Volume 7, No. 1 (1998), pp. 60–77 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

GOODNESS-OF-FIT TESTS FOR PARAMETRIC HYPOTHESES ON THE DISTRIBUTION OF POINT PROCESSES

N. O. Maglaperidze[1], Z. P. Tsigroshvili[1]*, and M. van Pul[2]

[1]A. Razmadze Mathematical Institute of Georgian Academy of Sciences
M. Alexidze str., 1, 380093 Tbilisi, Georgia
e-mail: zurats@imath.acnet.ge

[2]Measurement and Computational Applications
Koninklijke/Shell-Laboratorium Amsterdam
P.O. Box 38000, 1030 BN Amsterdam, The Netherlands

We construct asymptotically distribution-free goodness-of-fit tests for testing parametric hypothesis on the intensities of point processes. The technique for obtaining asymptotically distributed empirical process $\tilde{W}_n$ is based on the theory of innovation martingales.

Key words: parametric hypotheses, innovation martingales, goodness-of-fit tests

AMS 1991 Subject Classification: Primary 62G10; Secondary 62M07, 60G44, 60G55.

0. Introduction

The aim pursued in this paper is to construct asymptotically distribution-free goodness-of-fit tests for testing parametric hypotheses on the distribution of point processes, i.e., our object is to derive goodness-of-fit tests for the hypotheses

$$\lambda_n(\cdot) \in \Lambda_n = \{\lambda_n(\cdot,\theta), \theta \in \Theta \subset R^m\} \tag{1}$$

on the intensity process $\lambda_n(\cdot)$ of a point process.

For the particular case of the general multiplicative model of Aalen (see Andersen, Borgan, Gill, Keiding (abbreviated below as ABGK) [1], p. 128, or Hjort [4])

*Research supported by the International Science Foundation, Grants MX1000 and MX1200.

when $\lambda_n(t) = \alpha(t)Y_n(t)$, where Y_n is an observable random process, it is sufficient to consider the hypothesis

$$\alpha(\cdot) \in A = \{\alpha(\cdot,\theta),\, \theta \in \Theta \subset R^m\}.$$

The hypothesis (1) means that there exist θ and a sequence of $\lambda_n \in \Lambda_n$ such that the process

$$N_n(t) - \int_0^t \lambda_n(s,\theta)\, ds$$

is a martingale; here and everywhere N_n denotes a sequence of point processes.

Later we will assume that there exists an increasing sequence of constants a_n such that

$$M_n(t,\theta) = M_n(t) = a_n^{-1}\Big(N_n(t) - \int_0^t \lambda_n(s,\theta)\, ds\Big)$$

is a square integrable martingale which under the hypothesis converges to a Gaussian martingale M with mean zero and variance

$$EM^2(t) = F(t,\theta).$$

For the definition of $F(t,\theta)$ and conditions on λ_n and a_n see Section 1.

If we knew the true value of the parameter θ, taking various functionals of the process M_n we could obtain asymptotically distribution-free goodness-of-fit tests like χ^2, ω^2, Kolmogorov–Smirnov, etc. However, in our case (when testing a parametric hypothesis), we do not know the true value of θ and it seems natural to consider the process

$$\widetilde{M}_n(t) = M_n(\cdot,\widetilde{\theta}) = a_n^{-1}\Big(N_n(t) - \int_0^t \lambda_n(s,\widetilde{\theta})\, ds\Big),$$

where $\widetilde{\theta}$ is an estimate of θ (see conditions (c5) in Section 1). However, unlike M_n the process $\widetilde{M}_n$ is no longer a martingale and moreover, the classical change of time $u = F(t,\theta)$ does not lead to an asymptotically distribution-free process: the limit distribution of $\widetilde{M}_n(F^{-1}(\cdot,\theta))$ depends on $F(\cdot,\theta)$, $\theta \in \Theta$. Therefore the traditional way to construct asymptotically distribution-free tests is no more valid.

In this paper we propose another way of avoiding this problem and obtaining asymptotically distribution-free nonparametric goodness-of-fit tests. The guideline in this work is the idea of Khmaladze [6, 7] to base goodness-of-fit tests not on the parametric empirical process (in the classical i.i.d. case) but on its innovation process (see also Nikabadze, Khmaladze [9]).

One can find the description of Khmaladze's idea in the context of point processes in the monograph ABGK [1], pp. 464–469. Among recent papers on asymptotically distribution-free tests for parametric hypotheses we mention Hjort [4]. The reader may discover a very direct connection between the χ^2 tests of Hjort [4] and Nikulin [10] and the innovation of the present paper. This connection will be described in more detail in Tsigroshvili [12] (to appear).

In Section 1 we introduce some notation and give conditions (c1)–(c5) which we need throughout the paper. Furthermore we describe the nature of the process M_n

and its limiting process L^l. In Section 2 the Doob–Meyer decomposition is given for the limiting process L^l as a linear transformation Q of this process and some of its properties are studied. In Section 3 we consider the transformation Q of the process M_n and prove a limit theorem for QM_n. In Section 4 we construct goodness-of-fit statistics in two cases: when the function $F(\cdot,\theta)$ is known (only θ is unknown) and when the function F is unknown. Finally, in Section 5 we present some examples of the situations described in Section 4.

1. Description of the Limiting Process for $M_n(\cdot,\widetilde{\theta})$

Consider the process

$$\widetilde{M}_n(t) = M_n(\cdot,\widetilde{\theta}) = a_n^{-1}\Big(N_n(t) - \int_0^t \lambda_n(s,\widetilde{\theta})\,ds\Big), \tag{2}$$

where $\widetilde{\theta}$ is some estimate of the true value of the parameter θ.

In this section we want to clarify the nature of this process and show that $M_n(\cdot,\widetilde{\theta})$ is asymptotically a projection of $M_n(\cdot,\theta)$ for a wide class of estimates $\widetilde{\theta}$ (see (c5) below).

First, let us formulate regularity assumptions on the intensity process $\lambda_n(\cdot,\theta)$:

(c1) There exists a neighbourhood Θ_0 of θ such that for all $t\in[0,T]$ there exist first and second continuous derivatives of $\lambda_n(\cdot,\theta)$ and $\log\lambda_n(\cdot,\theta)$ w.r.t. θ (see also ABGK [1], p. 420).

(c2) There exist functions $f(t,\theta)$ and $h_1(t,\theta),\dots,h_m(t,\theta)$ and a sequence of increasing nonnegative constants $\{a_n\}, n=1,2,\dots$, such that for all $\varepsilon>0$ and for all $\theta,\bar{\theta}\in\Theta_0$

$$\left\|a_n^{-2}\frac{\lambda_n(\cdot,\theta)}{f(\cdot,\theta)} - 1\right\|_0^{T-\varepsilon} = o_p(1), \qquad \left\|a_n^{-2}\frac{\lambda_n(\cdot,\theta)}{f(\cdot,\theta)} - 1\right\|_{T-\varepsilon}^{T} = O_p(1),$$

$$\left\|a_n^{-2}\frac{\frac{\partial}{\partial\theta}\lambda_n(s,\bar{\theta}) - \frac{\partial}{\partial\theta}\lambda_n(s,\theta)}{f(\cdot,\theta)}\right\|_0^T = o_p(1) \qquad \text{as} \quad n\to\infty.$$

(c3) for all $i=1,2,\dots,m$

$$\left\|\frac{\partial}{\partial\theta_i}\log\lambda_n(\cdot,\theta) - h_i(\cdot,\theta)\right\|_0^T = o_p(1) \qquad \text{as} \quad n\to\infty.$$

Here and everywhere $\|f(\cdot)\|_a^b$ is the sup norm of the real-valued function f considered on the interval $[a,b]$.

Introduce some notation:

$$F_n(t,\theta) = a_n^{-2}\int_0^t \lambda_n(s,\theta)ds, \quad h_n(s,\theta) = \frac{\partial}{\partial\theta}\log\lambda_n(s,\theta), \quad F(t,\theta) = \int_0^t f(s,\theta)ds,$$

$$C_n(t,\theta) = \int_t^T h(s,\theta)h^T(s,\theta)dF_n(s,\theta), \quad C(t,\theta) = \int_t^T h(s,\theta)h^T(s,\theta)dF(s,\theta).$$

The superscript T denotes the transposition. We assume that

(c4) The matrix $C(t,\theta)$ is positive definite and finite and the matrix $C_n(t,\theta)$ is nonnegative definite and finite for all $t < T$ and $\theta \in \Theta$ if and only if $\lambda_n(t,\theta) \neq 0$.

To derive the asymptotic form of the process for $M_n(\cdot,\widetilde{\theta})$ we need the following assumptions about the estimate $\widetilde{\theta}$:

(c5) There exists an m-dimensional vector-function $l(\cdot,\theta)$ such that:

$$\begin{aligned} &\int_0^T l^T(s,\theta)l(s,\theta)dF_n(s,\theta) < \infty \quad P\text{-a.s.}, \\ &\int_0^T l(s,\theta)h^T(s,\theta)dF(s,\theta) = I_m, \end{aligned} \tag{3}$$

with I_m denoting the m-dimensional identity matrix, for which $\widetilde{\theta}$ admits the representation:

$$a_n(\widetilde{\theta}-\theta) = \int_0^T l(s,\theta)dM_n(s,\theta) + o_p(1). \tag{4}$$

The estimates with such properties can be exemplified by M-estimates, i.e., by the estimates which are solutions of the equation:

$$\int_0^T \Psi_n(s,\theta)dM_n(s,\theta) = 0,$$

where $\Psi_n(\cdot,\theta)$ is an m-dimensional vector-function square integrable w.r.t. $F_n(\cdot,\theta)$. In particular, the vector-function $\Psi_n(\cdot,\theta) = h_n(\cdot,\theta)$ corresponds to the maximum likelihood estimate and in this case, as it is easy to show, the representation (4) takes the form:

$$a_n(\widehat{\theta}-\theta) = C^{-1}(0,\theta)\int_0^T h(s,\theta)dM_n(s,\theta) + o_p(1).$$

We use the notation $\widehat{\theta}$ for the maximum likelihood estimate, using $\widetilde{\theta}$ for a general estimate satisfying (c5).

By the Taylor series expansion around the true value of θ we get

$$M_n(t,\widetilde{\theta}) = M_n(t,\theta) - a_n(\widetilde{\theta}-\theta)^T\int_0^t h_n(s,\theta)\,dF_n(s,\theta) + o_p(1),$$

and under conditions (c1)–(c5) one can easily prove the following theorem (for the proof see Hjort [4] or ABGK [1], p. 457).

Theorem 1. *If conditions* (c1)–(c5) *are fulfilled, then on* $D[0,T]$

$$M_n(\cdot,\widetilde{\theta}) \xrightarrow{D} L^l(\cdot), \qquad M_n(\cdot,\widehat{\theta}) \xrightarrow{D} L(\cdot), \qquad n\to\infty,$$

where

$$L^l(t) = M(t) - \int_0^t h^T(s)\,dF(s)\int_0^T l(s)\,M(ds) \tag{5}$$

and

$$L(t) = M(t) - \int_0^t h^T(s)\, dF(s) C^{-1}(0) \int_0^T h(s)\, M(ds) \tag{6}$$

are Gaussian processes with zero mean and covariance functions

$$\begin{aligned} R_{L^l}(t,u) = F(t\wedge u) &- \int_0^t h^T(s)\, dF(s) \int_0^u l(s)\, dF(s) \\ &- \int_0^u h^T(s)\, dF(s) \int_0^t l(s)\, dF(s) \\ &+ \int_0^t h^T(s)\, dF(s) \int_0^T l(s) l^T(s)\, dF(s) \int_0^u h^T(s)\, dF(s); \end{aligned}$$

and

$$R_L(t,u) = F(t\wedge u) - \int_0^t h^T(s)\, dF(s) C^{-1}(0) \int_0^u h(s)\, dF(s).$$

Here M is a Wiener process with covariance function $E\Big(M(t)M(u)\Big) = F(t\wedge u)$, and $\xrightarrow{D}$ denotes the convergence in distribution. All functions above depend on the parameter θ but for the sake of brevity we omit it from notation. Hopefully this will not lead to confusion.

Denote

$$\int_0^t h(s)\, dF(s) = g_t$$

and

$$\widetilde{\Pi} M_t = L^l(t), \qquad \widehat{\Pi} M_t = L(t),$$

so that we have

$$\widetilde{\Pi} M_t = M_t - g_t^T \int_0^T l(s)\, dM(s), \tag{7}$$

$$\widehat{\Pi} M_t = M_t - g_t^T C^{-1}(0) \int_0^T h(s)\, dM(s). \tag{8}$$

Introduce the class $C^2[0,T]$ of functions $\phi \in D[0,T]$ such that the derivative w.r.t. $F(\cdot)$ exists and

$$\Big(\|\phi\|_c\Big)^2 \equiv \int_0^T \left(\frac{d\phi(s)}{dF(s)}\right)^2 dF(s) < \infty.$$

Then, as it is well known, the bilinear functional on $C^2[0,T] \times D[0,T]$ of the form

$$\langle \phi; X\rangle = \int_0^T \left(\frac{d\phi(s)}{dF(s)}\right) dX(s) \tag{9}$$

is well defined for almost all trajectories X of the Gaussian martingale M, while on $C^2[0,T] \times C^2[0,T]$ this functional is a scalar product. Now, we are ready to describe the structure of the limiting processes (7) and (8):

Lemma 1. *The transformation $\widetilde{\Pi}M_t$ is a projection and the transformation $\widehat{\Pi}M_t$ is an orthogonal projection if and only if* (3) *holds. Projections are parallel to the function $g(t)$ on the subspace which is orthogonal w.r.t.* (9) *to the function $\int_0^t l(s)\,dF(s)$ or $\int_0^t h(s)\,dF(s)$ respectively.*

Proof. Indeed, for all $\beta \in R^m$

$$\widetilde{\Pi}g_t^T\beta = g_t^T\beta - g_t^T\int_0^T l(s)h^T(s)\,dF(s)\beta = 0,$$

i.e.,

$$\widetilde{\Pi}\widetilde{\Pi}M_t = \widetilde{\Pi}M_t \qquad \text{iff} \quad \int_0^T l(s)h^T(s)\,dF(s) = I_m$$

and

$$\begin{aligned}\Big\langle \beta^T\int_0^{\cdot} l(s)\,dF(s)\,;\,\widetilde{\Pi}M(\cdot)\Big\rangle &= \beta^T\int_0^T l(s)\,dM(s)\\ &\quad - \beta^T\int_0^T l(s)h^T(s)\,dF(s)\int_0^T l(s)\,dM(s) = 0. \quad \square\end{aligned}$$

Note that in the case of $\widehat{\Pi}$, the processes $\widehat{\Pi}M_t$ and $g^T(t)C^{-1}(0)\int_0^T h(s)\,dM(s)$ are independent. By the nature of this lemma we will call the estimates which satisfy (c5) "projecting" estimates (see Khmaladze [5]).

2. Doob–Meyer Decomposition for the Limiting Processes and Its Properties

In this section we derive a one-to-one transformation Q of the processes L^l and L into a Wiener process and give some properties of this transformation.

Consider the filtration $\{\mathcal{H}_t\}_{0\le t\le T}$,

$$\mathcal{H}_t = \mathcal{F}_t^M \vee \sigma\Big\{\int_0^T h(s)\,dM(s)\Big\} = \mathcal{F}_t^M \vee \sigma\Big\{\int_t^T h(s)\,dM(s)\Big\}. \tag{10}$$

We would like to derive the Doob–Meyer decomposition for the process $M(\cdot)$ w.r.t. $\{\mathcal{H}_t\}_{0\le t\le T}$. To do this, we proceed as follows: since $M(dt)$ and $\int_t^T h(s)dM(s)$ are independent of $\mathcal{F}_t^M$, by redundant conditioning (see, e.g., Bremaud [3], p. 281) we have

$$E(M(dt)\mid\mathcal{H}_t) = E\Big[M(dt)\,\Big|\int_t^T h(s)dM(s)\Big].$$

Now using the fact that for all $t \in [0,T]$, $M(dt)$ and $\int_t^T h(s)\,dM(s)$ are jointly Gaussian random variables and

$$E\Big[M(dt)\,\Big|\int_t^T h(s)\,dM(s)\Big] = h^T(t)C^{-1}(t)\int_t^T h(s)\,dM(s)\,dF(t),$$

we obtain the following Doob–Meyer decomposition for $M(\cdot)$:

$$W(t) = M(t) - \int_0^t h^T(s)C^{-1}(s)\int_s^T h(u)\,dM(u)\,dF(s). \tag{11}$$

Equation (11) defines a linear transformation Q of M into W. Let us apply the same transformation to processes L^l and L. Then the following equalities hold.

Property 1. $QM(t) = QL(t) = QL^l(t)$.

Proof. The proof of these equalities would follow from

$$QX(t) = 0 \qquad \text{iff} \quad \frac{dX(t)}{dF(t)} = h^T(t)\beta \quad \text{for all } \beta \in R^m.$$

To prove this fact, denote

$$V(t) = \int_t^T h(s)\,dX(s),$$

so that $QX(t) = 0$ is equivalent to

$$-dV(t) = h(t)h^T(t)C^{-1}(t)V(t)dF(t)$$

or

$$d[C^{-1}(t)V(t)] = 0,$$

which leads to $C^{-1}(t)V(t) = \beta$ or $V(t) = C(t)\beta$. This equality is valid if and only if $\frac{dX(t)}{dF(t)} = h^T(t)\beta$. □

Note that this property clarifies our choice of the filtration. As a consequence of it, an empirical analogue of $QL^l(\cdot)$ (which we will consider in the following section) will be independent of the estimate, and will have the same form for all projecting estimates. This is certainly an advantage from the computational viewpoint.

By the following property we obtain a one-to-one relationship between the processes $L(\cdot)$ and $QL(\cdot)$.

Property 2. *If condition* (c4) *is satisfied, then the process*

$$W(t) = L(t) - \int_0^t h^T(s)C^{-1}(s)\int_s^T h(u)\,dL(u)\,dF(s) \tag{12}$$

is a Wiener process w.r.t. the filtration $\{\mathcal{H}_t\}_{0\le t\le T}$ *with*

$$E\big(W(t)W(u)\big) = F(t\wedge u),$$

and (12) *has a unique inverse of the form*

$$L(t) = W(t) + \int_0^t h^T(s)\int_0^s C^{-1}(u)h(u)dW(u)dF(s).$$

Proof. Being a linear transform of the Gaussian process $L(\cdot)$, the process $W(\cdot)$ is also Gaussian. A direct calculation gives us

$$E\big(W(t)W(u)\big) = F(t \wedge u).$$

To prove that L is the only inverse on the space of functions orthogonal to $g(\cdot)$ w.r.t. (9), it is sufficient to show that the kernel of the transformation Q contains only functions of the form $g^T(t)\beta$ for all $\beta \in R^m$. However, this is the same as what was shown in the proof of Property 1. □

Rewrite transformation (12) in the following form:

$$W(t) = \int_0^T k(t,s)dL(s) = \int_0^T \big(I(s \le t) - r(t,s)\big)\, dL(s), \tag{13}$$

where

$$r(t,s) = \int_0^{t\wedge s} h^T(u)C^{-1}(u)\, dF(u)h(s) \tag{14}$$

and $I(A)$ denotes the indicator of the event A.

Now we state some easily verified properties of the kernel $r(t,s)$ (cf. Khmaladze [6]):

Property 3. *If the kernel $r(t,s)$ is defined by (14), then*

$$\int_0^T r^2(t,s)\, dF(s) = 2\int_0^t r(t,s)\, dF(s).$$

Proof. It is an easy consequence of the fact that

$$\int_0^T k^2(t,s)dF(s) = F(t). \quad \square$$

Property 4. *$r(t,s)$ is a Hilbert-Schmidt kernel.*

Proof. From Property 3 by the Cauchy–Schwarz inequality we have

$$\int_0^T r^2(t,s)\, dF(s) = 2\int_0^t r(t,s)\, dF(s) \le 2(F(t))^{1/2}\Big(\int_0^T r^2(t,s)dF(s)\Big)^{1/2},$$

and therefore

$$\int_0^T r^2(t,s)dF(s) \le 4F(t),$$

so that finally

$$\int_0^T\int_0^T r^2(t,s)\, dF(s)dF(t) \le 2F^2(T) < \infty. \quad \square$$

Property 5. *On the functional space $(C^2[0,T]; \|\cdot\|_c)$ the linear transformation Q is continuous.*

Proof. Consider

$$\|Q\phi\|_c = \Big(\int_0^T \Big(\frac{dQ\phi(s)}{dF(s)}\Big)^2 dF(s)\Big)^{1/2} \tag{15}$$

$$= \Big(\int_0^T \Big(\frac{d\phi(s)}{dF(s)} - h^T(s)C^{-1}(s)\int_s^T h(u)\,d\phi(u)\Big)^2 dF(s)\Big)^{1/2}$$

$$= \Big(\|\phi\|_c^2 + \int_0^T h^T(u)\,d\phi(u)C^{-1}(0)\int_0^T h(u)\,d\phi(u)\Big)^{1/2}.$$

Consider now the quadratic form

$$\psi(s) = \int_s^T h^T(u)\,d\phi(u)C^{-1}(s)\int_s^T h(u)\,d\phi(u).$$

By (c5), the matrix $C(s)$ is positive definite, so $\psi(s)$ is positive for all $s \in [0,T]$, and we have

$$\psi(s) = \int_s^T \Big(\int_s^T h^T(v)\,d\phi(v)C^{-1}(s)h(u)\Big)\,d\phi(u)$$

$$\le \Big(\int_s^T \Big(\int_s^T h^T(v)\,d\phi(v)C^{-1}(s)h(u)\Big)^2 d\phi(u)\Big)^{1/2}\Big(\int_s^T \Big(\frac{d\phi(u)}{dF(u)}\Big)^2 dF(u)\Big)^{1/2}$$

$$= (\psi(s))^{1/2}\Big(\int_s^T \Big(\frac{d\phi(u)}{dF(u)}\Big)^2 dF(u)\Big)^{1/2},$$

i.e.,

$$\psi(s) \le \int_s^T \Big(\frac{d\phi(u)}{dF(u)}\Big)^2 dF(u)$$

for all $s \in [0,T]$ and therefore $\psi(T) = 0$.

Now from (15) we have

$$\|Q\|_c = \sup_{\phi\in c^2[0,T]} \frac{\|Q\phi\|_c}{\|\phi\|_c} \le \sqrt{2}.$$

So the linear transformation Q is bounded on $C^2[0,T]$. □

3. Weak Convergence of The Process $QM_n(\cdot,\widetilde{\theta})$.

In this section we prove that under the hypothesis the transformation Q of $M_n(\cdot,\widetilde{\theta})$ converges in distribution in $D[0,T]$ to a Wiener process W with covariance function $E(W(t)W(u)) = F(t\wedge u)$.

Consider the process

$$W_n(\cdot,\widetilde{\theta}) = M_n(\cdot,\widetilde{\theta}) - \int_0^t h^T(s)C^{-1}(s)\int_s^T h(u)\,dM_n(u,\widetilde{\theta})\,dF(s), \tag{16}$$

or equivalently,

$$(17) \qquad W_n(t,\widetilde{\theta}) = \int_0^T k(t,s)\,dM_n(s,\widetilde{\theta}) = \int_0^T (I(s \le t) - r(t,s))\,dM_n(s,\widetilde{\theta}).$$

Theorem 2. *Under the conditions* (c1)–(c5), *when the hypothesis holds,*

$$W_n(\cdot,\widetilde{\theta}) \xrightarrow{D} W(\cdot)$$

on the space $D[0,T]$.

To prove this theorem we need an auxiliary lemma.

Lemma 2. *Let* $\phi_n(\cdot) \in C^2[0,T]$ *and* $\|\phi_n(\cdot)\|_c \to 0$, *as* $n \to \infty$. *Then*

$$\|Q\phi_n(\cdot)\|_0^T \to 0.$$

Proof. By Property 5 of the previous section, $\|\phi_n(\cdot)\|_c \to 0$ implies $\|Q\phi_n(\cdot)\|_c = o(1)$, i.e.,

$$\int_0^T \Big(\frac{d\phi_n(s)}{dF(s)}\Big)^2 dF(s) \to 0 \qquad \text{as} \quad n \to \infty.$$

Now

$$\|Q\phi_n(\cdot)\|_0^t = \left\| \int_0^T \Big(\frac{dQ\phi_n(s)}{dF(s)}\Big)\,dF(s)\right\|_0^T \le \|Q\phi_n(\cdot)\|_c F(T)^{1/2} \to 0$$

as $n \to \infty$. □

Proof of Theorem 2. By the Taylor expansion

$$M_n(t,\widetilde{\theta}) = M_n(t,\theta) - a_n(\widetilde{\theta}-\theta)^T \int_0^t h_n(s,\theta)\,dF_n(s,\theta) + \varepsilon_n(t),$$

where

$$\varepsilon_n(t) = a_n(\widetilde{\theta}-\theta)^T a_n^{-2} \int_0^t \Big(\frac{\partial}{\partial\theta}\lambda_n(s,\theta) - \frac{\partial}{\partial\theta}\lambda_n(s,\bar{\theta})\Big)\,ds$$

and $\bar{\theta}$ is some point between θ and $\widetilde{\theta}$.

By the linearity of Q and the fact that

$$Q\Big(\int_0^t h(s,\theta)\,dF(s,\theta)\Big) = 0$$

we have

$$W_n(t,\widetilde{\theta}) = W_n(t,\theta) + Q\varepsilon_n(t) - a_n(\widetilde{\theta}-\theta)^T Q\Big(\int_0^t h_n(s,\theta)\,dF_n(s,\theta) - \int_0^t h(s,\theta)\,dF(s,\theta)\Big).$$

Here $W_n(t,\theta) = QM_n(t,\theta)$ is a martingale w.r.t. $\{\mathcal{F}_t\}_{0\le t\le T}$ with jumps a_n^{-1} and quadratic variation $\langle QM_n\rangle(t,\theta) = F_n(t,\theta)$ which converges in probability to $F(t,\theta)$ for all $t \in [0,T]$, i.e., the conditions of the CLT for martingales are fulfilled (see, e.g., Liptser and Shiryaev [8]) and so, on the space $D[0,T]$, we have

$$QM_n(\cdot,\theta) \xrightarrow{D} QM(\cdot).$$

By the condition (c2) both

$$a_n(\widetilde{\theta}-\theta)^T \Big(\int_0^t h_n(s,\theta)\, dF_n(s,\theta) - \int_0^t h(s,\theta)\, dF(s,\theta) \Big)$$

and $\varepsilon_n(t)$ are the functions from $C^2[0,T]$ converging in probability in $\|\cdot\|_c$-norm to 0 and an application of Lemma 2 gives the assertion. □

4. Construction of The Goodness-of-Fit Statistics

The kernel of the transformation (16) depends on the limiting distribution function F and also on the unknown value of parameter $\theta, r(t,s) = r(t,s,\theta)$ (or $k(t,s) = k(t,s,\theta)$). Since the true value of θ remains unknown we cannot use (17) directly for the testing purposes. It is possible that the limiting function $F(\cdot,\theta)$ is also unknown. As an example of such situation we consider the Aalen's general multiplicative model (see Section 5 below). In this section we will show how to construct goodness-of-fit statistics in both cases.

First, let us consider the situation when $F(\cdot,\theta)$ is known and only θ is unknown. An example of such situation is also given in Section 5, this is Musa's model of ABGK [1], p. 133, or van Pul [12]. In this case it seems natural to replace θ by any a_n-consistent estimate (like $\widetilde{\theta}$ or $\widehat{\theta}$) in kernel $k(\cdot,\cdot,\theta)$ and consider the process

$$\widetilde{W}_n(t) = \int_0^T k(t,s,\widetilde{\theta})\, dM_n(s,\widetilde{\theta}). \tag{18}$$

One can see that

$$W_n(t,\widetilde{\theta}) - \widetilde{W}_n(t) = \int_0^T \big(r(t,s,\widetilde{\theta}) - r(t,s,\theta)\big)\, dM_n(s,\widetilde{\theta}), \tag{19}$$

where $W_n(t,\widetilde{\theta})$ is given by (17). To prove that

$$\widetilde{W}_n(\cdot) \xrightarrow{D} W(\cdot)$$

on $D[0,T]$, we need some smoothness conditions on $r(t,\cdot,\theta)$ w.r.t. θ for all $t < T$. These conditions are given below (see (c6) and Remark 3).

If now $F(t,\theta)$ is an unknown function (i.e., the limiting $\lambda(t,\theta)$ is unknown), it is reasonable to use the kernel $r_n(t,s,\theta)$ instead of $r(t,s,\theta)$, where

$$r_n(t,s,\theta) = \int_0^{t\wedge s} h^T(u,\theta) C_n^{-1}(u,\theta)\, dF_n(u,\theta) h(s,\theta) \tag{20}$$

(and plug in $\widetilde{\theta}$ for θ). Hence, it seems that the following process is suitable in this case:

$$\widetilde{W}_n(t) = \int_0^T k_n(t,s,\widetilde{\theta})\, dM_n(s,\widetilde{\theta}) = Q_n M_n(t,\widetilde{\theta}), \tag{21}$$

where

$$k_n(t,s,\theta) = I(s \le t) - r_n(t,s,\theta).$$

However, it is possibile to simplify the transformation Q first. Instead of calculating

$$E\big(M(dt) \mid \mathcal{H}_t\big) = E\Big(M(dt) \Big| \int_t^T h(s) dM(s)\Big)$$

in Section 2, let us consider the $(m+1)$-dimensional vector-function $H^T(\cdot) = (1, h^T(\cdot))$. Then

$$E\Big[M(dt) \Big| \int_t^T H(s)\, dM(s)\Big] = H^T(t)\mathcal{C}^{-1}(t) \int_t^T H(u)\, dM(u) dF(t) \tag{22}$$

and instead of (11) of Section 2 we obtain the process

$$W(t) = M(t) - \int_0^t H^T(s)\mathcal{C}^{-1}(s) \int_s^T H(u)\, dM(u) dF(s) = QM(t), \tag{23}$$

or equivalently,

$$W(t) = \int_0^T K(t,s)\, dM(s) = QM(t) \tag{24}$$

with

$$K(t,s) = I\{s \le t\} - R(t,s)$$

and

$$R(t,s,\theta) = \int_0^{t\wedge s} H^T(u)\mathcal{C}^{-1}(u)\, dF(u) H(s). \tag{25}$$

Here $\mathcal{C}^{-1}(t)$ is the inverse of the $(m+1) \times (m+1)$ matrix

$$\mathcal{C}(t,\theta) = \int_t^T H(s)H^T(s)\, dF(s) = \int_t^T \begin{pmatrix} 1 & h^T(s) \\ h(s) & h(s)h^T(s) \end{pmatrix} dF(s).$$

It is easy to show that $W(t)$ still has all properties of Section 2. At the same time the pre-limiting process has a computational advantage. Namely, as a suitable process to base the testing procedure on we will take the process

$$\widetilde{W}_n(t) = \int_0^T K_n(t,s,\widetilde{\theta})\, dM_n(s,\widetilde{\theta}) = Q_n M_n(t,\widetilde{\theta}) \tag{26}$$

with

$$K_n(t,s,\theta) = I(s \le t) - R_n(t,s,\theta)$$

and

$$R_n(t,s,\theta) = \int_0^{t\wedge s} H^T(u)\mathcal{C}_n^{-1}(u)\,dF_n(u)H(s), \tag{27}$$

where

$$\mathcal{C}_n(t,\theta) = \int_t^T H(s)H^T(s)dF_n(s) = \int_t^T \begin{pmatrix} 1 & h^T(s) \\ h(s) & h(s)h^T(s) \end{pmatrix} dF_n(s).$$

Then unlike the process defined by (21), the process (26) has the following

Property 6.

$$\widetilde{W}_n(t) = a_n^{-1}\int_0^T K_n(t,s,\widetilde{\theta})\,dN_n(s) = a_n^{-1}Q_nN_n(t).$$

Proof. The kernel of transformation Q_n contains the functions of the form $\int_0^t \eta(s)\,dF_n(s,\widetilde{\theta})$, with $\eta(s) = H^T(t,s,\widetilde{\theta})\xi$ for any $(m+1)$-dimensional vector ξ. If we choose $\xi^T = (1,0,\dots,0)$, then we get

$$\int_0^t H^T(t,s,\widetilde{\theta})\,dF_n(s,\widetilde{\theta})\xi = F_n(t,\widetilde{\theta}), \tag{28}$$

which means that Q_n annulates the compensator of N_n. □

Remark 1. In order R_n in (27) to be well defined, we assume that

$$\mathcal{C}_n^{-1}(u)\lambda_n(u) = \Big(\int_u^T H(s)H^T(s)\lambda_n(s)\,ds\Big)^{-1}\lambda_n(u) = 0 \quad \text{iff} \quad \mathcal{C}_n(u) = 0.$$

By Property 6 we see finally that the test statistic can be based on the process

$$\widetilde{W}_n(t) = a_n^{-1}\Big(N_n(t) - \int_0^T R_n(t,s,\widetilde{\theta})\,dN_n(s)\Big). \tag{29}$$

Now we formulate the smoothness assumptions on $R_n(t,\tau,\theta)$ w.r.t. parameter θ.
(c6) For $n\to\infty$, $\varepsilon\to 0$

$$a_n^{-2}\int_0^T \sup_{\|\vartheta-\theta\|_E<\varepsilon}\Big|\frac{\partial}{\partial\vartheta_i}R_n(t,\tau,\vartheta) - \frac{\partial}{\partial\theta_i}R_n(t,\tau,\theta)\Big|\,dN_n(\tau) = o_p(1) \tag{30}$$

and

$$\int_0^T \sup_{0\le t\le T}\Big|\frac{\partial}{\partial\theta_i}R_n(t,\tau,\theta)dN_n(\tau)\Big|\,dN_n(\tau) = O_p(1). \tag{31}$$

Here $||\vartheta - \theta||_E$ denotes the Euclidean norm in R^m.

In Section 5 we will see what these rather traditional conditions look like in examples, while at the end of this section we prove the following

Theorem 3. *If the conditions* (c1)–(c6) *are fulfilled, then, under the hypothesis, on the functional space* $D[0,T]$

$$\widetilde{W}_n(\cdot) \xrightarrow{D} W(\cdot),$$

where $W(\cdot)$ *is a Wiener process with covariance function*

$$E(W(t)W(u)) = F(t \wedge u).$$

To prove this theorem we need an auxiliary lemma.

Lemma 3. *If the assumption* (c2) *holds, then for all* $\varepsilon > 0$,

$$\sup_{0 \le t \le T-\varepsilon} \left| \mathcal{C}_n^{-1}(s) \left[\frac{a_n^{-2}\lambda_n(s)}{f(s)} \right] - \mathcal{C}^{-1}(s) \right| = o_p(1) \qquad \text{as} \quad n \to \infty.$$

Proof. Consider the (i,j)th component of the matrix $\mathcal{C}_n(u)$,

$$\begin{aligned} \int_u^T H_i(s)H_j(s)a_n^{-2}\lambda_n(s)\,ds &= \int_u^T H_i(s)H_j(s)\frac{a_n^{-2}\lambda_n(s)}{f(s)}\,dF(s) \\ &= \int_u^{T-\varepsilon} H_i(s)H_j(s)\left[\frac{a_n^{-2}\lambda_n(s)}{f(s)} - 1\right] dF(s) \\ &\quad + \int_u^{T-\varepsilon} H_i(s)H_j(s)\,dF(s) \\ &\quad + \int_{T-\varepsilon}^T H_i(s)H_j(s)\left[\frac{a_n^{-2}\lambda_n(s)}{f(s)}\right] dF(s). \end{aligned}$$

Now by (c2) the first and third summands in the right-hand side are small, while the second summand tends to the (i,j)th component of the matrix $\mathcal{C}(u)$, whenever ε is small. Therefore all components of the matrix $\mathcal{C}_n$ are close to the corresponding components of the matrix $\mathcal{C}$, and hence the determinants and all corresponding minors of these matrices are close to each other, i.e., we have

$$\sup_{0 \le t \le T-\varepsilon} |\mathcal{C}_n^{-1}(s) - \mathcal{C}^{-1}(s)| = o_p(1) \qquad \text{as} \quad n \to \infty.$$

An application of assumption (c2) again gives the assertion. □

Proof of Theorem 3. By the Taylor series expansion we obtain using Property 6

$$\widetilde{W}_n(t) = a_n^{-1}\left[N_n(t) - \int_0^T R_n(t,s,\widetilde{\theta})\,dN_n(s)\right]$$

$$= M_n(t) - \int_0^T R_n(t,s,\theta)\, dM_n(s)$$

$$- a_n(\widetilde{\theta} - \theta)^T a_n^{-2} \int_0^T \frac{\partial}{\partial\theta} R_n(t,s,\bar{\theta})\, dN_n(s) + o_p(1)$$

$$= QM_n(t) - \int_0^T [R_n(t,s,\theta) - R(t,s,\theta)]\, dM_n(s)$$

$$- a_n(\widetilde{\theta} - \theta)^T a_n^{-2} \int_0^T \frac{\partial}{\partial\theta} R_n(t,s,\bar{\theta}) dN_n(s) + o_p(1).$$

By Property 5 of Section 2, $M_n(\cdot)$ converges to the Wiener process and it is sufficient to show that the other two terms are small (see, e.g., Billingsley [2], Theorem 4.1). But

$$\int_0^T (R_n(t,s,\theta) - R(t,s,\theta)) dM_n(s)$$

$$= \int_0^t H^T(s)\Big(C_n^{-1}(s)\frac{a_n^{-2}\lambda_n(s)}{f(s)} - C^{-1}(s)\Big) \int_s^T H(u)\, dM_n(u) dF(s)$$

and by Lemma 3 it suffices to consider only $t > T - \varepsilon$ and to show that the random variables

$$\int_{T-\varepsilon}^T \Big| H^T(s) C_n^{-1}(s) \int_s^T H(u)\, dM_n(u) \Big|\, dF_n(s)$$

and

$$\int_{T-\varepsilon}^T \Big| H^T(s) C^{-1}(s) \int_s^T H(u) dM_n(u) \Big|\, dF(s)$$

are small in probability. But the first term is small because of Remark 1 above. To show that the second term is small, we only need to apply Chebyshev's inequality.

Now consider

$$a_n^{-2} \int_0^T \frac{\partial}{\partial\theta} R_n(t,s,\bar{\theta}) dN_n(s) = a^{-2} \int_0^T \Big(\frac{\partial}{\partial\theta} R_n(t,s,\bar{\theta}) - \frac{\partial}{\partial\theta} R_n(t,s,\theta)\Big)\, dN_n(s)$$

$$+ a_n^{-1} \int_0^T \frac{\partial}{\partial\theta} R_n(t,s,\theta)\, dM_n(s) + a_n^{-1} \int_0^T \frac{\partial}{\partial\theta} R_n(t,s,\theta)\, dF_n(s,\theta).$$

Assumption (c6) guarantees that the first two integrals in the right-hand side are small in sup-norm, while the third integral is equal to 0. Indeed, by Property 6, $Q_n F_n(t,\theta) = 0$, or

$$\frac{\partial}{\partial\theta} F_n(t,\theta) = \int_0^T \frac{\partial}{\partial\theta} R_n(t,s,\theta)\, dF_n(s,\theta) + a_n^{-2} \int_0^T R_n(t,s,\theta) \frac{\partial}{\partial\theta} \lambda_n(s,\theta)\, ds$$

with

$$\frac{\partial}{\partial\theta} F_n(t,\theta) = \int_0^t h_n(s,\theta)\, dF_n(s,\theta),$$

so that

$$\int_0^T \frac{\partial}{\partial \theta} R_n(s,\theta)\, dF_n(s,\theta) = Q_n\Big(\int_0^t h_n(s,\theta)\, dF_n(s,\theta)\Big) = 0. \quad \square$$

Remark 2. Without condition (c6) we can still prove that

$$\widetilde{W}_n(\cdot) \xrightarrow{D} W(\cdot)$$

on $D[0, T-\varepsilon]$ for any $\varepsilon > 0$.

Remark 3. In the case of a known limiting $F(t,\theta)$ we need $R(t,\tau,\theta)$ to satisfy (30) and (31), and also (30) with $a_n^{-2} dN_n(\tau)$ replaced by $dF(\tau,\theta)$. This additional requirement is needed because now $QF_n(t,\theta) \neq 0$. Thus we have an additional condition, but on the essentially simpler kernel.

5. Examples

In this section we will derive the forms of the transformations Q for two different models: for the so-called Musa's model, in which the limiting function F is known, and for Aalen's multiplicative model, where this limiting function is unknown.

• Musa's model. In this model $a_n = n^{1/2}$ and the intensity of the point process is

$$\lambda_n(t,\theta_1,\theta_2) = n\theta_1\big(\theta_2 - n^{-1}N_n(t^-)\big)$$

(see ABGK [1], p. 133, also van Pul [11]). According to this model the larger number $N(t)$ of "errors" we have detected, the smaller is the intensity of detecting a new one.

It is easy to see that in this case

$$h_n^T(t,\theta_1,\theta_2) = \Big(\theta_1^{-1}, \big(\theta_2 - n^{-1}N_n(t^-)\big)^{-1}\Big)$$

and

$$h(t,\theta_1,\theta_2) = \big(\theta_1^{-1}, (\theta_2 - F(t,\theta_1,\theta_2))\big),$$

where the limiting function F is known and has quite a simple form:

$$F(t,\theta_1,\theta_2) = \theta_2\big(1 - \exp(-\theta_1 t)\big).$$

The kernel $R(t,\tau,\theta_1,\theta_2)$ of the transformation Q in this model has the following form:

$$R(t,\tau,\theta_1,\theta_2) = \theta_1\Big(\int_0^{t\wedge\tau} G_1(\mu(u))\, du + e^{-\mu(\tau)} \int_0^{t\wedge\tau} G_2(\mu(u))\, du\Big),$$

where

$$G_1(u) = \frac{e^u - 1 - u}{\big(e^{u/2} - e^{-u/2}\big)^2 - u^2}, \qquad G_2(u) = \frac{e^u - 1 - ue^u}{\big(e^{u/2} - e^{-u/2}\big)^2 - u^2},$$

and

$$\mu(u) = \mu(u, \theta_1) = \theta_1(T - u).$$

Now we see that for all $t < T$ and $\tau < T$ the function $R(t, \tau, \theta_1, \theta_2)$ does not depend on θ_2 and is sufficiently smooth w.r.t. parameter θ_1,

$$\frac{\partial}{\partial\theta_1} R(t, \tau, \theta_1, \theta_2) = \frac{1}{\theta_1}\Big[\mu(0)(G_1(\mu(0)) + e^{-\mu\tau}G_2(\mu(0)))$$
$$- \mu(t \wedge \tau)\Big(G_1(\mu(t \wedge \tau)) + e^{-\mu\tau}G_2(\mu(t \wedge \tau))\Big) - \mu(\tau)e^{-\mu\tau}\int_{\mu\tau}^{\mu(0)} G_2(x)\, dx\Big].$$

Clearly, condition (c6) is satisfied for $R(t, \tau, \theta_1, \theta_2)$.

• Aalen's multiplicative model. In this model $a_n = n^{1/2}$ again and the intensity of the point process $N_n(t)$ is

$$\lambda_n(t, \theta) = \alpha(t, \theta)Y_n(t)$$

(see, e.g., Hjort [4], ABGK [1]). In this case

$$h_n(t, \theta) = h(t, \theta) = \frac{\partial}{\partial\theta}\log\alpha(t, \theta)$$

and

$$H^T(t, \theta) = \Big(1, \frac{\partial}{\partial\theta}\log\alpha(t, \theta)\Big).$$

However, unlike the previous example, we do not know the limiting function for $n^{-1}\lambda_n(t, \theta)$, i.e., for $n^{-1}Y_n(t)$ (we only assume that this limit exists). Now the kernel of the transformation has the following form:

$$R_n(t, \tau, \theta) = \int_0^{t\wedge\tau} H^T(u)B_n^{-1}(u)Y_n(u)\alpha(u)\, du H(\tau),$$

where

$$B_n(u) = \int_u^T Y_n(s)\, dI(s)$$

and

$$I(s) = \int_0^s \begin{pmatrix} 1 & h^T(\sigma) \\ h(\sigma) & h(s)h^T(\sigma) \end{pmatrix} \alpha(\sigma)\, d\sigma$$

is an "extended partial information matrix".

Let us demonstrate the construction of the goodness-of-fit process in the special case when the parameter θ is one-dimensional and the function $\alpha(t, \theta) = \psi(\theta)$ does not depend on t. In such situation the extension of function h to function H gives nothing new, because the components 1 and $\psi(\theta)$ of H are linearly dependent (for possible degeneration of the matrix I see Tsigroshvili [12]. In this case we have

$$B_n(u) = \psi(\theta)\Big[\frac{\partial}{\partial\theta}\log\psi(\theta)\Big]\int_u^T Y_n(s)\, ds$$

and the kernel $R_n(t,\tau,\theta)$ becomes free of θ:

$$R_n(t,\tau,\theta) = \int_0^{t\wedge\tau} \left(\int_u^T Y_n(s)ds\right)^{-1} Y_n(u)\,du = -\log\left(1 - \frac{\int_0^{t\wedge\tau} Y_n(s)ds}{\int_0^T Y_n(s)\,ds}\right).$$

It is easy to verify that Q_n annulates the compensator $\psi(\theta)\int_0^t Y_n(s)\,ds$ of $N_n(t)$ and finally the process W_n has the following form:

$$W_n(t) = n^{-1/2}\left[N_n(t) + \int_0^T \log\left(1 - \frac{\int_0^{t\wedge\tau} Y_n(s)\,ds}{\int_0^T Y_n(s)\,ds}\right) dN_n(\tau)\right].$$

Hence $\widetilde{W}_n(\cdot)$ does not depend on the value of the parameter θ.

Acknowledgements: The authors would like to thank Professors E. V. Khmaladze and R. D. Gill for their helpful comments and constant attention to their work.

References

[1] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer, New York, 1993.

[2] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1966.

[3] P. Bremaud, *Point Processes and Queues: Martingale Dynamics*, Springer, New York, 1981.

[4] N. L. Hjort, *Goodness-of-fit tests in models for life history data based on cumulative hazard rates*, Ann. Statist., 18 (1990), 1221–1258.

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

[6] E. V. Khmaladze, *Martingale approach to the theory of goodness-of-fit tests*, Theory Probab. Appl., 26 (1981), 240–257.

[7] E. V. Khmaladze, *Goodness of fit problem and scanning innovation martingales*, Ann. Statist., 21 (1993), 798–829.

[8] R. S. Liptser and A. N. Shiryaev, *A functional Central Limit Theorem for semimartingales*, Theory Probab. Appl., 25 (1980), 667–688.

[9] A. M. Nikabadze and E. V. Khmaladze, *On goodness of fit tests for parametric hypothesis in R^m*, Soviet Math. Dokl., 35 (1987), 627–629.

[10] M. S. Nikulin, *Chi-square test for continuous distributions with shift and scale parameters*, Theory Probab. Appl., 18 (1973), 559–568.

[11] M. C. J. van Pul, *Asymptotic properties of a class of statistical models in software reliability*, Scand. J. Statist., 19 (1992), 1–23.

[12] Z. P. Tsigroshvili, *Some notes on innovation martingales and goodness-of-fit tests*, Proc. A. Razmadze Math. Inst., 117, no.2 (1998) (to appear).

[*Received* December 1995, *revised* November 1997]