L_1-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

Model checks in statistics: An innovation process approach

Winfried Stute

University of Giessen, Germany

Abstract: In this paper we study a class of Gaussian processes which typically appear as limits of marked empirical processes when composite models need to be checked. A transformation to their martingale part is derived which when applied to the empirical process gives rise to asymptotically distribution-free tests for composite models.

Key words: Model checks, innovation martingale, gaussian process.

AMS subject classification: 62G10, 62G30, 60G15.

1 Introduction

In this paper we will develop a general methodology for nonparametrically testing the goodness-of-fit of a parametric or a semiparametric model. To begin with the simplest example, assume one observes independent identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$ on the real line, from some unknown distribution function (d.f.) F. Furthermore, let $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ be a given family of distribution functions parametrized by some vector $\theta \in \Theta \subset \mathbb{R}^k$. To keep the discussion as simple as possible, we will assume that no nuisance parameters are present so that F_θ is uniquely determined by θ. The problem of how to test for the hypothesis

$$H_0 : F \in \mathcal{F}$$

has attracted many researchers over the past decades. Most of the test statistics are certain functionals of the underlying empirical process. More precisely, denote with

$$F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \le x\}}, \quad x \in \mathbb{R},$$

the empirical distribution function of the data. The by now classical invariance principle of Donsker (1952) then asserts that the empirical process

$$\alpha_n(x) = n^{1/2}[F_n(x) - F(x)], \tag{1}$$

in the Skorokhod space $D[-\infty, \infty]$, converges in distribution to

$$\alpha_\infty := B^0 \circ F.$$

Here, B^0 is a Brownian Bridge on the unit interval, i.e., a centered Gaussian process with covariance function

$$\mathrm{Cov}[B^0(s), B^0(t)] = \min(s,t) - st.$$

For details and extensions, see Gaenssler and Stute (1979) and Shorack and Wellner (1986). To test for a simple hypothesis, $F = F_{\theta_0}$, one needs to replace F in (1) by F_{θ_0} so that under H_0

$$\alpha_n^0 \equiv n^{1/2}[F_n - F_{\theta_0}]$$

equals α_n. In particular, critical values if not available for finite sample size may be obtained from the distribution of the limit α_∞. For composite hypotheses, things unfortunately become more complicated. Under $H_0, F = F_{\theta_0}$ for some unknown $\theta_0 \in \Theta$, the true parameter. Since now θ_0 remains unspecified, it needs to be estimated from the data by some θ_n, say. We thus come up with the so-called empirical process with estimated parameters

$$\hat{\alpha}_n \equiv n^{1/2}[F_n - F_{\theta_n}].$$

This process may be viewed as a basic device to measure the deviance between a completely nonparametric and a parametric fit. It has been extensively studied by Durbin (1973). To briefly recall its ingredients, assume that θ_n has, under H_0, a linear expansion

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{i=1}^{n} l(X_i, \theta_0) + o_{\mathbb{P}}(1),$$

where l is a proper vector-valued function with expectation zero and finite covariance matrix. Then, under appropriate smoothness assumptions,

$$\hat{\alpha}_n(x) = \alpha_n(x) - G^t(x, \theta_0) \int l(y, \theta_0) \alpha_n(dy) + o_{\mathbb{P}}(1)$$

uniformly in x, where

$$G(x) \equiv G(x, \theta_0) = \left.\frac{\partial F_\theta(x)}{\partial \theta}\right|_{\theta=\theta_0}.$$

From this we readily get

$$\hat{\alpha}_n \to B^0 \circ F - G^t V \equiv \hat{\alpha}_\infty$$

with

$$V = \int l(y, \theta_0) B^0 \circ F(dy).$$

The limit $\hat{\alpha}_\infty$ is again a centered Gaussian process, but its covariance function is more complicated, and tables for critical values may and will depend on θ_0 and are not readily available. In such a situation a parametric bootstrap may offer a useful possibility to approximate the distribution of $\hat{\alpha}_n$ under H_0; see Stute et al. (1993).

Though from a computational point of view, this seems to be quite satisfactory, it is worthwhile considering also another approach which not only provides an approximation in distribution, but also leads to a deeper understanding of the involved processes. For $\hat{\alpha}_n$, this approach has been initiated, in a landmark paper, by Khmaladze (1981). As to this, recall that B^0 has the representation

$$B^0(t) = B(t) - tB(1), \qquad 0 \le t \le 1,$$

in terms of a Brownian Motion B and, vice versa,

$$B(t) = B^0(t) + \int_0^t \frac{B^0(x)}{1-x} dx. \tag{2}$$

In the latter equation B may be viewed as the innovation martingale and the integral as the compensator in the Doob-Meyer decomposition of B^0. Now, Khmaladze (1981) was able to also find the corresponding decomposition for $\hat{\alpha}_\infty$. Replacing $\hat{\alpha}_\infty$ by its innovation martingale then leads to a new process, say $T\hat{\alpha}_\infty$, which is a Gaussian martingale and hence a Brownian Motion w.r.t. proper time. In particular, this process is distribution-free modulo a transformation in time and therefore is a good candidate for giving rise to distribution-free test statistics.

It is the purpose of the present paper to extend Khmaladze's (1981) approach to a much more general setting. This will enable us to design model checks in the context of regression, times series, multivariate analysis and survival analysis, among others. Now, rather than (2), our starting point will be the following representation of B^0 in terms of B, which incorporates a transformation in time and a scale factor:

$$B^0(t) = (1-t)B\left(\frac{t}{1-t}\right). \tag{3}$$

To show that the right hand side has the same covariance structure as B^0, just use the monotonicity of the time transformation and apply

$$\mathrm{Cov}[B(s), B(t)] = \min(s,t).$$

Monotonicity will also be a crucial issue in the examples which will be shortly discussed. In each case the limit process will be of the following type:

$$R_\infty = G_1 B \circ \psi - G_2^t V. \tag{4}$$

Here, G_1 and G_2 are two deterministic functions, ψ denotes the aforementioned nondecreasing nonnegative time transformation and V is a normal vector, which may and will depend on B. Conclude from the introductory remarks that for $R_\infty = \hat{\alpha}_\infty$, i.e., for the empirical process with estimated parameters,

$$G_1 = 1 - F \qquad \psi = F/(1-F)$$

and

$$G_2 = \frac{\partial F_\theta}{\partial \theta} \qquad \text{at } \theta = \theta_0.$$

In our second example we discuss a situation which typically comes up when the X-data represent lifetimes. Under random right censorship one observes, due to other causes of failure, variables $Z_i = \min(X_i, Y_i), 1 \le i \le n$, where the censoring variables are independent and also independent of the X's, with the common d.f. G. Also available are 0-1 variables $\delta_i = 1_{\{X_i \le Y_i\}}$ indicating whether X_i has been observed or not. Since under censorship F_n may not be available, it needs to be replaced by the nonparametric MLE adapted to the new framework:

$$1 - \hat{F}_n(x) = \prod_{i=1}^{n} \left[1 - \frac{\delta_{[i:n]}}{n-i+1}\right]^{1_{\{Z_{i:n} \le x\}}}, x \ge 0. \tag{5}$$

This is the famous product-limit estimator due to Kaplan-Meier (1958). In (5), $Z_{1:n} \le \ldots \le Z_{n:n}$ are the order statistics of the observed Z's. Finally $\delta_{[i:n]}$ denotes the δ-variable associated with $Z_{i:n}$. Note that $\hat{F}_n$ boils down to F_n if all δ's equal one. Breslow and Crowley (1974) extended Donsker's invariance principle to the present setup. They showed that the so-called Kaplan-Meier process

$$\beta_n(x) = n^{1/2}[\hat{F}_n(x) - F(x)]$$

converges in distribution to a centered Gaussian process β_∞. In our notation it admits the representation

$$\beta_\infty = (1 - F)B \circ C,$$

where, under a continuity assumption,

$$C(x) = \int_0^x \frac{F(dy)}{(1 - F(y))^2(1 - G(y))}.$$

Hence, in terms of (4), the Kaplan-Meier process with estimated parameters converges to R_∞ with

$$G_1 = 1 - F \quad \text{and} \quad \psi = C.$$

The function G_2 is the same as before. A detailed analysis of this example may be found in Nikabadze and Stute (1997).

In our next example, we will discuss the important problem of model checks in regression. For this, let (X, Y) be a bivariate random vector such that $\mathbb{E}|Y| < \infty$. Denote with

$$m(x) = \mathbb{E}\{Y|X = x\}$$

the regression function of Y w.r.t. $X = x$. Also, let $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ be a given parametric family of candidates for m. For example, the m_θ's may consist of all functions spanned by a given basis $g_1, \ldots, g_k$:

$$m_\theta(x) = \theta_1 g_1(x) + \ldots + \theta_k g_k(x).$$

This includes, e.g., all polynomials or trigonometric polynomials with a given bound on the degree. To test for the hypothesis

$$H_0 : m \in \mathcal{M},$$

let θ_n be, under H_0, any estimator of θ_0, computed from a sample of independent replicates of (X, Y), admitting a representation

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{i=1}^n l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1).$$

The residuals

$$\hat{\varepsilon}_{in} = Y_i - m_{\theta_n}(X_i), \qquad 1 \le i \le n,$$

traditionally play an important role in model diagnostics for regression. In our approach they will be embedded into a marked point process

$$\hat{\gamma}_n(x) = n^{-1/2} \sum_{i=1}^{n} \hat{\varepsilon}_{in} 1_{\{X_i \le x\}}, \qquad x \in \mathbb{R}.$$

Under H_0, one can show that

$$\hat{\gamma}_n \to \hat{\gamma}_\infty = B \circ \psi - G_2^t V,$$

where

$$\psi(x) = \int_{-\infty}^{x} \sigma^2(u) F(du)$$

and

$$G_2(x) = \int_{-\infty}^{x} \frac{\partial m_\theta(u)}{\partial \theta} F(du) \qquad \text{at } \theta = \theta_0,$$

with $\sigma^2(u) = \text{Var}\{Y|X = u\}$ denoting the conditional variance and F being the marginal distribution of X. See Stute (1997) and Stute, Thies and Zhu (1996) for details. We thus see that (4) applies again with $G_1 \equiv 1$ and G_2, ψ from above.

Another example to which our methodology will apply is in a time series context. For this, let $X_1, X_2, \ldots$ be a stationary sequence of observations. We are interested in the dynamics of the process. One possibility would be to decompose a future observation X_i into the part explained by the past observations and the i-th innovation:

$$X_i = m(X_{i-1}, X_{i-2}, \ldots) + \varepsilon_i.$$

Thus m is the regression function of X_i given $\mathcal{F}_{i-1} = \sigma(X_{i-1}, X_{i-2}, \ldots)$. If we are, e.g., interested in testing whether the X-sequence is first order autoregressive with $m \in \mathcal{M}$, a pre-specified parametric model, we could form, similar to the regression case, the process

$$\hat{\delta}_n(x) = n^{-1/2} \sum_{i=1}^{n} [X_i - m_{\theta_n}(X_{i-1})] \, 1_{\{X_{i-1} \le x\}}.$$

Due to dependencies some little extra work is needed to show that also in this case $\hat{\delta}_n \to \hat{\delta}_\infty$, where $\hat{\delta}_\infty$ is of type (4) with $G_1 = 1$ and some appropriate ψ and G_2.. Note that the stationary distribution now also depends on θ_0. See Koul and Stute (1997) for details.

Our final example concerns a generalized linear model. Here one observes a sequence of multivariate data $(X_i, Y_i), 1 \le i \le n$, from $\mathbb{R}^{k+1}$, for which it is assumed that the regression function of Y_1 given X_1 has a decomposition into a linear form of X_1 and a specified link function h:

$$m(\underset{\sim}{x}) = \mathbb{E}[Y_1|X_1 = \underset{\sim}{x}] = h(\theta_{10}x_1 + \ldots + \theta_{k0}x_k).$$

The associated process for testing that m is of this form becomes

$$\hat{\varepsilon}_n(x) = n^{-1/2}\sum_{i=1}^{n}[Y_i - h(< \theta_n, X_i >)]1_{\{<\theta_n, X_i> \leq x\}},$$

where θ_n is an estimator of $\theta_0 = (\theta_{10}, \ldots, \theta_{k0})$ and $< \cdot, \cdot >$ is the scalar product in $\mathbb{R}^k$.

Again it can be shown that under standard regularity assumptions $\hat{\varepsilon}_n$ in the limit is of the form (4). The case when h is unspecified requires nonparametric estimation of the (univariate) link function.

This list of examples indicates that the class of Gaussian processes considered in (4) is rich enough to cover many interesting cases which typically appear as limit processes when parameters need to be estimated. Since their distributional character is not readily understood, we propose to transform R_∞ from (4) to another process, which has a much nicer structure, namely a Brownian Motion in proper time. This will be the content of the following section.

2 Transformation of Gaussian processes

As we have seen in the first section Gaussian processes of type (4)

$$R_\infty = G_1 B \circ \psi - G_2^t V$$

frequently appear as limits of certain marked empirical processes when parameters need to be estimated. In this section we introduce a transformation T which maps R_∞ into a Brownian Motion in proper time. This transformation will be a composition of two linear operators T_1 and T_2 which will be defined now.

Assume that G_1 is a function of bounded variation which is positive on its support. For the sake of simplicity only a continuous G_1 will be considered. Put

$$(T_1 f)(x) = f(x) - \int_{-\infty}^{x} \frac{f(y)}{G_1(y)} G_1(dy). \tag{6}$$

Here f varies in the class of functions for which the integral is defined.

Lemma 1 *The stochastic process $T_1G_1B\circ\psi$ is a Brownian Motion w.r.t. time*

$$\varphi(x) = \int_{-\infty}^{x} G_1^2(y)\psi(dy).$$

Proof: We have

$$T_1G_1B\circ\psi(x) = G_1B\circ\psi(x) - \int_{-\infty}^{x} B\circ\psi dG_1 = \int_{-\infty}^{x} G_1 dB\circ\psi.$$

It follows that $T_1G_1B\circ\psi$ is a centered Gaussian process with covariance function $\min\{\varphi(x_1),\varphi(x_2)\}$ at x_1, x_2. □

For the empirical process and the Kaplan-Meier process the function G_1 equals $1-F$ so that

$$T_1f = f + \int^{\bullet} \frac{f}{1-F}dF,$$

which corresponds to (2). For the other examples, $G_1 \equiv 1$ in which case the integral in (6) vanishes and T_1 reduces to the identity operator.

Since T_1 is a linear operator and since V does not depend on x, we obtain

$$\begin{aligned} T_1R_\infty &= T_1G_1B\circ\psi - T_1G_2^tV \\ &= B\circ\varphi - (T_1G_2)^tV \equiv B\circ\varphi - G_3^tV, \end{aligned}$$

say, where

$$\begin{aligned} G_3 &= T_1G_2 = G_2 - \int^{\bullet} \frac{G_2}{G_1}dG_1 \\ &= \int^{\bullet} \left[\frac{dG_2}{dG_1} - \frac{G_2}{G_1}\right] dG_1, \end{aligned}$$

provided the Radon-Nikodym derivative of G_2 w.r.t. G_1 exists. In the next step we construct a linear operator T_2 with the following two properties:

$$T_2G_3 \equiv 0 \tag{7}$$

$$T_2B\circ\varphi = B\circ\varphi \quad \text{in distribution.} \tag{8}$$

Putting $T = T_2\circ T_1$ we therefore get in distribution

$$TR_\infty = T_2(B\circ\varphi - G_3^tV) = B\circ\varphi,$$

i.e., TR_∞ is a Brownian Motion w.r.t. time φ.

To define T_2, let $R^0_\infty = B \circ \varphi$ be a Brownian Motion w.r.t. time φ. Also, let G be a given vector-valued function. Define the matrix

$$A(x) = \int_x^\infty \left(\frac{dG}{d\varphi}\right)\left(\frac{dG}{d\varphi}\right)^t d\varphi$$

and

$$T_2 f(x) = f(x) - \int_{-\infty}^x \left(\frac{dG}{d\varphi}\right)^t (y) A^{-1}(y) \left[\int_y^\infty \frac{dG}{d\varphi}(z) f(dz)\right] \varphi(dy) \qquad (9)$$

assuming that A is nonsingular.

Lemma 2 *We have*

(i) $T_2 G^t \equiv 0$

(ii) $T_2 R^0_\infty = R^0_\infty$ in distribution

Proof: (i) is trivial; as to (ii), we have for $s \le t$,

$$\begin{aligned}
&\mathrm{Cov}[\mathrm{T_2 R^0_\infty(s), T_2 R^0_\infty(t)}] = \mathrm{E}[\mathrm{R^0_\infty(s) R^0_\infty(t)}] \\
&- \mathbb{E}\left\{R^0_\infty(t) \int_{-\infty}^s \left(\frac{dG}{d\varphi}\right)^t (y) A^{-1}(y) \left[\int_y^\infty \frac{dG}{d\varphi}(z) R^0_\infty(dz)\right] \varphi(dy)\right\} \\
&- \mathbb{E}\left\{R^0_\infty(s) \int_{-\infty}^t \left(\frac{dG}{d\varphi}\right)^t (y) A^{-1}(y) \left[\int_y^\infty \frac{dG}{d\varphi}(z) R^0_\infty(dz)\right] \varphi(dy)\right\} \\
&+ \mathbb{E}\left\{\int_{-\infty}^s \int_{-\infty}^t \left(\frac{dG}{d\varphi}\right)^t (y_1) A^{-1}(y_1) \left[\int_{y_1}^\infty \frac{dG}{d\varphi}(z) R^0_\infty(dz)\right] \varphi(dy_1)\right. \\
&\qquad \left.\left[\int_{y_2}^\infty \left(\frac{dG}{d\varphi}\right)^t (z) R^0_\infty(dz)\right] A^{-1}(y_2) \left(\frac{dG}{d\varphi}\right)(y_2) \varphi(dy_2)\right\}.
\end{aligned}$$

The first expectation equals $\varphi(s)$, while the second is easily seen to be

$$\int_{-\infty}^s \left(\frac{dG}{d\varphi}\right)^t (y) A^{-1}(y) \int_y^t \frac{dG}{d\varphi}(z) \varphi(dz) \varphi(dy).$$

Finally, the third and fourth expectations equal

$$\int_{-\infty}^s \left(\frac{dG}{d\varphi}\right)^t (y) A^{-1}(y) \int_y^s \frac{dG}{d\varphi}(z) \varphi(dz) \varphi(dy)$$

and

$$\int_{-\infty}^{s}\int_{-\infty}^{t}\left(\frac{dG}{d\varphi}\right)^t (y_1)A^{-1}(y_1)A(y_1 \vee y_2)A^{-1}(y_2)\frac{dG}{d\varphi}(y_2)\varphi(dy_2)\varphi(dy_1),$$

respectively. Summation and an application of Fubini complete the proof. □

Theorem 1 *Assume that*

$$R_\infty = G_1 B \circ \psi - G_2^t V.$$

Define T_1 *through (6) and* T_2 *through (9), with* $G = G_3$. *Then* $T \equiv T_2 \circ T_1$ *satisfies*

$$TR_\infty = B \circ \varphi \qquad \textit{in distribution.}$$

Proof: Apply Lemma 1, (7) and (8). □

We now briefly discuss further issues needed before Theorem 1 can be applied to a real data situation. Let R_n be one of the processes $\hat{\alpha}_n - \hat{\varepsilon}_n$ considered in the previous section, or any other marked empirical process admitting a limit R_∞ as given in (4). The next step to verify is that along with

$$R_n \to R_\infty$$

one has

$$TR_n \to TR_\infty = B \circ \varphi. \tag{10}$$

Finally, observe that typically T incorporates quantities which are unknown in practice and need to be estimated from the data. Hence we come up with a random operator T_n, for which it remains to show that

$$T_n R_n \to B \circ \varphi. \tag{11}$$

The proof of (10) and (11) requires some extra work and uses special properties of the underlying processes. It is therefore beyond the scope of the present paper. For the aforementioned examples technical details as well as simulation results may be found in the cited papers.

We finally discuss an application of (11) which is designed to derive tests for H_0 when the alternative is specified. As has been noted by Stute (1997) in the regression case, the Radon-Nikodym derivative of the underlying test process R_n w.r.t. the hypothesis and local alternatives may often be expressed, in the limit, in terms of the principal components of R_∞. While these are not readily available and some numerical work is required to approximate them, the transformed processes converge to a Brownian Motion, for which the principal components are readily available. In other words, Theorem 1 together with (11) may be used to yield optimal Neyman-Pearson tests for composite models when local alternatives are specified.

References

[1] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimators under random censorship. *Ann. Statist.* **2** 437-453.

[2] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277-281.

[3] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279-290.

[4] Gaenssler, P. and Stute, W. (1979). Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Prob.* **7** 193-243.

[5] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53** 457-481.

[6] Khmaladze, E.V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theor. Prob. Appl.* **26** 240-257.

[7] Koul, H.L. and Stute, W. (1997). Nonparametric model checks for time series. Preprint. Univ. of Giessen.

[8] Nikabadze, A. and Stute, W. (1997). Model checks under random censorship. *Statist. and Prob Letters* **32** 249-259.

[9] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley

[10] Stute, W., González Manteiga, W. and Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit tests. *Metrika* **40** 243-256.

[11] Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**. In press.

[12] Stute, W., Thies, S. and Zhu, L.X. (1996). Nonparametric model checks for regression. Preprint. Univ. of Giessen.