Goodness-of-fit problem for errors in non-parametric regression: distribution free approach

Estate Khmaladze and Hira L. Koul[1]
Victoria University of Wellington and Michigan State University

Abstract

This paper discusses asymptotically distribution free tests for the classical goodness-of-fit hypothesis of an error distribution in nonparametric regression models. These tests are based on the same martingale transform of the residual empirical process as used in the one sample location model. The residuals used are those obtained from estimating the regression function by the local linear polynomial method. The results of this paper are made feasible by a recent result of Müller, Schick and Wefelmeyer that establishes an asymptotic uniform linearity of the nonparametric residual empirical process at the rate $n^{-1/2}$.

1 Introduction

Consider a sequence of i.i.d. pairs of random variables $\{(X_i, Y_i)_{i=1}^n\}$ where X_i are d-dimensional covariates and Y_i are the one dimensional responses. Suppose Y_i has regression in mean on X_i, i.e., there is a regression function $m(\cdot)$ and a sequence of i.i.d. innovations $\{e_i, 1 \leq i \leq n\}$ such that

$$Y_i = m(X_i) + e_i,\ i = 1, \ldots, n.$$

This regression function, as in most applications, is generally unknown and we do not make assumptions about its possible parametric form, so that we need to use a non-parametric estimator $\hat{m}_n(\cdot)$ of $m(\cdot)$ based on $\{(X_i, Y_i)_{i=1}^n\}$.

The problem of interest here is to test the hypothesis that the common distribution function of e_i is a given F. Since $m(\cdot)$ is unknown we can only use residuals

$$\hat{e}_i = Y_i - \hat{m}_n(X_i),\ i = 1, \ldots, n,$$

which, obviously, are not i.i.d. any more. Let F_n and $\hat{F}_n$ denote the empirical distribution functions of the errors e_i, $1 \leq i \leq n$, and the residuals $\hat{e}_i$, $1 \leq i \leq n$, respectively, and let

$$v_n(x) := \sqrt{n}[F_n(x) - F(x)], \quad \hat{v}_n(x) := \sqrt{n}[\hat{F}_n(x) - F(x)], \quad x \in \mathbb{R}$$

[1]Research of this author partly supported by the NSF grant DMS-O704130.

denote empirical and "estimated" empirical processes.

Akritas and Van Keilegom (2001) and Müller, Schick and Wefelmayer (2006) established, under the null hypothesis and some assumptions, the following uniform asymptotic expansion of $\hat{v}_n$:

$$\hat{v}_n(x) = v_n(x) - f(x)\, R_n + \xi_n(x), \quad R_n = O_p(1), \quad \sup_x |\xi_n(x)| = o_p(1). \tag{1.1}$$

Basically, the term R_n is made up by the sum

$$R_n = n^{-1/2} \sum_{i=1}^{n} [\hat{m}_n(X_i) - m_n(X_i)],$$

but using special form of the estimator $\hat{m}_n$, Müller, Schick and Wefelmayer (2006) obtained especially simple form for it:

$$R_n = n^{-1/2} \sum_{i=1}^{n} \varepsilon_i.$$

In the case of parametric regression where the regression function is of the parametric form, $m(\cdot) = m(\cdot, \theta)$, and the unknown parameter θ is replaced by its estimator $\hat{\theta}_n$, similar asymptotic expansion have been established in Loyns (1980), Koul (2002), and Khmaladze and Koul (2004). However, the non-parametric case is more complex and it is remarkable that the asymptotic expansion (1.1) is still true.

The above expansion (1.1) leads to the central limit theorem for the process $\hat{v}_n$, and, hence, produces the null limit distribution for test statistics based on this process. However, the same expansion makes it clear that the statistical inference based on $\hat{v}_n$ is inconvenient in practice and even infeasible: not only does the limit distribution of $\hat{v}_n$ after time transformation $t = F(x)$ still depend on the hypothetical distribution F, but it depends also on the estimator $\hat{m}_n$, (and, in general, on the regression function m itself), that is, it is different for different estimators. Since goodness-of-fit statistics are essentially non-linear functionals of the underlying process with difficult to calculate limit distributions, it is practically inconvenient to be obliged to do substantial computational work to evaluate their null distributions every time we test the hypothesis. Note, in particular, that if we try to use some kind of bootstrap simulations, we would have to compute the non-parametric estimator $\hat{m}_n$ for every simulated sub-sample, which makes it especially time consuming.

The goal of this paper is to show that this complication can be avoided in the way, which is technically surprisingly simple. Namely, we present the transformed process w_n, which, after time transformation $t = F(x)$, converges in distribution to a standard Brownian motion, for any estimator $\hat{m}_n$ for which (1.1) is valid. One would expect that this is done at the cost of some power. We will see, however, somewhat unexpectedly, that tests based on this transformed process w_n should, typically, have better power than those based on $\hat{v}_n$.

2 Transformed Process

Let the error d.f. F have finite Fisher information for location, i.e., let $\psi_f = -\dot{f}/f$ denote the score function for location family $F(\cdot - \theta), \theta \in \mathbb{R}$ at $\theta = 0$ – we can assume that $\theta = 0$ without loss of generality. Then

$$\int \psi_f(x)^2 dF(x) < \infty. \tag{2.1}$$

Consider augmented score function

$$h(x) = \begin{pmatrix} 1 \\ \psi_f(x) \end{pmatrix},$$

and augmented incomplete information matrix

$$\Gamma_{F(x)} = \int_x^\infty h(x) h^T(x) dF(x) = \begin{pmatrix} 1 - F(x) & f(x) \\ f(x) & \sigma_f^2(x) \end{pmatrix}, \qquad x \in \mathbb{R}, \tag{2.2}$$

with $\sigma_f^2(x) = \int_x^\infty \psi_f^2(y) dF(y)$.

For any signed measure ν for which the following integral is well defined, let

$$K(x, \nu) \;=\; \int_{-\infty}^x h^T(y) \Gamma_{F(y)}^{-1} \int_y^\infty h(z) d\nu(z) dF(y), \quad x \in \mathbb{R}.$$

Our process w_n is defined as

$$w_n(x) \;=\; \sqrt{n}[\hat{F}_n(x) - K(x, \hat{F}_n)], \quad x \in \mathbb{R}. \tag{2.3}$$

We shall show that w_n converges in distribution to the Brownian motion w in time F, that is, to Gaussian process with mean 0 and covariance function $Ew(x)w(x') = F(\min(x, x'))$. In other words, we will show that time transformed process $b_n(t) = w_n(x)$, with $t = F(x)$, converges in distribution to standard Brownian motion on the interval $[0, 1]$.

To begin with observe that the process w_n can be rewritten as

$$w_n(x) = \sqrt{n}[\hat{v}_n(x) - K(x, \hat{v}_n)]. \tag{2.4}$$

Indeed, $F(x)$ is the first coordinate of the vector-function $H(x) = \int_{-\infty}^x h(y) dF(y) = (F(x), -f(x))^T$, and we will see that

$$H^T(x) - K(x, H^T) = 0, \qquad \forall\, x \in \mathbb{R}. \tag{2.5}$$

Subtracting this identity from (2.3) yields (2.4). Using asymptotic expansion (1.1) we can rewrite

$$(2.6)\qquad w_n(x) = v_n(x) - K(x, v_n) + \eta_n(x), \quad \eta_n(x) = \xi_n(x) - K(x, \xi_n),$$

where one expects η_n to be "small" (see Sec. 4), and the main part on the right does not contain the term $f(F^{-1}(t))R_n$ of that expansion. This is true again because of (2.5) and the fact that the second coordinate of $H(x)$ is $-f(x)$.

The transformation and the process b_n is very similar to the one studied in *Khmaladze, Koul (2004)*. However, asymptotic behavior of the empirical distribution function $\hat{F}_n$ here is more complicated. As a result, we have to prove the smallness of the "residual process" η_n in (2.6) differently - see Sec.4. Besides, here we explicitly consider the case of possibly degenerate matrix $\Gamma_{F(x)}$ and show that w_n and b_n are still well defined - see Lemma 2.1. Also in this section, we demonstrate that although, in this transformation, singularity at $t = 1$ exists, the process b_n converges to its weak limit on the closed interval $[0, 1]$ - see Theorems 2.2 and 4.1, *(ii)*.

Now we shall show that (2.5) holds and the process w_n is well defined even if $\Gamma_{F(x)}$ is not of full rank and the inverse $\Gamma^{-1}_{F(x)}$ is not unique.

If $\Gamma_{F(x)}$ is of the full rank, then (2.5) is obvious. For most distribution functions F, the matrix $\Gamma_{F(x)}$ indeed is not degenerate, that is, the coordinates 1 and ψ_f of h are linearly independent functions on tail set $\{x > x_0\}$ for every $x_0 \in \mathbb{R}$. However, if for x greater than some x_0, the density f has the form $f(x) = \alpha e^{-\alpha x}$, $\alpha > 0$, the function $\psi_f(x)$ equals the constant α so that 1 and $\psi_f(x)$ become linearly dependent for $x > x_0$. In this case

$$(2.7)\qquad \Gamma_{F(x)} = (1 - F(x)) \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}, \quad x > x_0.$$

Conversely, one can prove that if (2.7) holds for some $x_0 \in \mathbb{R}$, then for some $\alpha > 0$, $f(x) = \alpha e^{-\alpha x}$, $x > x_0$.

The lemma below shows, that although in this case $\Gamma^{-1}_{F(x)}$ can not be uniquely defined, the function $h^T(x)\Gamma^{-1}_{F(x)} \int_x^\infty h(y) dv_n(y)$ is well defined. Here it is more transparent and simple to use also time transformation $t = F(x)$. Accordingly, let $u_n(t) = v_n(F^{-1}(t))$, $\gamma(t) := h(F^{-1}(t))$, and $\Gamma_t = \int_t^1 \gamma(s)\gamma(s)^T ds$, $0 \le t \le 1$.

Lemma 2.1 *If, for some x_0, such that $0 < F(x_0) < 1$, the matrix $\Gamma_{F(x)}$, for $x > x_0$ degenerates to the form (2.7), then the equalities (2.5) and, therefore, (2.4) are still valid. Besides,*

$$h^T(x)\Gamma^{-1}_{F(x)} \int_x^\infty h(y) dv_n(y) = -\frac{v_n(x)}{1 - F(x)}, \quad \forall\, x \in \mathbb{R},$$

or

$$\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s) = -\frac{u_n(t)}{1-t}, \quad \forall\, 0 \le t < 1.$$

A similar fact holds with $v_n(u_n)$ replaced by $\hat{v}_n(\hat{u}_n)$.

Remark 2.1 The argument that follows is an adaptation of a quite general treatment of the case of degenerate matrices $\Gamma_{F(x)}$, given in Nikabadze (1987) and Tsigroshvili (1998).

Proof. For $t > t_0$ we have $\gamma(t) = (1, \alpha)^T$, α a positive real number, and

$$\Gamma_t = (1-t)\begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}.$$

Then its image and kernel, or, rather, image and kernel of the corresponding linear operator in $\mathbb{R}^2$, are

$$\mathcal{I}(\Gamma_t) \;=\; \{b : b = \Gamma_t a \text{ for some } a \in \mathbb{R}^2\} = \{b : b = \beta(1-t)(1,\alpha)^T,\ \beta \in \mathbb{R}\}$$

and

$$\mathcal{K}(\Gamma_t) = \{a : \Gamma_t a = 0\} = \{a : a = c(-\alpha, 1)^T,\ c \in \mathbb{R}\}.$$

Moreover, $\int_t^1 \gamma du_n$ and both coordinates of $H(t)$ are in $\mathcal{I}(\Gamma_t)$ and if $b \in \mathcal{I}(\Gamma_t)$ then $\Gamma_t b = (1-t)(1+\alpha^2)b$. Then Γ_t^{-1} is *any* (matrix of) linear operator on $\mathcal{I}(\Gamma_t)$ such that

$$\Gamma_t^{-1} b = \frac{1}{(1-t)(1+\alpha^2)} b + a, \qquad a \in \mathcal{K}(\Gamma_t).$$

But $\gamma(t) = (1, \alpha)^T$ is orthogonal to an $a \in \mathcal{K}(\Gamma_t)$ and therefore

$$\gamma^T(t)\Gamma_t^{-1} b = \frac{1}{(1-t)(1+\alpha^2)}\gamma^T(t) b \tag{2.8}$$

does not depend on the choice of $a \in \mathcal{K}(\Gamma_t)$ and, hence, is defined uniquely. For $b = \int_t^1 \gamma(s)du_n(s)$ this gives the equality in the lemma. Besides, for any $b \in \mathcal{I}(\Gamma_t)$, $a \in \mathcal{I}(\Gamma_t)$,

$$\gamma^T(t)\Gamma_t^{-1}\Gamma_t(b+a) \;=\; \gamma^T(t)\Gamma_t^{-1}\Gamma_t b = \gamma^T(t) b = \gamma^T(t)(b+a),$$

which gives (2.5). The rest of the claim is obvious. □

Now consider the leading term of (2.6) in time $t = F(x)$. It is useful to consider its function parametric version, defined as

$$b_n(\varphi) = u_n(\varphi) - K_n(\varphi), \qquad \varphi \in L_2[0,1], \tag{2.9}$$

where $u_n(\varphi) = \int_0^1 \varphi(s) du_n(s)$, and

$$K_n(\varphi) = K(\varphi, u_n) = \int_0^1 \varphi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(s) du_n(s)\, dt.$$

With slight abuse of notation, denote $b_n(\varphi)$ when $\varphi(\cdot) = I(\cdot \le t)$ by

$$b_n(t) = u_n(t) - \int_0^t \gamma^T(\tau)\Gamma_u^{-1} \int_u^1 \gamma(s) du_n(s)\, du. \tag{2.10}$$

Conditions for weak convergence of u_n are well known: if $\Phi \subset L_2[0,1]$ is a class of functions, such that the sequence $u_n(\varphi), n \ge 1$, is uniformly in n equi-continuous on Φ, then $u_n \to_d u$ in $l_\infty(\Phi)$ where u is standard Brownian bridge, see, e.g., van der Vaart and Wellner (1996). The conditions for the weak convergence of K_n to great extent must be simpler, because, unlike u_n, K_n is continuous linear functional in φ on the whole of $L_2[0,1]$, however, not uniformly in n. We will see, Proposition 2.1 below, that although, for every $\epsilon > 0$, the provisional limit in distribution of $K_n(\varphi)$, viz,

$$K(\varphi) = K(\varphi, u) = \int_0^1 \varphi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(s) du(s)\, dt$$

is continuous on $L_{2,\epsilon}$, the class of functions in $L_2[0,1]$ which are equal 0 on the interval $(1-\epsilon, 1]$, it is not continuous on $L_2[0,1]$. Therefore it is unavoidable to use some condition on φ at $t = 1$. The condition (2.11) below still allows $\varphi(t) \to \infty$ as $t \to 1$ (see examples below).

Theorem 2.1 *(i) Let $L_{2,\epsilon} \subset L_2[0,1]$ be the subspace of all square integrable functions which are equal to 0 on the interval $(1-\epsilon, 1]$. Then, $K_n \to_d K$, on $L_{2,\epsilon}$, for any $0 < \epsilon < 1$.*

(ii) Let, for an arbitrary small but fixed $\epsilon > 0$ and fixed C and $\alpha < 1/2$, $\Phi_\epsilon \subset L_2[0,1]$ be a class of all square integrable functions satisfying the following right tail condition:

$$|\varphi(s)| \le C[\gamma^T(s)\Gamma_s^{-1}\gamma(s)]^{-1/2}(1-s)^{-1/2-\alpha}, \quad \forall\, s > 1-\epsilon. \tag{2.11}$$

Then, $K_n \to_d K$, on Φ_ϵ.

Proof. (i) The integral $\int_t^1 \gamma\, du_n$ as process in t, obviously, converges in distribution to the Gaussian process $\int_t^1 \gamma\, du$. Therefore, all finite-dimensional distributions of $\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma\, du_n$, for $t < 1$, converge to corresponding finite-dimensional distributions of the Gaussian process $\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma\, du$. Hence, for any fixed $\varphi \in L_{2,\epsilon}$, distribution of $K_n(\varphi)$ converges to that of $K(\varphi)$. So, we only need to show tightness, or, equivalently, equicontinuity of $K_n(\varphi)$ in φ. We have

$$
\begin{aligned}
|K_n(\varphi)| &\leq \int_0^1 |\varphi(t)||\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)|dt \\
&\leq \sup_{t\leq 1-\epsilon} |\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)| \int_0^{1-\epsilon} |\varphi(t)|dt,
\end{aligned}
$$

while

$$
\sup_{t\leq 1-\epsilon} |\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)| \to_d \sup_{t\leq 1-\epsilon} |\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du(s)| = O_P(1).
$$

This proves that $K_n(\varphi)$ is equi-continuous in $\varphi \in L_{2,\epsilon}$ and (i) follows.

(ii) To prove (ii), what we need is to show the equi-continuity of $K_n(\varphi)$ on Φ_ϵ. But for this we need only to show that for a sufficiently small $\epsilon > 0$, and uniformly in n,

$$
\sup_{\varphi\in\Phi_\epsilon} \Big| \int_{1-\epsilon}^1 \varphi(t)\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)dt\Big|,
$$

is arbitrarily small in probability. Denote the envelope function for $\varphi \in \Phi_\epsilon$ by Ψ. Then

$$
\int_{1-\epsilon}^1 |\varphi(t)||\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)|dt \leq \int_{1-\epsilon}^1 |\Psi(t)|\,|\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)|dt.
$$

However, bearing in mind that

$$
E|\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)|^2 \leq \gamma^T(t)\Gamma_t^{-1}\gamma(t), \quad \forall\, t \in [0,1],
$$

we obtain that

$$
\begin{aligned}
&E\int_{1-\epsilon}^1 |\Psi(t)|\,|\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)|dt \\
&= \int_{1-\epsilon}^1 |\Psi(t)|E|\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)du_n(s)|dt \\
&\leq \int_{1-\epsilon}^1 |\Psi(t)||\gamma^T(t)\Gamma_t^{-1}\gamma(t)|^{1/2}dt \\
&\leq \int_{1-\epsilon}^1 \frac{1}{(1-t)^{1/2+\alpha}}dt.
\end{aligned}
$$

The last integral can be made arbitrarily small for sufficiently small ϵ. □

Consequently, we obtain the following limit theorem for b_n. Recall, say from van der Vaart and Wellner (1996), that the family of Gaussian random variables $b(\varphi), \varphi \in L_2[0,1]$ with covariance function $Eb(\varphi)b(\varphi') = \int_0^1 \varphi(t)\varphi'(t)dt$ is called (function parametric) standard Brownian motion on Φ if $b(\varphi)$ is continuous on Φ.

Theorem 2.2 *(i) Let Φ be a Donsker class, i.e., let $u_n \to_d u$ in $l_\infty(\Phi)$. Then, for every $\epsilon > 0$,*

$$b_n \to_d b \quad in \quad l_\infty(\Phi \cap \Phi_\epsilon),$$

where $\{b(\varphi), \varphi \in \Phi\}$ is standard Brownian motion.
(ii) If the envelope function $\Psi(t)$ of (2.11) tends to positive (finite or infinite) limit at $t = 1$, then for the process (2.10) we have

$$b_n \to_d b \quad on \quad [0, 1].$$

The condition of *(ii)* is satisfied in all examples below.

Examples. Here we consider four examples of F. In all of them $\gamma^T(s)\Gamma_s^{-1}\gamma(s)$ is of order $(1-s)^{-1}$ and, hence, the upper bound in (2.11) is of the order $(1-s)^{-\alpha}$, $\alpha \le 1/2$, as $s \to 1$.

Consider logistic d.f. F with the scale parameter 1, or equivalently $\psi_f(x) = 2F(x) - 1$. Then $h(x) = (1, 2F(x) - 1)^T$ or $\gamma(s) = (1, 2s - 1)^T$ and

$$\Gamma_s = (1-s)\begin{pmatrix} 1 & s \\ s & (1-2s+4s^2)/3 \end{pmatrix}, \quad \det(\Gamma_s) = \frac{(1-s)^4}{3},$$
$$\Gamma_s^{-1} = \frac{3}{(1-s)^3}\begin{pmatrix} (1-2s+4s^2)/3 & -s \\ -s & 1 \end{pmatrix},$$

so that indeed $\gamma^T(s)\Gamma_s^{-1}\gamma(s) = 4(1-s)^{-1}$, for all $0 \le s < 1$.

Next, suppose F is a normal d.f. with variance 1. Because here $\psi_f(x) = x$, one obtains $h(x) = (1, x)^T$ and $\sigma_f^2(x) = xf(x) + 1 - F(x)$. Denote $\mu(x) = f(x)/(1 - F(x))$. Then

$$\Gamma_{F(x)} = (1 - F(x))\begin{pmatrix} 1 & \mu(x) \\ \mu(x) & x\mu(x) + 1 \end{pmatrix},$$
$$\Gamma_{F(x)}^{-1} = \frac{1}{(1 - F(x))}\,\frac{1}{(x\mu(x) + 1 - \mu^2(x))}\begin{pmatrix} x\mu(x) + 1 & -\mu(x) \\ -\mu(x) & 1 \end{pmatrix}.$$

Hence

$$h^T(x)\Gamma_{F(x)}^{-1}h(x) = \frac{1}{(1 - F(x))}\,\frac{(1 - x\mu(x) + x^2)}{(x\mu(x) + 1 - \mu^2(x))}.$$

However, using asymptotic expansion for the tail of the normal distribution function (see, e.g., Feller (1966), p.179), for $\mu(x)$ we obtain

$$\mu(x) = \frac{x}{1 - S(x)}, \quad \text{where } S(x) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}(2i-1)!!}{x^{2i}} = \frac{1}{x^2} - \frac{3}{x^4} + \dots .$$

From this one can derive that

$$\frac{1 - x\mu(x) + x^2}{x\mu(x) + 1 - \mu^2(x)} = \frac{1 - x^2 S(x) - S(x)}{1 - x^2 S(x) - 2S(x) + S^2(x)}(1 - S(x)) \sim 2,$$

and therefore

$$h^T(x)\Gamma_{F(x)}^{-1}h(x) \sim \frac{2}{1 - F(x)}, \quad x \to \infty.$$

Next, consider the Cauchy d.f. In this case, for $x \to \infty$,

$$f(x) = \frac{1}{\pi}\frac{1}{1 + x^2} \sim \frac{1}{\pi x^2} \text{ and } 1 - F(x) \sim \frac{1}{\pi x},$$

so that

$$\psi_f(x) = \frac{2x}{1 + x^2} \sim \frac{2}{x}, \qquad \sigma_f^2(x) \sim \frac{4}{3\pi x^3}$$

As a consequence of this we get

$$\Gamma_{F(x)} \sim \frac{1}{\pi x^3}\begin{pmatrix} x^2 & x \\ x & 4/3 \end{pmatrix}, \qquad \Gamma_{F(x)}^{-1} \sim \frac{\pi x}{3}\begin{pmatrix} 4/3 & -x \\ -x & x^2 \end{pmatrix}$$

and

$$h^T(x)\Gamma_{F(x)}^{-1}h(x) \sim 4\pi x/9 \sim (4/9)[1 - F(x)]^{-1}$$

as in all previous cases.

Finally, let F be double exponential, or Laplace, d.f. with the density $f(x) = \alpha e^{-\alpha|x-\theta|}, \alpha > 0$, and put $\theta = 0$. For $x > 0$ we get $h(x) = (1, \alpha)^T$ and $\gamma(s) = (1, \alpha)^T$, and Γ_s becomes degenerate, equal to (2.7). Therefore again, see (2.8) with vector $b = \gamma(t)$, for $s > 1/2$

$$\gamma^T(s)\Gamma_s^{-1}\gamma(s) = (1 - s)^{-1}.$$

Next, in this section we wish to clarify the question of a.s. continuity of K_n and K as linear functionals and thus justify the presence of tail condition (2.11). For this purpose it is sufficient to consider particular case, when $\gamma(s) = 1$ is one-dimensional and $\Gamma_s = 1 - s$. In this case

$$K_n(\varphi) = -\int_0^1 \varphi(s)\frac{u_n(s)}{1 - s}ds, \qquad K(\varphi) = -\int_0^1 \varphi(s)\frac{u(s)}{1 - s}ds.$$

The proposition below is of independent interest.

Proposition 2.1 *(i) $K_n(\varphi)$ is continuous linear functional in φ on $L_2[0,1]$ for every finite n.*
(ii) However, the integral

$$\int_0^1 \frac{u^2(s)}{(1-s)^2} ds$$

is almost surely infinite. Moreover,

$$\frac{1}{-\ln(1-s)} \int_0^s \frac{u^2(t)}{(1-t)^2} dt \to_p 1, \quad \text{as} \quad s \to 1.$$

Therefore, $K(\varphi)$ is not continuous on $L_2[0,1]$.

Remark 2.2 It is easy to see that

$$E \int_0^1 \frac{u^2(s)}{(1-s)^2} ds = \infty,$$

but this would not resolve the question of a.s. behaviour of the integral and, hence, of K.

Proof. (i) From the Cauchy-Schwarz inequality we obtain

$$|K_n(\varphi)| \leq \Big(\int_0^1 \varphi^2(s) ds \Big)^{1/2} \Big(\int_0^1 \frac{u_n^2(s)}{(1-s)^2} ds \Big)^{1/2}$$

and the question reduces to whether the integral $\int_0^1 [u_n(s)/(1-s)]^2 ds$ is a.s. finite or not. However, it is, as even $\sup_s |u_n(s)/(1-s)|$ is a proper random variable for any finite n, which proves (i).

(ii) Recall that $u(s)/(1-s)$ is a Brownian motion: if b denotes standard Brownian motion on $[0,\infty)$, then, in distribution,

$$\frac{u(t)}{1-t} = b(\frac{t}{1-t}), \qquad \forall\, t \in [0,1].$$

Hence, in distribution,

$$\int_0^s \frac{u^2(t)}{(1-t)^2} dt = \int_0^s b^2(\frac{t}{1-t}) dt = \int_0^\tau \frac{b^2(z)}{(1+z)^2} dz, \quad \tau = s/(1-s).$$

Integrating the last integral by parts yields

$$(2.12) \qquad \begin{aligned} \int_0^\tau \frac{b^2(z)}{(1+z)^2} dz &= -\frac{b^2(\tau)}{1+\tau} + 2\int_0^\tau \frac{b(z)}{1+z} db(z) + \int_0^\tau \frac{1}{1+z} dz \\ &= -\frac{b^2(\tau)}{1+\tau} + 2\int_0^\tau \frac{b(z)}{1+z} db(z) + \ln(1+z). \end{aligned}$$

Consider the martingale

$$M(t) = \int_0^t \frac{b(z)}{1+z} db(z), \quad t \geq 0.$$

Its quadratic variation process is

$$\langle M \rangle_t = \int_0^t \frac{b^2(z)}{(1+z)^2} dz.$$

Note that $\langle M \rangle_\tau$ equals the term on the left side of (2.12). Divide the equation (2.12) by $\ln(1+\tau)$ to obtain

$$\frac{\langle M \rangle_\tau}{\ln(1+\tau)} = -\frac{b^2(\tau)}{(1+\tau)\ln(1+\tau)} + 2\frac{M(\tau)}{\ln(1+\tau)} + 1.$$

The equalities

$$EM^2(t) = E\langle M \rangle_t = \int_0^t \frac{z}{(1+z)^2} dz = \ln(1+t) - \frac{1}{1+t}, \quad Eb^2(t) = t,$$

imply that

$$\frac{b^2(\tau)}{(1+\tau)\ln(1+\tau)} = o_p(1) \text{ and } \frac{M(\tau)}{\ln(1+\tau)} = o_P(1) \text{ as } \tau \to \infty.$$

Hence, $\langle M \rangle_\tau / \ln(1+\tau) \to_p 1$, as $\tau \to \infty$. □

3 Power

Consider, for the sake of comparison, the problem of fitting a distribution in the one sample location model up to an unknown location parameter. More precisely, consider the problem of testing that $X_1, \cdots, X_n$ is a random sample from $F(\cdot - \theta)$, for some $\theta \in \mathbb{R}$, against the class of all contiguous alternatives, i.e. such sequences of alternative distributions $A_n(\cdot - \theta)$, where

$$\Big(\frac{dA_n(x)}{dF(x)}\Big)^{1/2} = 1 + \frac{1}{2\sqrt{n}} g(x) + r_n(x),$$
$$\int g^2(x) dF(x) < \infty, \qquad \int r_n^2(x) dF(x) = o(\frac{1}{n}).$$

As is known, and as can intuitively be easily understood, we should be interested only in the class of functions $g \in L_2(F)$ that are orthogonal to ψ_f:

(3.1) $$\int g(x)\psi_f(x) dF(x) = 0.$$

Indeed, as g describes a functional "direction" in which the alternative A_n deviates from F, if it has a component collinear with ψ_f,

$$g(x) = g_{\perp}(x) + c\psi_f(x), \quad \int g_{\perp}(x)\psi_f(x)dF(x) = 0,$$

then infinitesimal changes in the direction $c\psi_f$ will be explained by, or attributed to, the infinitesimal changes in the value of parameter, that is, "within" parametric family. Hence it can not (and should not) be detected by a test for our parametric hypothesis. So, we assume that g and ψ_f are orthogonal.

Since θ remains unspecified, we still need to estimate it. Suppose $\bar{\theta}$ is its MLE under F and consider empirical process $\bar{v}_n$ based on $\bar{e}_i := X_i - \bar{\theta}, i = 1, 2, \ldots, n$:

$$\bar{v}_n(x) = \sqrt{n}[\bar{F}_n(x) - F(x)], \qquad \bar{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{\bar{e}_i \leq x\}}.$$

If we assume the hypothetical θ known, we would come back to the empirical process v_n.

It is known, see, e.g., Khmaladze (1979), that the asymptotic shift of $\bar{v}_n$ and v_n under the sequence of alternatives A_n with orthogonality condition (3.1) is the same and equals the function

$$G(x) = \int_{-\infty}^{x} g(y)dF(y).$$

However, the process $\bar{v}_n$ has asymptotic representation

$$\bar{v}_n(x) = v_n(x) - \frac{dF}{d\theta}(x - \theta) \int \psi_f(y)dv_n(y) + o_P(1)$$

and, the main part on the right is orthogonal projection of v_n - see Khmaladze (1979) for precise statement, see also Tjurin (1974). Heuristically speaking, it implies that the process $\bar{v}_n$ is "smaller" than v_n. In particular, $Var\, \bar{v}_n(x) \leq Var\, v_n(x)$ for all x. Therefore, tests based on omnibus statistics, which typically measure an "overall" deviation of an empirical distribution function from F, or of empirical process from 0, will have better power if based on $\bar{v}_n$ than v_n. From a certain point of view this may seem a paradox, as it implies that, even if we know the parameter θ, it would still be better to replace it by an estimator, because the power of many goodness of fit tests will thus increase.

Transformation of the process $\bar{v}_n$ asymptotically coincides with the process w_n we study here, and moreover, the relationship between the two processes is one-to-one. Therefore, any statistic from one is, asymptotically, a statistic from the other, and the processes yield the same inference.

With the process $\hat{v}_n$ the situation is different: although it can be shown that the shift of this process under alternatives A_n with orthogonality condition (3.1) is again function G, with general estimator $\hat{m}_n$ and, therefore, the general form of R_n, this process is not a transformation of v_n only, and therefore is not its projection. In other words, it is not as "concentrated" as $\bar{v}_n$. The bias part of R_n brings in additional randomisation. As a result, one will have less power in tests based on omnibus statistics from $\hat{v}_n$.

We must add that with the estimator, used by Müller, Schick and Wefelmeyer, and therefore, with their simple form of R_n, the process $\hat{v}_n$ is again asymptotically a projection, although in general a skew one, of the process v_n. As described in Khmaladze (1979), it is asymptotically in one-to-one relationship with the process $\bar{v}_n$, and, therefore w_n. Hence a statistic from $\hat{v}_n$ is, in this case, also a statistic from each of the other two, and the only difference between this processes is that $\hat{v}_n$ and $\bar{v}_n$ are not asymptotically distribution free, while w_n is.

4 Weak convergence of w_n

In this section we prove weak convergence for the process w_n, given by (2.3) and (2.4). In view of (2.6), (2.9) and the fact that the weak convergence of the first part in the right hand side of (2.6) was proved in Theorem 2.1, it suffices to show that the process η_n of (2.6) is asymptotically small. Being the transformation of "small" process ξ_n, the smallness of η_n is plausible. However, the transformation $K(\cdot, \xi_n)$ is not continuos in ξ_n in uniform metric. Indeed, although in the integration by parts formula

$$\int_t^1 \gamma(s) d\xi_n(F^{-1}(s)) = \xi_n(F^{-1}(s))\gamma(s)|_{s=t}^1 - \int_t^1 \xi_n(F^{-1}(s)) d\gamma(s),$$

we can show, that $\xi_n(F^{-1}(1))\gamma(1) = 0$, the integral on the right side is not necessarily small if $\gamma(t)$ is not bounded at $t = 1$. However, the time transformed score function $\psi_f(F^{-1}(t))$, the second coordinate of $\gamma(t)$, is unbounded at $t = 1$ already for normal d.f. F. Therefore, one can not prove the smallness of η_n in sufficient generality, using only uniform smallness of ξ_n.

If we use, however, quite mild additional assumption on the right tail of ξ_n, or rather of $\hat{v}_n$ and f, we can get the weak convergence of w_n basically iunder the same conditions as in Theorem 2.2. Namely, assume that for some positive $\beta < 1/2$,

$$\text{(4.1)} \qquad \sup_{y>x} \frac{|\hat{v}_n(y)|}{(1-F(y))^\beta} = o_P(1), \quad \text{as } x \to \infty,$$

uniformly in n. Recall that the same condition for v_n is satisfied for all $\beta < 1/2$.

Denote tail expected value and variance of ψ_f by

$$E[\psi_f|x] := E[\psi_f(e_1)|e_1 > x], \quad Var[\psi_f|x] := Var[\psi_f(e_1)|e_1 > x].$$

Now we formulate two more conditions on F.

a) For any $\epsilon > 0$ the function $\psi_f(F^{-1})$ is of bounded variation on $[\epsilon, 1-\epsilon]$ and for some $\epsilon > 0$ it is monotone on $[1-\epsilon, 1]$.

b) For some $\delta > 0$, $\epsilon > 0$ and some $C < \infty$,

$$\frac{(\psi_f(x) - E[\psi_f|x])^2}{Var[\psi_f|x]} < C(1 - F(x))^{-2\delta}, \quad \forall\, x : F(x) > 1 - \epsilon.$$

Note that in terms of the above notation,

$$\gamma(t)^T \Gamma_t^{-1} \gamma(t) = \frac{1}{1 - F(x)} [1 + \frac{\Big(\psi_f(x) - E[\psi_f|x]\Big)^2}{Var[\psi_f|x]}], \quad t = F(x). \tag{4.2}$$

Hence, condition b) is equivalent to

$$\gamma(t)^T \Gamma_t^{-1} \gamma(t) \leq C(1 - t)^{-1-2\delta}.$$

Condition a) implies that $\psi_f(F^{-1}(t))$ and $\gamma(t)^T \Gamma_t^{-1} \gamma(t)$ are bounded on $[\epsilon, 1-\epsilon]$. Both conditions are easily satisfied in all examples of Sec. 2, the latter – even with $\delta = 0$.

Our last condition is as follows.

c) For some $C < \infty$ and $\beta > 0$ as in (4.1)

$$\Big|\int_x^\infty [1 - F(y)]^\beta d\psi_f(y)\Big| \leq C|\psi_f(x) - E[\psi_f|x]|.$$

Condition c) is also easily satisfied in all examples of Sec. 2 for arbitrarily small β. For example, for logistic distribution, with $t = F(x)$, $\psi_f(x) = 2t - 1$ and

$$\Big|\int_x^\infty [1 - F(y)]^\beta d\psi_f(y)\Big| = 2\int_t^1 (1 - s)^\beta ds = \frac{2}{\beta + 1}(1 - t)^{\beta+1}$$

while

$$|\psi_f(x) - E[\psi_f|x]| = (1 - t)$$

and their ratio tends to 0, as $t \to 1$. For normal distribution,

$$\int_x^\infty [1 - F(y)]^\beta d\psi_f(y) \sim \int_x^\infty \frac{1}{y^\beta} f^\beta(y) dy \leq \frac{1}{x} \int_x^\infty y^{1-\beta} f^\beta(y) dy$$

while

$$|\psi_f(x) - E[\psi_f|x])| = |x - \frac{f(x)}{1 - F(x)}| \sim \frac{x}{x^2 - 1}, \quad x \to \infty,$$

and the ratio again tends to 0, as $x \to \infty$, etc.

Let us recall the notation

$$K(\varphi, \xi_n) = \int_0^1 \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^1 \gamma(s)\xi_n(F^{-1}(ds))dt,$$

and for a given indexing class Φ of functions from $L_2[0,1]$ let $\Phi \circ F = \{\varphi(F(\cdot)), \varphi \in \Phi\}$.

Theorem 4.1 *(i) Suppose conditions (4.1) and a)-c) are satisfied with $\beta > \delta$. Then, on the class Φ_ϵ as in Theorem 2.1 but with $\alpha < \beta - \delta$ we have*

$$\sup_{\varphi \in \Phi_\epsilon} |K(\varphi, \xi_n)| = o_P(1), \quad n \to \infty.$$

Therefore, if Φ is a Donsker class, then, for every $\epsilon > 0$,

$$w_n \to_d b \quad in \quad l_\infty(\Phi \cap \Phi_\epsilon \circ F),$$

where $\{b(\varphi), \varphi \in \Phi\}$ is standard Brownian motion.
(ii) If, in addition, $\delta \le \alpha$, then for the time transformed process $w_n(F^{-1}(\cdot))$ (2.3) we have

$$w_n(F^{-1}(\cdot)) \to_d b(\cdot) \quad in \quad D[0,1].$$

Proof. Note, that

$$\gamma(t)^T \Gamma_t^{-1}(0,a)^T = \frac{1}{1-F(x)} \frac{\Big(\psi_f(x) - E[\psi_f|x]\Big)\, a}{Var[\psi_f|x]}, \quad t = F(x),\ \forall\, a \in \mathbb{R}.$$

Use this equality for $a = \int_t^1 (1-s)^\beta d\psi_f(F^{-1}(s))$. Then condition c) implies that

$$|\gamma(t)^T \Gamma_t^{-1}(0,a)^T| \le C\gamma(t)^T \Gamma_t^{-1}\gamma(t), \quad \forall\, t < 1. \tag{4.3}$$

Now we prove the first claim.

(i) Use the notation $\xi_n'(t) = \xi_n(x)$ with $t = F(x)$. Since we expect singularities at $t = 0$ and, especially, at $t = 1$ in both integrals in $K(\varphi, \xi_n)$ we will isolate the neighbourhood of these points and consider it separately. Mostly we will take care of the neighbourhood of $t = 1$. The neighbourhood of $t = 0$ can be treated more easily (see below). First assume Γ_t^{-1} non-degenerate for all $t < 1$. We have

$$\begin{aligned} & \int_0^1 \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^1 \gamma(s)\xi_n'(ds)dt \\ &= \int_0^{1-\epsilon} \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^{1-\epsilon} \gamma(s)\xi_n'(ds)dt \\ &\quad + \int_0^{1-\epsilon} \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_{1-\epsilon}^1 \gamma(s)\xi_n'(ds)dt \\ &\quad + \int_{1-\epsilon}^1 \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^1 \gamma(s)\xi_n'(ds)dt. \end{aligned} \tag{4.4}$$

Since γ has bounded variation we can apply to the inner integral integration by parts. Consider the third summand on the right side. First note that, when proving that it is small, we can replace ξ_n by the difference $\hat{v}_n - v_n$ only: indeed, since $df(F^{-1}(s)) = \psi_f(x)f(x)dx$, according to (2.5) the integral

$$\int_{1-\epsilon}^{1} \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^1 \gamma(s) df(F^{-1}(s))dt$$

is the second coordinate of $\int_{1-\epsilon}^1 \varphi(t)\gamma(t)dt$, and is small for ϵ small anyway. Denote $\hat{u}_n(t) = \hat{v}_n(x)$ and consider

$$\begin{aligned} &\int_{1-\epsilon}^{1} \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^1 \gamma(s)\hat{u}_n(ds)dt \\ &= \int_{1-\epsilon}^{1} \varphi(t)\gamma(t)^T \Gamma_t^{-1}[-\gamma(t)\hat{u}_n(t) - \int_t^1 \hat{u}_n(s)d\gamma(s)]dt. \end{aligned}$$

Using assumption on φ and conditions (4.1) and b), from (4.2) we obtain

$$\begin{aligned} &\Big|\int_{1-\epsilon}^{1} \varphi(t)\gamma(t)^T \Gamma_t^{-1}\gamma(t)\hat{u}_n(t)dt\Big| \\ &\leq \int_{1-\epsilon}^{1} [\gamma(t)^T \Gamma_t^{-1}\gamma(t)]^{1/2} \frac{1}{(1-t)^{1/2+\alpha-\beta}}dt \sup_{t>1-\epsilon} \frac{|\hat{u}_n(t)|}{(1-t)^\beta} \\ &\leq \int_{1-\epsilon}^{1} \frac{1}{(1-t)^{1+\alpha+\delta-\beta}}dt \sup_{t>1-\epsilon} \frac{|\hat{u}_n(t)|}{(1-t)^\beta}, \end{aligned}$$

which is small for small ϵ as soon as $\alpha < \beta - \delta$.

Now note that $\int_t^1 \hat{u}_n(s)d\gamma(s) = (0, \int_t^1 \hat{u}_n(s)d\psi_f(F^{-1}(s)))^T$. Using monotonicity of $\psi_f(F^{-1})$ for small enough ϵ we obtain

$$(4.5)\quad |\int_t^1 \hat{u}_n(s)d\psi_f(F^{-1}(s))| < C|\int_t^1 (1-s)^\beta d\psi_f(F^{-1}(s))| \quad \sup_{s>1-\epsilon} \frac{|\hat{u}_n(s)|}{(1-s)^\beta}$$

Therefore, using (4.3), for the double integral we get

$$\begin{aligned} &\Big|\int_{1-\epsilon}^{1} \varphi(t)\gamma(t)^T \Gamma_t^{-1} \int_t^1 \hat{u}_n(s)d\gamma(s)dt| \\ &\qquad \leq C\int_{1-\epsilon}^{1} |\varphi(t)|\gamma(t)^T \Gamma_t^{-1}\gamma(t)dt \sup_{s>1-\epsilon} \frac{|\hat{u}_n(s)|}{(1-s)^\beta}, \end{aligned}$$

and the integral on the right side, as we have seen in above, is small as soon as $\alpha < \beta - \delta$. The same conclusion is true for $\hat{u}_n$ replaced by u_n.

Since (4.5) implies the smallness of $\int_{1-\epsilon}^1 \hat{u}_n(s)d\psi_f(F^{-1}(s))$ and $\int_{1-\epsilon}^1 u_n(s)d\psi_f(F^{-1}(s))$, to prove that the middle summand on the right side of (4.4) is small one needs only finiteness of ψ_f in each x with $0 < F(x) < 1$, which follows from a). This and uniform in x smallness of ξ_n proves smallness of the first summand as well.

The smallness of integrals

$$\int_0^\epsilon \varphi(t)\gamma(t)^T\Gamma_t^{-1}\gamma(t)\int_t^1 \gamma(s)\xi_n'(ds)dt$$

follows from $\Gamma_t^{-1} \sim \Gamma_0^{-1}$ and square integrability of φ and γ.

If Γ_t^{-1} becomes degenerate after some t_0, for these t we get

$$\gamma(t)^T\Gamma_t^{-1}\int_t^1 \gamma(s)\xi_n'(ds) = \frac{\xi_n'(t)}{1-t},$$

and the smallness of all tail integrals easily follows for our choice of the indexing functions φ.

(ii) Since for $\delta \le \alpha$ the envelope function $\Psi(t)$ of (2.11) satisfies inequality

$$\Psi(t) \ge (1-t)^{\delta-\alpha},$$

it has positive finite or infinite lower limit at $t = 1$. But then it is possible to choose as an indexing class the class of indicator functions $\varphi(t) = \mathbb{I}_{\{t\le\tau\}}$ and the claim follows. □

References

[1] Akritas, M.G. and van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–567.

[2] Feller, W. (1978) *Introduction to probability theory and its applications*. J.Wiley, New York.

[3] Khmaladze, E. V. (1979). The use of ω^2 tests for testing parametric hypotheses. *Theory of Probab. & Appl.*, **24**(2), 283 - 301.

[4] Khmaladze, E. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** (1981), 246–265.

[5] Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32**, 995-1034.

[6] Koul, Hira L. (2002). *Weighted empirical processes in dynamic nonlinear models*. Second edition of *Weighted empiricals and linear models* [Inst. Math. Statist., Hayward, CA, 1992; MR1218395 (95c:62061)]. Lecture Notes in Statistics, 166. *Springer-Verlag, New York*.

[7] Loynes, R.M. (1980). The empirical distribution function of residuals from generalised regression. *Ann. Statist.*, **8**, 285–298.

[8] Müller, U.U., Schick, A. & Wefelmeyer, W. (2006). Estimating the error distribution function in semi-parametric regression. A preprint. To appear in *Statistics and Decisions.*

[9] Nikabadze, A. M. (1987). A method for constructing likelihood tests for parametric hypotheses in R^m. (Russian) Teor. Veroyatnost. i Primenen. **32**, no. 3, 594–598.

[10] Tjurin, T. Conditional probability distributions. Lecture Notes, No. 2. Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, 1974.

[11] Tsigroshvili, Z. (1998). Some notes on goodness-of-fit tests and innovation martingales. (English. English, Georgian summary) *Proc. A. Razmadze Math. Inst.* **117**, 89–102.

[12] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York.