lim_{h\downarrow 0} \frac{1 - F(x_F - xh)}{1 - F(x_F - h)} = x^{\alpha}, \quad \alpha > 0. \tag{11}$$

A distribution function F belongs to the domain of max-*attraction of a law of Type III if the derivative $F'(x) = f(x)$ exists, the derivative $f'(x)$ is negative in the interval* (x_0, x_F) (*whatever* $x_F \le \infty$), $f(x) = 0$ *for* $x > x_F$, *and*

$$\lim_{x\uparrow x_F} \frac{f'(x)(1 - F(x))}{f^2(x)} = -1. \tag{12}$$

Note that the sufficient conditions (10) and (11) in Theorem B are also necessary for the max-attraction to Types I and II, while (12) is close to a necessary condition for the max-attraction to Type III.

The constants a_n and b_n in relation (8) can be chosen as follows:

$$\text{Type I} \qquad (\Phi_\alpha): \qquad b_n = (\gamma_n)^{-1}, \qquad a_n = 0; \tag{13}$$

$$\text{Type II} \quad (\Psi_\alpha): \quad b_n = (x_F - \gamma_n)^{-1}, \qquad a_n = x_F; \tag{14}$$

If condition (12) holds, then

$$\text{Type III} \quad (\Lambda): \qquad a_n = \gamma_n, \qquad b_n = r(\gamma_n), \tag{15}$$

where

$$\gamma_n = F^{-1}\left(1 - n^{-1}\right) = \inf\left(x\colon F(x) \ge 1 - n^{-1}\right),$$

$F(x) = 1 - \exp(-R(x))$, and $R'(x) = r(x)$ (see [5, p. 40]).

Example 1. If (ε_j) are independent identically distributed standard Gaussian random variables, then condition (12) holds. If we choose

$$b_n = (2\ln n)^{1/2}, \qquad a_n = (2\ln n)^{1/2} - \frac{\ln\ln n + \ln(4\pi)}{2(2\ln n)^{1/2}} \qquad \text{for } n > 1, \tag{16}$$

then

$$\lim_{n\to\infty} \mathsf{P}\{b_n(Z_n - a_n) < x\} = \Lambda(x)$$

(see [4]–[6]).

Theorem 1. *Let a distribution function F of random variables ε_j in model* (1) *belong to the domain of the* max-*attraction of a law G. Assume that conditions* (i) *and* (ii) *hold. If*

$$b_n n^{-1/2} \to 0, \qquad n \to \infty, \tag{17}$$

then

$$\lim_{n\to\infty} \mathsf{P}\{b_n(\hat{Z}_n - a_n) < x\} = G(x), \qquad x \in \mathbf{R}^1, \tag{18}$$

where G is a member of one of the three types of extreme value distribution functions (9), *that is, G is either Φ_α, Ψ_α, or Λ.*

Proof of Theorem 1. Using the elementary inequality

$$\left| \max_{1\le j\le n} c_j - \max_{1\le j\le n} d_j \right| \le \max_{1\le j\le n} |c_j - d_j|,$$

we obtain

$$|\hat{Z}_n - Z_n| \le \max_{1\le j\le n} |\hat{\varepsilon}_j - \varepsilon_j| \le \sum_{i=1}^{q} |\hat{\theta}_{in} - \theta_i| \max_{1\le j\le n} |x_{ji}|,$$

whence

$$\Delta_n = \mathsf{E}\,|\hat{Z}_n - Z_n| \le \sum_{i=1}^{q} (\operatorname{Var} \hat{\theta}_{in})^{1/2} \max_{1\le j\le n} |x_{ji}|. \tag{19}$$

Condition (ii) implies that the inverse matrix $J_n^{-1} = \mathcal{L}_n = (\mathcal{L}_n^{kl})_{k,l=1}^q$ exists and that

$$\det J_n = \lambda_1 \dots \lambda_q \ge \lambda_{\min}^q \ge \lambda_0^q > 0,$$
$$(\det J_n)^{-1} \le \lambda_0^{-q}, \qquad n > n_0,$$

where $\lambda_1, \dots, \lambda_q$ are the eigenvalues of the matrix J_n.

Each of the cofactors for the matrix J_n is a product of some entries of the matrix J_n Every entry of the matrix J_n does not exceed 1 by the Cauchy–Bunyakovskiĭ inequality and there are $(q-1)!$ terms in the corresponding minor. Thus, condition (ii) implies that each entry of the matrix $\mathcal{L}_n$ is estimated from above by

$$|\mathcal{L}_n^{kl}| \le \lambda_0^{-q}(q-1)!, \qquad k, l = 1, \dots, q. \tag{20}$$

Since the covariance matrix of the normalized least squares estimator $d_n(\hat{\theta}_n - \theta)$ is given by

$$\mathsf{E}\, d_n(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^T d_n = \sigma^2 \left(d_n^{-1} X^T X d_n^{-1}\right)^{-1} = \sigma^2 \mathcal{L}_n,$$

the upper bound (20) implies that

$$\operatorname{Var} \hat{\theta}_{in} = \sigma^2 \mathcal{L}_n^{ii} d_{in}^{-2} \le \sigma^2 \lambda_0^{-q} (q-1)!\, d_{in}^{-2} \tag{21}$$

for $n > n_0$. Relations (19), (21) and condition (i) imply that

$$\Delta_n \le \sigma \lambda_0^{-q/2} \sqrt{(q-1)!} \left(\sum_{i=1}^{q} k_i \right) n^{-1/2}. \tag{22}$$

Inequality (22) and condition (17) yield

$$\mathsf{E}\,|b_n(\hat{Z}_n - a_n) - b_n(Z_n - a_n)| = b_n \Delta_n \to 0, \qquad n \to \infty. \qquad \square$$

Corollary 1. *Let conditions* (i) *and* (ii) *hold. Then*

a) *If a distribution function F satisfies condition* (10), *then equality* (18) *holds with the limit law $G(x) = \Phi_\alpha(x)$, where the constants a_n and b_n are defined by equality* (13);

b) *If a distribution function F satisfies condition* (11) *for $\alpha > 2$, then equality* (18) *holds with the limit law $G(x) = \Psi_\alpha(x)$, where the constants a_n and b_n are defined by equality* (14);

c) *If a distribution function F satisfies condition* (12), *then equality* (18) *holds with the limit law $G(x) = \Lambda(x)$, where the constants a_n and b_n are defined by equality* (15).

If we prove relation (17) for each of the cases a)–c), then Corollary 1 follows directly from Theorem 1.

Proof of (17) *in the case of* a). According to conditions (10) and (13), $x_F = \infty$ and $\gamma_n \to \infty$, $b_n = \gamma_n^{-1} \to 0$ as $n \to \infty$, that is, (17) holds.

Proof of (17) *in the case of* b). Condition (11) means that

$$1 - F(x_F - x) = x^\alpha L(x), \qquad x \downarrow 0,$$

where $L(x)$ is a slowly varying function at zero. Considering equality (14) we deduce

$$\gamma_n = x_F - u_n, \qquad u_n^\alpha L(u_n) = \frac{1}{n}.$$

We rewrite the latter equality as follows:

$$U_1\left(\frac{1}{u_n}\right) = n, \quad \text{where } U_1(x) = x^\alpha L_1(x),\ L_1(x) = \frac{1}{L(1/x)}.$$

It is clear that $L_1(x)$ slowly varies at infinity. It is known that there exists a function $L_2(x)$ that slowly varies at infinity and such that

$$U_1(U_2(x)) \sim U_2(U_1(x)) \sim x, \qquad x \to \infty,$$

where $U_2(x) = x^{1/\alpha} L_2(x)$ (see [7]). Then

$$n^{1/\alpha} L_2(n) = U_2(n) = U_2\left(U_1\left(\frac{1}{u_n}\right)\right) \sim \frac{1}{u_n}$$

as $n \to \infty$. Therefore,

$$b_n = (x_F - \gamma_n)^{-1} = \frac{1}{u_n} \sim n^{1/\alpha} L_2(n)$$

and thus condition (17) holds for $\alpha > 2$.

Proof of (17) *in the case of* c). This case follows from the following auxiliary result that may have its own value.

Lemma 1. *Assume that condition* (12) *holds. If the sequences a_n and b_n are defined by equalities* (15), *then, given $\delta > 0$, there exists a constant $C = C_\delta < \infty$ such that*

$$a_n \le C n^\delta, \qquad b_n \le C n^\delta. \tag{23}$$

Proof of Lemma 1. Let

$$R(x) = -\ln(1 - F(x)), \qquad r(x) = R'(x) = \frac{F'(x)}{1 - F(x)}.$$

This together with (12) implies

$$-\frac{r'}{r^2} = \left(\frac{1}{r}\right)' = \left(\frac{1-F}{F'}\right)' = -1 - \frac{(1-F)F''}{(F')^2} \to 0, \qquad x \to x_F. \tag{24}$$

Hence, for any $\delta > 0$, there exists $x_1 < x_F$ such that

$$\frac{r'(x)}{r(x)} \le \delta r(x)$$

for all $x \in (x_1, x_F)$ or, equivalently,

$$(\ln r(x))' \le \delta R'(x).$$

If $C = C_\delta$ is such that $\ln C > \ln r(x_1)$, then

$$\ln r(x) \le \delta R(x) + \ln C$$

for $x \in (x_1, x_F)$, whence

$$r(x) \le C \exp(\delta R(x)).$$

Substituting $x = R^{-1}(\ln n)$ in the latter inequality, we obtain $r(R^{-1}(\ln n)) \le Cn^\delta$. It remains to recall that $a_n = \gamma_n = R^{-1}(\ln n)$ and $b_n = r(a_n)$ in (15) are such that inequality in (23) holds.

To establish the first bound again we use relation (24). Given an arbitrary $\delta > 0$, there exists a number $x_2 < x_F$ such that

$$\frac{r'(x)}{r(x)} \ge -\delta r(x)$$

for all $x \in (x_2, x_F)$ or

$$(\ln r(x))' \ge -\delta R'(x).$$

If $\ln C \le \ln r(x_2)$, then

$$\ln r(x) \ge -\delta R(x) + \ln C$$

and

$$r(x) \ge C \exp(-\delta R(x)).$$

Substituting $x = R^{-1}(y)$ in the latter inequality we get

$$r(R^{-1}(y)) \ge C \exp(-\delta y).$$

This together with the equality $(R^{-1}(y))' = (r(R^{-1}(y)))^{-1}$ implies that

$$(R^{-1}(y))' \le C' \exp(\delta y)$$

or

$$R^{-1}(y) \le (C'/\delta) \exp(\delta y) + C''$$

if $C'' \ge R^{-1}(y_2) = x_2$. Let $y = \ln n$. Then

$$a_n = R^{-1}(\ln n) \le (C'/\delta) n^\delta + C''.$$

It is clear that this upper bound implies the first inequality in (23). □

Example 1 (continuation)**.** If (ε_j) in model (1) are independent identically distributed standard Gaussian random variables and if conditions (i) and (ii) hold, then

$$\lim_{n\to\infty} \mathsf{P}\left\{ b_n \left(\frac{\hat{Z}_n}{\sigma} - a_n \right) < x \right\} = \Lambda(x), \qquad x \in \mathbf{R}^1,$$

where the normalizing sequences a_n and b_n are defined by (16).

Example 2. Let (ε_j) be uniformly distributed random variables in the interval $[-1, 1]$. Condition (11) holds in this case with $\alpha = 1$. Then (14) implies that $a_n = 1$ and $b_n = \frac{n}{2}$. Thus,

$$\lim_{n\to\infty} \mathsf{P}\left\{ \frac{n}{2}(Z_n - 1) < x \right\} = \Psi_1(x), \qquad x \in \mathbf{R}^1.$$

Since $b_n n^{-1/2} \nrightarrow 0$, the assumptions of Theorem 1 do not hold. We show that equality (18) also does not hold in this case.

Consider the simplest case of model (1), namely

$$y_j = \theta + \varepsilon_j, \qquad j = 1, \dots, n.$$

Then the least squares estimator equals $\hat{\theta} = \bar{y} = \theta + \bar{\varepsilon}$ and

$$\hat{y}_j = \bar{y}, \qquad \hat{\varepsilon}_j = y_j - \hat{y}_j = \varepsilon_j - \bar{\varepsilon}, \qquad \hat{Z}_n = \max_{1\le j\le n} \varepsilon_j - \bar{\varepsilon} = Z_n - \bar{\varepsilon}.$$

For the case under consideration,

$$\frac{n}{2}(\hat{Z}_n - 1) = \frac{n}{2}(Z_n - 1) - \frac{1}{2}\sum_{j=1}^{n} \varepsilon_j.$$

Since $Z_n < 1$ almost surely, we have

$$\begin{aligned} \mathsf{P}\left\{\frac{n}{2}(\hat{Z}_n - 1) > x\right\} &\ge \mathsf{P}\left\{-\frac{1}{2}\left(\frac{2n}{3}\right)^{1/2}\left(\frac{3}{2n}\right)^{1/2}\sum_{j=1}^{n}\varepsilon_j > x\right\} \\ &\sim 1 - \Phi\left(\left(\frac{6}{n}\right)^{1/2} x\right) \sim \frac{1}{2} \end{aligned}$$

for $x > 0$ and sufficiently large n, where Φ is the standard Gaussian distribution function. This means that the distribution of $\frac{n}{2}(\hat{Z}_n - 1)$ does not converge to Ψ_1, since $1 - \Psi_1(x) = 0$ for $x > 0$.

Example 3. Assume that random variables ε_j have the student distribution with s degrees of freedom, that is,

$$F_s(x) = K_s \int_{-\infty}^{x} \left(1 + \frac{u^2}{s}\right)^{-(s+1)/2} du, \qquad K_s = \frac{\Gamma((s+1)/2)}{\sqrt{\pi s}\,\Gamma(s/2)}.$$

Applying the l'Hospital rule for $x > 0$ we get

$$\lim_{t\to\infty} \frac{1 - F_s(tx)}{1 - F_s(t)} = x^{-s}.$$

Hence, according to (10), the distribution function F_s belongs to the domain of the max-attraction of the distribution Φ_s. Moreover, Theorem 1 holds if conditions (i) and (ii) are satisfied.

2. Nonlinear regression model

Consider the following nonlinear regression model

$$y_j = g(j, \theta) + \varepsilon_j, \qquad j = 1, \dots, n, \tag{25}$$

where the sequence of functions $g(j, \cdot)$ is defined in some open set $\mathcal{O} \subset \mathbf{R}^q$ such that $\Theta^c \subset \mathcal{O}$, where Θ is a bounded open convex set that contains the true (unknown) value of the parameter θ. We assume that the functions $g(j, \cdot)$ are twice continuously differentiable in $\mathcal{O}$. We also assume that the errors of observations ε_j, $j = 1, \dots, n$, are independent identically distributed random variables such that $\mathsf{E}\,\varepsilon_j = 0$.

Every vector $\hat{\theta}_n \in \Theta^c$ such that

$$Q(\hat{\theta}_n) = \inf_{\tau\in\Theta^c} Q(\tau), \qquad Q(\tau) = \sum_{j=1}^{n} (y_j - g(j, \tau))^2$$

is called the least squares estimator of the parameter θ constructed from the observations (25). In contrast to the case of linear models, we cannot write the least squares estimator $\hat{\theta}_n$ in an explicit form.

Introduce some notation. Let $\alpha = (\alpha_1, \dots, \alpha_q)$ be a multi-index,

$$|\alpha| = \alpha_1 + \dots + \alpha_q, \qquad d_n^\alpha(\theta) = d_{1n}^{\alpha_1}(\theta) \dots d_{qn}^{\alpha_q}(\theta),$$

where

$$d_{in}^2(\theta) = \sum_{j=1}^n g_i^2(j, \theta), \qquad g_i = \frac{\partial}{\partial \theta_i} g$$

for $i = 1, \dots, q$. Below is a somewhat more general notation:

$$d_n(\alpha, \theta) = \sum_{j=1}^n g^{(\alpha)}(j, \theta)^2, \qquad g^{(\alpha)} = \frac{\partial^{|\alpha|}}{(\partial \theta_1)^{\alpha_1} \dots (\partial \theta_q)^{\alpha_q}} g.$$

If $e_i \in \mathbf{R}^q$ is the i-th unit coordinate vector, then $d_n(e_i, \theta) = d_{in}(\theta)$, $i = 1, \dots, q$.

We consider the case where

$$0 < \liminf_{n \to \infty} n^{-1/2} d_{in}(\theta), \qquad i = 1, \dots, q. \tag{26}$$

Put $d_n(\theta) = \operatorname{diag}(d_{in}(\theta), i = 1, \dots, q)$,

$$f^{(\alpha)}(j, u) = g^{(\alpha)}\left(j, \theta + n^{1/2} d_n^{-1}(\theta) u\right), \qquad U_n^c(\theta) = n^{-1/2} d_n(\theta)(\Theta^c - \theta),$$

$$\Phi_n^{(\alpha)}(u_1, u_2) = \sum_{j=1}^n \left(f^{(\alpha)}(j, u_1) - f^{(\alpha)}(j, u_2)\right)^2, \quad u_1, u_2 \in U_n^c(\theta),$$

$$J_n(\theta) = \left(d_{kn}^{-1}(\theta) d_{ln}^{-1}(\theta) \sum_{j=1}^n g_k(j, \theta) g_l(j, \theta)\right)_{k,l=1}^q,$$

$$\Lambda_n(\theta) = \left(\Lambda_n^{kl}(\theta)\right)_{k,l=1}^q = J_n^{-1}(\theta).$$

We also put

$$S(r) = \{u \in \mathbf{R}^q \colon \|u\| < r\}, \quad r > 0; \qquad u_n(\theta) = n^{-1/2} d_n(\theta)(\hat{\theta}_n - \theta).$$

To prove an assertion similar to Theorem 1 for nonlinear regression models, one needs a stronger property than just the consistence of the least squares estimator $\hat{\theta}_n$, since the upper bounds for moments of the estimator $\hat{\theta}_n$ in a nonlinear model require more restrictions imposed on the regression function g than those needed in the poof of the consistency [2].

Below are the conditions under which the least squares estimator $\hat{\theta}_n$ possesses the property mentioned above.

$\mathcal{A}_1$. $\mu_m = \mathsf{E}\,|\varepsilon_j|^m < \infty$ for some integer $m \geq 3$.

$\mathcal{A}_2$. For all $r > 0$, there are constants $C_i(\alpha, r)$, $i = 1, 2, 3$, such that

$$\sup_{u \in S^c(r) \cap U_n^c(\theta)} n^{(|\alpha| - 1)/2} (d_n^\alpha(\theta))^{-1} d_n\left(\alpha, \theta + n^{1/2} d_n^{-1}(\theta) u\right) \leq C_1(\alpha, r), \quad |\alpha| = 1, 2; \tag{27}$$

$$\sup_{u_1, u_2 \in S^c(r) \cap U_n^c(\theta)} n (d_n^\alpha(\theta))^{-2} \Phi_n^{(\alpha)}(u_1, u_2) \|u_1 - u_2\|^{-2} \leq C_2(\alpha, r), \qquad |\alpha| = 2; \tag{28}$$

$$\sup_{u \in S^c(r) \cap U_n^c(\theta)} n^{1/2} (d_n(\alpha, \theta))^{-1} \max_{1 \leq j \leq n} \left|f^{(\alpha)}(j, u)\right| \leq C_3(\alpha, r), \qquad |\alpha| = 1, 2. \tag{29}$$

$\mathcal{A}_3$. If $g^{(\alpha)}(j, \theta) \neq 0$, then

$$\liminf_{n \to \infty} n^{1/2} (d_n^\alpha(\theta))^{-1} d_n(\alpha, \theta) > 0, \qquad |\alpha| = 2. \tag{30}$$

$\mathcal{A}_4$. $\lambda_{\min}(J_n(\theta)) \geq \lambda_0 > 0$ for $n > n_0$.

$\mathcal{A}_5$. The number m involved in condition $\mathcal{A}_1$ is such that

$$\mathsf{P}\{\|u_n(\theta)\| \geq r)\} = O\left(\frac{1}{n^{(m-2)/2}}\right) \tag{31}$$

for all $r > 0$.

Condition (29) generalizes (i), while condition $\mathcal{A}_4$ is similar to condition (ii) introduced in Section 1. Relation (31) which generalizes a result due to Malinvaud [8] is stronger than the property of the weak consistence of the least squares estimator $\hat{\theta}_n$. Sufficient conditions for (31) and the corresponding proof are given in Theorem 8 of [2, pp. 30–32].

The following result is a generalization of (31); it is, in fact, Theorem 19 of [2, pp. 91] stated in a form convenient for the purpose of this paper. Put

$$\chi(\delta) = 2\lambda_0^{-1}\sigma q(m-2+\delta)^{1/2},$$

where $\sigma^2 = \mu_2$.

Theorem 2. *If conditions $\mathcal{A}_1$–$\mathcal{A}_5$ hold, then*

$$\mathsf{P}\left\{\|u_n(\theta)\| \geq \chi(\delta) n^{-1/2} \ln^{1/2} n\right\} = O\left(\frac{1}{n^{(m-2)/2}}\right) \tag{32}$$

for all fixed $\delta > 0$.

One can rewrite relation (32) as follows:

$$\mathsf{P}\left\{\left\|d_n(\theta)(\hat{\theta}_n - \theta)\right\| \geq \chi(\delta) \ln^{1/2} n\right\} = O\left(\frac{1}{n^{(m-2)/2}}\right). \tag{33}$$

Thus, Theorem 2 can be viewed as a result on moderate deviations for the normalized least squares estimator $d_n(\theta)(\hat{\theta}_n - \theta)$.

Put $\hat{y}_j = g(j, \hat{\theta}_n)$. We are ready to state the main result of this section by using the notation introduced in Section 1.

Theorem 3. *Let the distribution function F of random variables ε_j in model* (25) *belong to the domain of the* max-*attraction of a law G. Assume that conditions $\mathcal{A}_1$–$\mathcal{A}_5$ hold and*

$$\left(\frac{\ln n}{n}\right)^{1/2} b_n \to 0, \qquad n \to \infty. \tag{34}$$

Then relation (18) *holds, where the limit law G equals one of the three extreme distribution functions Φ_α, Ψ_α, or Λ.*

Proof of Theorem 3. As above,

$$|\hat{Z}_n - Z_n| \leq \max_{1\leq j\leq n} |\hat{\varepsilon}_j - \varepsilon_j| = \max_{1\leq j\leq n} \left|g(j, \hat{\theta}_n) - g(j, \theta)\right|.$$

On the other hand,

$$g(j, \hat{\theta}_n) - g(j, \theta) = f(j, u_n(\theta)) - f(j, 0) = \sum_{i=1}^{q} f_i(j, u_n^*(\theta)) n^{1/2} d_{in}^{-1}(\theta) u_{in}(\theta),$$

where $\|u_n^*(\theta)\| < \|u_n(\theta)\|$. If $\|u_n(\theta)\| < 1$, then condition (29) implies

$$\begin{aligned}|\hat{Z}_n - Z_n| &\leq \sum_{i=1}^{q} n^{1/2} d_{in}^{-1}(\theta) \max_{1\leq j\leq n} |f_i(j, u_n^*(\theta))| \cdot |u_{in}(\theta)| \\ &\leq \sum_{i=1}^{q} C_3(e_i, 1) |u_{in}(\theta)| \leq \|\mathbb{C}\| \cdot \|u_n(\theta)\|,\end{aligned}$$

where $\mathbb{C} = (C_3(e_1, 1)), \ldots, C_3(e_q, 1))$.

Now

$$\mathsf{P}\left\{b_n|\hat{Z}_n - Z_n| \geq \varepsilon\right\} \leq \mathsf{P}\left\{b_n\|u_n(\theta)\| \geq \frac{\varepsilon}{\|\mathbb{C}\|}\right\} + \mathsf{P}\{\|u_n(\theta)\| \geq 1\} = \mathcal{P}_1 + \mathcal{P}_2$$

for an arbitrary $\varepsilon > 0$, where $\mathcal{P}_2 \to 0$ in view of $\mathcal{A}_5$. Now we estimate the first term on the right-hand side of the latter bound:

$$\begin{aligned}\mathcal{P}_1 &= \mathsf{P}\left\{b_n\|u_n(\theta)\| \geq \frac{\varepsilon}{\|\mathbb{C}\|}, \|u_n(\theta)\| \geq \chi(\delta)n^{-1/2}\ln^{1/2} n)\right\} \\ &\quad + \mathsf{P}\left\{b_n\|u_n(\theta)\| \geq \frac{\varepsilon}{\|\mathbb{C}\|}, \|u_n(\theta)\| < \chi(\delta)n^{-1/2}\ln^{1/2} n)\right\} \\ &= \mathcal{P}_3 + \mathcal{P}_4.\end{aligned}$$

According to condition (34), $\mathcal{P}_4 = 0$ starting with some number n, while

$$\mathcal{P}_3 = o\left(n^{-(m-2)/2}\right)$$

by Theorem 2. □

It is obvious that the presence of the factor $\ln^{1/2} n$ in condition (34) does not influence the proof of Corollary 1. Thus the following result is a corollary to Theorem 3.

Corollary 2. *Let conditions* $\mathcal{A}_1$*–*$\mathcal{A}_5$ *hold for model* (25)*. Then statements a), b), and c) of Corollary* 1 *are valid.*

In particular, if

$$g(j,\theta) \equiv g(x_j,\theta), \qquad x_j = (x_{j1},\dots,x_{jl}) \in \mathbb{X} \subset \mathbf{R}^l, \qquad j = 1,\dots,n,$$

for model (25), where $\mathbb{X}$ is a certain domain of regression experiment design, then model (25) is a generalization of model (1).

Assume that $\mathbb{X}$ is a compact set. If all derivatives of the regression function $g(x,\theta)$ (with respect to its parameters θ) are continuous with respect to the set of the variables $(x,\theta) \in \mathbb{X} \times \Theta^c$, then

$$\sup_{\theta\in\Theta^c} d_n(\alpha,\theta) \leq \bar{c}(\alpha)n^{1/2}, \qquad |\alpha| = 1,2, \tag{35}$$

where $\bar{c}(\alpha)$ are some constants. We restrict the consideration to the case where

$$d_{in}(\theta) \geq \underline{c}_i n^{1/2} \tag{36}$$

for some $\underline{c}_i > 0$, $i = 1,\dots,q$ (see (26)). If $|\alpha| = 2$ and $g^{(\alpha)}(x,\theta) \neq 0$ for all arguments, then we assume that

$$d(\alpha,\theta) \geq \underline{c}(\alpha)n^{1/2} \tag{37}$$

for some $\underline{c}(\alpha) > 0$.

If conditions (35) and (36) hold, then, without loss of generality, one can use the normalization $n^{1/2}\mathbb{I}_q$ instead of $d_n(\theta)$. The assumptions of Theorem 2 are also simplified in this case. For example, conditions (27), (29), and (30) follow from (35)–(37). Condition (28) becomes of the form

$$n^{-1}\sum_{j=1}^{n}\left(g^{(\alpha)}(x_j,\theta_1) - g^{(\alpha)}(x_j,\theta_2)\right)^2 \|\theta_1 - \theta_2\|^{-2} \leq c(\alpha),$$

θ_1, $\theta_2 \in \Theta^c$, $|\alpha| = 2$.

3. Some applications. A regression model adequacy test

Testing the goodness of fit for regression models is often needed in many practical problems. We consider this problem for a simple linear regression model by using the limit results obtained above for the maximal error term.

Let

$$y_j = \theta_0 + \theta_1 x_j + \varepsilon_j, \qquad j = 1, \dots, n, \tag{38}$$

where ε_j are independent identically distributed standard Gaussian random variables. Let $\hat\theta_0$ and $\hat\theta_1$ be the least squares estimators for the unknown parameters θ_0 and θ_1, respectively,

$$\hat y_j = \hat\theta_0 + \hat\theta_1 x_j, \qquad \hat\varepsilon_j = y_j - \hat y_j,$$

$$\hat Z_n = \max_{1\le j\le n} \hat\varepsilon_j, \qquad \bar x = \frac{1}{n}\sum_{j=1}^n x_j, \qquad s^2 = \frac{1}{n}\sum_{j=1}^n (x_j - \bar x)^2.$$

In what follows we assume that the sequences a_n and b_n are defined by relation (16). Similarly to Theorem 1, one can prove

$$\lim_{n\to\infty} \mathsf{P}\left\{ b_n\left(\frac{\hat Z_n}{\sigma} - a_n\right) < x\right\} = \Lambda(x), \qquad x \in \mathbf{R}^1, \tag{39}$$

under the condition that

$$\frac{\ln n}{n s_n^2} \max_{1\le j\le n} |x_j|^2 \to 0, \qquad n\to\infty. \tag{40}$$

The latter condition is weaker than assumptions of Theorem 1.

For our purposes, it is necessary to use a stronger assumption than limit equality (39).

Theorem 4. *If condition* (40) *holds for model* (38)*, then*

$$\lim_{n\to\infty} \mathsf{P}\left\{ b_n\left(\frac{\hat Z_n^*}{\hat\sigma_n} - a_n\right) < x\right\} = \Lambda^2(x), \qquad x \in \mathbf{R}^1, \tag{41}$$

where

$$\hat Z_n^* = \max_{1\le j\le n} |\hat\varepsilon_j|, \qquad \hat\sigma_n^2 = (n-2)^{-1}\sum_{j=1}^n \hat\varepsilon_j^2.$$

Proof of Theorem 4. First we prove the relation

$$\lim_{n\to\infty} \mathsf{P}\left\{ b_n\left(\frac{\hat Z_n^*}{\sigma} - a_n\right) < x\right\} = \Lambda^2(x). \tag{42}$$

Without loss of generality, we assume that $\sigma = 1$.

Put

$$Z_n^* = \max_{1\le j\le n} |\varepsilon_j|, \qquad W_n = \min_{1\le j\le n} \varepsilon_j.$$

Then

$$\mathsf{P}\{b_n(Z_n^* - a_n) < x\} = \mathsf{P}\{Z_n^* < u_n\} = \mathsf{P}\{Z_n < u_n, W_n > -u_n\} \to \Lambda^2(x)$$

for $u_n = x/b_n + a_n$ as $n\to\infty$. The latter asymptotic relation follows from Theorem 1.8.2 in [6], since

$$n\Phi(-u_n) = n(1-\Phi(u_n)) \to \exp(-x) = -\ln(\Lambda(x))$$

for a standard Gaussian distribution function $\Phi(x)$ (see [6]).

Now equality (42) follows from the inequality

$$|Z_n^* - \hat{Z}_n^*| \le \max_{1\le j\le n} |\varepsilon_j - \hat{\varepsilon}_j|$$

by follwing the lines of the proof of Theorem 1.

It remains to prove that (42) $\Rightarrow$ (41). Put

$$\zeta_n = b_n\left(\frac{\hat{Z}_n^*}{\sigma} - a_n\right), \qquad \zeta_n' = b_n\left(\frac{\hat{Z}_n^*}{\hat{\sigma}_n} - a_n\right).$$

According to equality (42),

$$\zeta_n \xrightarrow{\mathrm{Var}} \zeta, \qquad n\to\infty, \tag{43}$$

where $\mathsf{P}\{\zeta < x\} = \Lambda^2(x)$.

The definition of the random variables ζ_n and ζ_n' implies that

$$\zeta_n - \zeta_n' = b_n\frac{\hat{Z}_n^*}{\sigma}\left(\frac{\hat{\sigma}_n - \sigma}{\hat{\sigma}_n}\right) = \frac{(\zeta_n + a_n b_n)(\hat{\sigma}_n^2 - \sigma^2)}{\hat{\sigma}_n(\hat{\sigma}_n + \sigma)}. \tag{44}$$

Then we use the known estimate

$$\mathsf{E}\left|\hat{\sigma}_n^2 - \sigma^2\right| \le \sqrt{\operatorname{Var}\hat{\sigma}_n^2} = O\left(\frac{1}{\sqrt{n}}\right) \tag{45}$$

(see, for example, [2]). Thus $\hat{\sigma}_n^2 \xrightarrow{\mathsf{P}} \sigma^2$, whence

$$\zeta_n\left(\hat{\sigma}_n^2 - \sigma^2\right) \xrightarrow{\mathsf{P}} 0, \qquad n\to\infty,$$

in view of (43).

Similarly, relations (16) and (45) imply that

$$\mathsf{E}\, a_n b_n\left|\hat{\sigma}_n^2 - \sigma^2\right| = a_n b_n O\left(\frac{1}{\sqrt{n}}\right) \to 0, \qquad n\to\infty.$$

These relations together with (44) yield

$$\zeta_n - \zeta_n' \xrightarrow{\mathsf{P}} 0, \qquad n\to\infty.$$

Now equality (41) follows from (42). $\square$

Theorem 4 can easily be generalized to the case of multiple linear Gaussian regression (the proof follows the lines of that of Theorem 1).

Asymptotic equalities (41) and (42) can be helpful to construct a test of hypothesis

$$H_0 = \left\{y = \theta_0 + \theta_1 x + \varepsilon, \varepsilon\sim N\left(0,\sigma^2\right)\right\}$$

against an appropriate alternative H_1. The test statistic can be chosen as

$$T_n = b_n\left(\frac{\hat{Z}_n^*}{\hat{\sigma}_n} - a_n\right)$$

if σ is unknown, or

$$T_n' = b_n\left(\frac{\hat{Z}_n^*}{\sigma} - a_n\right)$$

if σ is known. If the null hypothesis H_0 is true, then the distributions of the statistics T_n and T_n' are approximately equal to $\Lambda^2(x)$ for large n.

Let the level of significance for this test be denoted by α. Then the critical region is given by

$$K = \left(-\infty, -\ln\ln\sqrt{\frac{2}{\alpha}}\right) \cup \left(-\ln\ln\sqrt{\frac{2}{2-\alpha}}, \quad +\infty\right).$$

If, for example, $\alpha = 0.05$, then

$$K = (-\infty, -0.61218) \cup (4.36939, +\infty)\,.$$

Below are two possible choices of the alternative hypothesis H_1 for the test based on the statistics T_n and $T_n^{'}$, respectively.

(i) $H_1 = \{y = \theta_0 + \theta_1 x + \varepsilon$, where the tails of the distribution of ε are different from the Gaussian ones$\}$.

(ii) $H_1 = \{y = g(x,\theta) + \varepsilon$, where the distribution of ε is Gaussian with parameters 0 and $\sigma^2\}$; here

$$g(x,\theta) = \sum_{k=0}^{m} \theta_k x^k$$

is a polynomial of order $m \geq 2$.

If the hypothesis H_1 is true in the second case, then the asymptotic behavior of T_n is not known yet in the most interesting case where σ^2 is unknown. Further investigation is needed to find the precise statement for the test.

Bibliography

1. G. A. F. Seber, *Linear Regression Analysis*, Wiley, New York, 1977. MR0436482 (55:9428)
2. A. V. Ivanov, *Asymptotic Theory of Nonlinear Regression*, Kluwer Academic Press, Dordrecht, 1997. MR1472234 (99h:62086)
3. B. V. Gnedenko, *Sur la distribution limit du terme maximum d'une serie aleatoire*, Ann. Math. **44** (1943), 423–453. MR0008655 (5:41b)
4. J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, John Wiley & Sons, New York–Chichester–Brisbane, 1978. MR489334 (80b:60040)
5. S. I. Resnick, *Extreme Values, Regular Variation and Point Processes*, Springer, Berlin, 1987. MR900810 (89b:60241)
6. M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York–Heidelberg–Berlin, 1983. MR691492 (84h:60050)
7. E. Seneta, *Regularly Varying Functions*, Springer-Verlag, Berlin–Heidelberg–New York, 1976. MR0453936 (56:12189)
8. E. Malinvaud, *The Consistency of Nonlinear Regression*, Ann. Math. Statist. **41** (1970), 953–969. MR0261754 (41:6367)

National Technical University of Ukraine "Kyiv Polytechnic Institute", Peremogy Avenue, 37, Kyiv 03056, Ukraine

E-mail address: alexntuu@gmail.com

Kyiv National Taras Shevchenko University, Glushkov Avenue 2,Building 6, Kyiv 03127, Ukraine

E-mail address: mik@unicyb.kiev.ua

Received 09/OCT/2011

Translated by S. V. KVASKO