Transactions of A. Razmadze
Mathematical Institute
Vol. **174** (2020), issue 2, 155–173

PROJECTION APPROACH TO DISTRIBUTION-FREE TESTING FOR POINT PROCESSES. REGULAR MODELS

ESTATE V. KHMALADZE

Abstract. We create the notion of equivalence between different martingale models for point processes. This allows to map one model into another model in the same equivalence class. Therefore the distribution of test statistics for goodness of fit testing needs to be calculated in only once, for "standard" model, in each equivalence class. The equivalence classes are surprisingly broad, and thus the economy on computational work is considerable.Namely, any such class includes a non-time homogeneous Poisson model. Therefore it is sufficient to know the distribution of test statistics only for Poisson models.

The situation, therefore, becomes comparable to testing simple hypothesis about a continuous distribution function for a sample of i.i.d. random variables with continuous distribution F, when it is sufficient to consider F, uniform on $[0, 1]$. However, for point processes we consider here parametric cases, and the nature of equivalence is entirely different.

1. In Place of Introduction

This text was mainly written as a basic background material for the project which I was working on with Dr. S. Umut Can and Prof. R. Laeven from the University of Amsterdam. The aim of the project is to establish equivalence between testing parametric models for point processes with different forms of random intensities. Eventually, we intend to show that a huge majority of testing such models is equivalent to that of the non-time-homogeneous Poisson process which involves estimated parameters.

The text is not yet the final version, it is even not completely finished, but as it is, it may be useful for many readers. It is the first general and unified text with the material, which can be either found in various papers, or is new.

Umut Can greatly helped in preparation of the text and Roger Laeven made a number of useful remarks and I am grateful to both.

2. Basic Asymptotic Set-up

The method we want to develop for the testing problems for intensities of point processes can be first explained by drawing parallels between point processes and empirical processes, as the method for the latter has already been developed (see [6–8]).

Given a sample, i.e., a collection of independent and identically distributed (i.i.d.) positive random variables $X_1, \dots, X_n$, let us first consider the so-called *binomial process*

$$Z_n(t) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}} = \sum_{i=1}^{n} \mathbb{1}_{\{X_{(i:n)} \le t\}}, \quad t \ge 0. \tag{1}$$

Here, $X_{(i:n)}$ denotes the i^{th} order statistic of the sample $X_1, \dots, X_n$, with $X_{(1:n)} = \min\{X_1, \dots, X_n\}$ and $X_{(n:n)} = \max\{X_1, \dots, X_n\}$. Also, $\mathbb{1}_E$ denotes the indicator function of the event E, so for example,

$$\mathbb{1}_{\{X_i \le t\}} = \begin{cases} 1 & \text{if } X_i \le t \\ 0 & \text{otherwise} \end{cases}, \quad t \ge 0.$$

2020 *Mathematics Subject Classification.* 62C07, 62E20, 62F03, 60G55.

Key words and phrases. Martingale models for point processes; Models with estimated parameters; Asymptotic methods; Unitary operators.

For a given X_i, the indicator function $\mathbb{1}_{\{X_i\le t\}}$ is a step function of t, and since $X_1,\dots,X_n$ are i.i.d. random variables, $\mathbb{1}_{\{X_1\le t\}},\dots,\mathbb{1}_{\{X_n\le t\}}$ are i.i.d. stochastic processes in t. If we fix the value of $t>0$, then $\mathbb{1}_{\{X_1\le t\}},\dots,\mathbb{1}_{\{X_n\le t\}}$ become independent Bernoulli random variables with

$$P[\mathbb{1}_{\{X_i\le t\}}=1]=P[X_i\le t]=F(t),$$

where F denotes the common distribution function of the X_i's. It now follows from the first equality in (1) that $Z_n(t)\sim \mathrm{Binom}(n,F(t))$ and, in particular, $\mathbb{E}[Z_n(t)]=nF(t)$. It also follows from the Central Limit Theorem that for any $t>0$,

$$v_n(t):=\frac{1}{\sqrt{n}}[Z_n(t)-nF(t)] \tag{2}$$

is asymptotically Gaussian as $n\to\infty$. In fact, we know from the Functional Limit Theorem that not just $v_n(t)$ for any given $t>0$ is asymptotically Gaussian, but the stochastic process $\{v_n(t):t\ge 0\}$ is asymptotically Gaussian as well, in the sense that it converges weakly to a Gaussian process v. The process v_n is called the empirical process associated with the sample $X_1,\dots,X_n$, and the limiting Gaussian process v is called the F-Brownian bridge. Occasionally, it will be convenient to use the notation $F_n(t)=Z_n(t)/n$ for an empirical distribution function and write empirical process v_n in the equivalent form

$$v_n(t):=\sqrt{n}[F_n(t)-F(t)].$$

For these and many more nice facts about empirical processes we refer the readers to the monograph [10]. Some of these facts may not be, however, very visible from the second definition in (1). Indeed, the random variables $X_{(1:n)},\dots,X_{(n:n)}$ are neither independent nor identically distributed. Although $\mathbb{1}_{\{X_{(1:n)}\le t\}},\dots,\mathbb{1}_{\{X_{(n:n)}\le t\}}$ are still the Bernoulli random variables for any fixed $t\ge 0$, they are now very much dependent, and the distribution functions

$$F_{(i:n)}(t):=P[X_{(i:n)}\le t]$$

are very different for different i. The properly centered form of $Z_n(t)$ taken from the second definition in (1) is, therefore,

$$Z_n(t)-\sum_{i=1}^{n}F_{(i:n)}(t),\quad t\ge 0, \tag{3}$$

and it is almost an accident that

$$\sum_{i=1}^{n}F_{(i:n)}(t)=nF(t).$$

The second definition in (1) has, however, the advantage that it represents $Z_n(t)$ as a point process with order statistics corresponding to arrival times: $X_{(i:n)}$ can be interpreted as the arrival time of the i^{th} event.

As the martingale theory of point processes is well-developed and widely known, almost nobody would center point processes by their unconditional expected values as in (3). What is done instead is the *conditional* centering of increments of $Z_n(t)$ given the past history of this process:

$$\mathrm{d}Z_n(t)-\mathbb{E}[\mathrm{d}Z_n(t)\,|\,Z_n(s),0\le s\le t]=:\mathrm{d}M_n(t). \tag{4}$$

The resulting process $\{M_n(t):t\ge 0\}$ is a martingale and the equality (4) itself is called the Doob-Meyer decomposition of $Z_n(t)$, which we now view as a submartingale.

Let us now define

$$\lambda_n(t)=\mathbb{E}[\mathrm{d}Z_n(t)\,|\,Z_n(s),0\le s\le t]/\mathrm{d}t,\quad t\ge 0,$$

which is called the *intensity* of the point process $Z_n(t)$, and

$$w_n(t)=\frac{1}{\sqrt{n}}M_n(t)=\frac{1}{\sqrt{n}}\left[Z_n(t)-\int_0^t\lambda_n(s)\,\mathrm{d}s\right],$$

which is also a martingale in t. Thus from $Z_n(t)$ we have produced, using different methods of centering, two different processes, namely, the empirical process $v_n(t)$ and the process $w_n(t)$, which we will refer to as an *innovation martingale* of the process $Z_n(t)$. Yet, we will see below that there is a very important similarity between the asymptotic behavior of v_n and w_n in the practically important case when the underlying distribution function F depends on some finite-dimensional parameter θ, and when the random intensity λ_n also depends on such a parameter.

In the context of goodness of fit testing, when the null hypothesis does not completely specify the distribution function F, but only states that it belongs to a parametric family $\{F_\theta : \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^m$, we call this hypothesis a parametric hypothesis. The same term is used if we hypothesize that the intensity of Z_n belongs to a parametric family of intensities $\{\lambda_{n,\theta} : \theta \in \Theta\}$. In the case of a parametric hypothesis, we will need to estimate the parameter θ and then to make a judgment on whether the hypothesis is true or not, by observing the behavior of the processes

$$\frac{1}{\sqrt{n}}[Z_n(t) - nF_{\widehat{\theta}}(t)] = v_{n,\widehat{\theta}}(t) = \widehat{v}_n(t)$$

and

$$\frac{1}{\sqrt{n}}\left[Z_n(t) - \int_0^t \lambda_{n,\widehat{\theta}}(s)\,\mathrm{d}s\right] = w_{n,\widehat{\theta}}(t) = \widehat{w}_n(t),$$

respectively. The 'similarity' that was alluded to above consists in the fact that $\widehat{v}_n$ is asymptotically a projection of v_n, and $\widehat{w}_n$ is asymptotically a projection of w_n; substituting the estimate $\widehat{\theta}$ in place of the true parameter θ is asymptotically equivalent to projecting the initial process. Thus if we have a method that exploits this geometric fact in the case a parametric hypothesis about distribution functions, it should be possible to develop a similar method in the situation with point processes.

Let us now review why we have a projection in the case of a parametric hypothesis about F. Suppose that $\{F_\theta : \theta \in \Theta\}$ is a *regular* parametric family of distributions in the following sense:

(a_1) the space Θ of feasible parameter values is an open subset of the Euclidean space $\mathbb{R}^m$;

(a_2) the vector of the derivatives

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = [\dot{f}/f]_\theta(x)$$

is square-integrable, i.e., the Fisher information matrix

$$R_\theta = \int [\dot{f}/f]_\theta(x)\,[\dot{f}/f]_\theta^{\mathsf{T}}(x)\,f_\theta(x)\,\mathrm{d}x$$

is finite and non-degenerate for every $\theta \in \Theta$,

(a_3) for any $\theta \in \Theta$;

$$\int [\dot{f}/f]_\theta(x)\,f_\theta(x)\,\mathrm{d}x = 0.$$

The openness of Θ is useful because then every θ has a neighborhood in Θ and we can differentiate at θ without worrying about boundary effects. Conditions (a_2) and (a_3) are ubiquitous in all asymptotic statistics with regular parametric families.

To describe the difference between $\widehat{v}_n$ and v_n we first need an asymptotic representation of the maximum likelihood estimator (MLE) $\widehat{\theta}$, or rather, of $\sqrt{n}(\widehat{\theta} - \theta)$. The MLE is the (correctly chosen) root of the maximum likelihood equation

$$\sum_{i=1}^n [\dot{f}/f]_{\widehat{\theta}}(X_i) = 0. \tag{5}$$

Using the regularity condition (a_3), we can rewrite (5) as

$$\int [\dot{f}/f]_{\widehat{\theta}}(x)\,[\mathrm{d}Z_n(x) - n\mathrm{d}F_{\widehat{\theta}}(x)] = 0,$$

that is,

$$\int [\dot{f}/f]_{\widehat{\theta}}(x)\,\mathrm{d}v_{n,\widehat{\theta}}(x) = 0. \tag{6}$$

Replacing the left-hand side of (6) by the Taylor expansion around θ, we obtain

$$0 = \int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) + \int \frac{\partial}{\partial\theta}[\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x)(\widehat\theta-\theta)$$
$$-\sqrt{n}\int [\dot f/f]_\theta(x)\,\dot f_\theta(x)^{\mathsf T}\mathrm{d}x(\widehat\theta-\theta) + o_P(1). \tag{7}$$

Here, the assumption that the residual term is indeed $o_P(1)$ is, actually, another regularity assumption, (a_4), on the family $\{F_\theta : \theta\in\Theta\}$. Note that we can write the second term in the right-hand side of (7) as

$$\frac{1}{\sqrt n}\int \frac{\partial}{\partial\theta}[\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x)\sqrt{n}(\widehat\theta-\theta)$$
$$= \int \frac{\partial}{\partial\theta}[\dot f/f]_\theta(x)\,\mathrm{d}\,[F_n(x)-F_\theta(x)]\,\sqrt{n}(\widehat\theta-\theta),$$

which is asymptotically negligible as long as the matrix

$$\frac{\partial}{\partial\theta}\left[\dot f/f\right]_\theta(x) = \frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)$$

is integrable with respect to F_θ – this follows from the Law of Large Numbers. Using the regularity assumption (a_2) for the third term on the right-hand side of (7), we obtain

$$0 = \int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) - R_\theta\sqrt{n}(\widehat\theta-\theta) + o_P(1),$$

or equivalently,

$$\sqrt{n}(\widehat\theta-\theta) = R_\theta^{-1}\int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) + o_P(1), \tag{8}$$

which is the asymptotic MLE representation we wanted.

Now, let us apply the Taylor expansion again and write

$$v_{n,\widehat\theta}(t) = v_{n,\theta}(t) - \left[\frac{\partial}{\partial\theta}F_\theta(t)\right]^{\mathsf T}\sqrt{n}(\widehat\theta-\theta) + o_P(1),$$

or, by virtue of (8),

$$v_{n,\widehat\theta}(t) = v_{n,\theta}(t) - \int_0^t [\dot f/f]_\theta^{\mathsf T}(s) f_\theta(s)\,\mathrm{d}s\; R_\theta^{-1}\int [\dot f/f]_\theta(x)\,\mathrm{d}v_{n,\theta}(x) + o_P(1). \tag{9}$$

The main part on the right-hand side of (9) is a linear transformation of $v_{n,\theta}$; moreover, as the proposition below shows, it is a projection.

Proposition 2.1. *The linear operator* Π *defined by*

$$\Pi\gamma(t) = \gamma(t) - \int_0^t [\dot f/f]_\theta^{\mathsf T}(s)\,\mathrm{d}F_\theta(s)\,R_\theta^{-1}\int [\dot f/f]_\theta(x)\,\mathrm{d}\gamma(x)$$

is an orthogonal projector, i.e., it satisfies the conditions

(i) $\Pi\Pi\gamma(t) = \Pi\gamma(t)$,

(ii) $\Pi\gamma(t)\equiv 0 \Leftrightarrow \dfrac{\mathrm{d}\gamma}{\mathrm{d}F}(t) = c^{\mathsf T}[\dot f/f]_\theta(t)$ *for some* $c\in\mathbb{R}^m$,

(iii) $\int [\dot f/f]_\theta(s)\,\mathrm{d}\Pi\gamma(s) = 0$.

This fact has several useful consequences which we will discuss later. Right now we would like to establish the analogous result for point processes.

Given a point process $\{N_n(t) : t\geq 0\}$, let

$$\lambda_{n,\theta}(t)\mathrm{d}t = \mathbb{E}[\mathrm{d}N_n(t)\,|\,N_n(s), 0\leq s\leq t]$$

denote the intensity as above, and let $t_0 = 0 < t_1 < \cdots < t_k = T$ be a partition of the interval $[0, T)$, where we are considering our point process. If the partition is sufficiently fine, the likelihood of the vector $(N_n(t_1), \ldots, N_n(t_k))^\mathsf{T}$ is given as a product of Bernoulli's likelihoods:

$$\begin{aligned}
&\prod_{j=1}^{k} \left[\lambda_{n,\theta}(t_j)\Delta t_j\right]^{\Delta N_n(t_j)} \left[1 - \lambda_{n,\theta}(t_j)\Delta t_j\right]^{1-\Delta N_n(t_j)} \\
&= \exp\left\{ \sum_{j=1}^{k} \Delta N_n(t_j) \ln\left[\lambda_{n,\theta}(t_j)\Delta t_j\right] + \sum_{j=1}^{k} \left(1 - \Delta N_n(t_j)\right) \ln\left[1 - \lambda_{n,\theta}(t_j)\Delta t_j\right] \right\}. \qquad (10)
\end{aligned}$$

The likelihood of the same vector under the assumption that N_n is a Poisson process with constant intensity λ has a similar form, with $\lambda_{n,\theta}$ replaced by λ. This Poisson likelihood is only a reference likelihood and we could have used many other measures in order to create likelihood ratios. If we take the limit of the likelihood in (10) as $k \to \infty$ and $\max_j\{\Delta t_j : 1 \le j \le k\} \to 0$, we will obtain zero, but the likelihood ratio below will have a non-trivial limit:

$$\begin{aligned}
&\frac{\prod_{j=1}^{k} \left[\lambda_{n,\theta}(t_j)\Delta t_j\right]^{\Delta N_n(t_j)} \left[1 - \lambda_{n,\theta}(t_j)\Delta t_j\right]^{1-\Delta N_n(t_j)}}{\prod_{j=1}^{k} \left[\lambda\Delta t_j\right]^{\Delta N_n(t_j)} \left[1 - \lambda\Delta t_j\right]^{1-\Delta N_n(t_j)}} \\
&\quad = \exp\left\{ \sum_{j=1}^{k} \Delta N_n(t_j) \ln\frac{\lambda_{n,\theta}(t_j)}{\lambda} + \sum_{j=1}^{k} \left(1 - \Delta N_n(t_j)\right) \ln\frac{1 - \lambda_{n,\theta}(t_j)\Delta t_j}{1 - \lambda\Delta t_j} \right\} \\
&\quad \to \exp\left\{ \int_0^T \ln\frac{\lambda_{n,\theta}(t)}{\lambda}\, \mathrm{d}N_n(t) - \int_0^T \left[\lambda_{n,\theta}(t) - \lambda\right] \mathrm{d}t \right\}.
\end{aligned}$$

Differentiating this log-likelihood ratio with respect to θ and setting the result equal to zero, we obtain the maximum likelihood equation

$$\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\, \mathrm{d}N_n(t) - \int_0^T \dot\lambda_{n,\theta}(t)\, \mathrm{d}t = 0,$$

which can be rewritten as

$$\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\, [\mathrm{d}N_n(t) - \lambda_{n,\theta}(t)\mathrm{d}t] = 0,$$

or

$$\int_0^T [\dot\lambda/\lambda]_{n,\theta}(t)\, \mathrm{d}w_{n,\theta}(t) = 0.$$

Now we need regularity assumptions on $\lambda_{n,\theta}$ as a function of θ, namely:

- (b_1) differentiation with respect to θ, integration with respect to $\mathrm{d}N_n(t)$, and $\mathrm{d}t$ can be interchanged,
- (b_2) the ratio $[\dot\lambda/\lambda]_{n,\theta}(t)$ is well-defined on $\{t : \lambda_{n,\theta}(t) > 0\}$, and can be defined as a constant on $\{t : \lambda_{n,\theta}(t) = 0\}$,
- (b_3) we have

$$\frac{1}{\sqrt{n}} \int_0^T \frac{\partial}{\partial\theta} [\dot\lambda/\lambda]_{n,\theta}(t)\, \mathrm{d}w_{n,\theta}(t) = \int_0^T \frac{\partial}{\partial\theta} [\dot\lambda/\lambda]_{n,\theta}(t) \left[\frac{\mathrm{d}N_n(t)}{n} - \frac{\lambda_{n,\theta}(t)}{n}\mathrm{d}t\right] = o_P(1),$$

which is a form of Law of Large Numbers for the process N_n and the vector function $[\dot\lambda/\lambda]_{n,\theta}(t)$.

To obtain a suitable asymptotic expansion for the MLE $\widehat{\theta}$, we use the Taylor expansion once again, and rewrite the maximum likelihood equation

$$\int_0^T [\dot{\lambda}/\lambda]_{n,\widehat{\theta}}(t)\,\mathrm{d}w_{n,\widehat{\theta}}(t) = 0$$

as follows:

$$0 = \int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t) + \frac{1}{\sqrt{n}}\int_0^T \frac{\partial}{\partial\theta}[\dot{\lambda}/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t)\sqrt{n}(\widehat{\theta}-\theta)$$
$$- \frac{1}{n}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)[\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(t)\lambda_{n,\theta}(t)\,\mathrm{d}t\sqrt{n}(\widehat{\theta}-\theta) + o_P(1).$$

Here, the last $o_P(1)$ is our regularity assumption (b_4) and we will also use

(b_5) the random matrix

$$R_{n,\theta} = \frac{1}{n}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\,[\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(t)\,\lambda_{n,\theta}(t)\,\mathrm{d}t$$

is well-defined and non-degenerate for all $\theta \in \Theta$ and all n sufficiently large. Moreover, there is a non-degenerate matrix R_θ such that $R_{n,\theta} \to R_\theta$ as $n \to \infty$.

This leads to the desired asymptotic representation

$$\sqrt{n}(\widehat{\theta}-\theta) = R^{-1}_{n,\theta}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t) + o_P(1), \tag{11}$$

which is an analog of (8) for point processes.

Now we turn to the difference between $w_{n,\theta}$ and $w_{n,\widehat{\theta}}$. Using Taylor's expansion again, we obtain

$$w_{n,\widehat{\theta}}(t) = w_{n,\theta}(t) - \frac{1}{n}\int_0^t [\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(s)\lambda_{n,\theta}(s)\,\mathrm{d}s\sqrt{n}(\widehat{\theta}-\theta) + o_P(1),$$

or, by virtue of (11),

$$w_{n,\widehat{\theta}}(t) = w_{n,\theta}(t) - \frac{1}{n}\int_0^t [\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(s)\lambda_{n,\theta}(s)\,\mathrm{d}s\,R^{-1}_{n,\theta}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\,\mathrm{d}w_{n,\theta}(t) + o_P(1),$$

an expression analogous to (9). The main part on the right-hand side is a linear transformation of $w_{n,\theta}$. Moreover, defining

$$\Lambda_{n,\theta}(t) = \int_0^t \lambda_{n,\theta}(s)\,\mathrm{d}s,$$

we have the following analog of Proposition 2.1.

Proposition 2.2. *The linear operator* Π_n *defined by*

$$\Pi_n\gamma(t) = \gamma(t) - \frac{1}{n}\int_0^t [\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(s)\,\mathrm{d}\Lambda_{n,\theta}(s)\,R^{-1}_{n,\theta}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(t)\,\mathrm{d}\gamma(t)$$

is an orthogonal projector, i.e., it satisfies the conditions

(i) $\Pi_n\Pi_n\gamma(t) = \Pi_n\gamma(t)$,

(ii) $\Pi_n\gamma(t) \equiv 0 \Leftrightarrow \dfrac{\mathrm{d}\gamma}{\mathrm{d}\Lambda_{n,\theta}}(t) = c^{\mathsf{T}}[\dot{\lambda}/\lambda]_{n,\theta}(t)$ *for some* $c \in \mathbb{R}^m$,

(iii) $\displaystyle\int_0^T [\dot\lambda/\lambda]_{n,\theta}(s)\,\mathrm{d}\Pi_n\gamma(s) = 0.$

Therefore, substitution of ML estimation in place of the 'true' value of the parameter again is asymptotically equivalent to taking a projection of the martingale $w_{n,\theta}$. Heuristically speaking, this makes the process $w_{n,\hat\theta}$ stochastically 'smaller', less volatile, less 'noisy', and makes the tests based on $w_{n,\hat\theta}$ better, more powerful, as it is the case for the process $v_{n,\hat\theta}$ (see, e.g., [4]).

3. Function Parametric Versions and Unitary Operators

We realised that the empirical process $v_{n,\widehat\theta}$ with estimated parameter $\widehat\theta$ is essentially a projection of the corresponding empirical process $v_{n,\theta}$ (see (9)). However, for different parametric families we have different score functions $\dot f/f$, and therefore different projections. Even within the same parametric family, different values of the parameter θ again lead to different $\dot f/f$, and $v_{n,\widehat\theta}$ will have different limit behavior. Consequently, the limiting distribution of any given test statistic $T(v_{n,\widehat\theta})$ will be different in any new testing problem.

In the goodness of fit problems, the test statistics T, as functionals of $v_{n,\widehat\theta}$, are non-linear and their limiting distributions are difficult to calculate, so that numerical methods have to be used. The theory becomes fragmented. Our eventual goal is to unify the theory again. We will see that what looks like many similar but different problems actually is one single problem, which requires the calculation of limiting distributions of test statistics $T(v_{n,\widehat\theta})$, for many "similar" $v_{n,\widehat\theta}$, only once. The same is true for testing parametric hypotheses about (random) intensities of point processes.

The main idea behind the methods we are going to employ consists in building a unitary operator, or rotation, of one testing problem into another, thus creating surprisingly broad families of equivalent testing problems. However, it may look awkward to "rotate" empirical processes. We introduce now a form of empirical processes which will create a natural setting to apply unitary operators.

Let v_F denote an F-Brownian bridge, i.e., a Gaussian process with mean zero and covariance function

$$\mathbb{E}[v_F(t)v_F(t')] = \min\{F(t), F(t')\} - F(t)F(t').$$

This process is the limit in distribution of the empirical process v_n in (2). It is convenient to recall that if w_F is an F-Brownian motion, i.e., a Gaussian process with mean zero and covariance

$$\mathbb{E}[w_F(t)w_F(t')] = \min\{F(t), F(t')\},$$

then one well-known connection between w_F and v_F is

$$v_F(t) \overset{d}{=} w_F(t) - F(t)w_F(\infty), \tag{12}$$

and if we agree to choose F supported on the unit interval $[0, 1]$ so that $F(0) = 0$ and $F(1) = 1$, then we can write $w_F(1)$ instead of $w_F(\infty)$.

If $\{F_\theta : \theta \in \Theta\}$ is a regular parametric family, then from (9) it is possible to derive that the Gaussian process

$$\widehat v_F(t) = v_F(t) - \int_0^t [\dot f/f]_\theta^{\mathsf T}(s)\,\mathrm{d}F_\theta(s) R_\theta^{-1} \int [\dot f/f]_\theta(x)\,\mathrm{d}v_F(x)$$

is the limit in distribution of the parametric empirical process $v_{n,\widehat\theta}$ with θ denoting the true parameter value. Now let us rewrite $\widehat v_F$ in what is called function parametric form.

Suppose, as before, that θ is an m-dimensional parameter. Then $[\dot f/f]_\theta(\cdot)$ is an m-dimensional vector function with linearly independent components. Let us introduce the notation

$$q_\theta(\cdot) = R_\theta^{-1/2}[\dot f/f]_\theta(\cdot)$$

for the ortho-normalised form of the score function $[\dot f/f]_\theta$. Indeed, q_θ is a vector function with orthogonal and normalized components in the space $L_2(F)$, since

$$\int q_\theta\, q_\theta^\top \,\mathrm{d}F_\theta = R_\theta^{-1/2} \int [\dot f/f]_\theta [\dot f/f]_\theta^\top \,\mathrm{d}F_\theta\, R_\theta^{-1/2} = I.$$

Below, we will drop the subscript θ in F_θ when F_θ is used as a subscript.

Now, given a function $\varphi \in L_2(F_\theta)$, let us introduce what is called *function parametric* version of our processes. Consider the integral

$$\widehat{v}_F(\varphi) := \int \varphi(x)\,\mathrm{d}\widehat{v}_F(x) = \int \varphi(x)\,\mathrm{d}v_F(x) - \int \varphi(x) q_\theta^\top(x)\,\mathrm{d}F_\theta(x) \int q_\theta(y)\,\mathrm{d}v_F(y). \tag{13}$$

This is a Wiener stochastic integral, well-defined on $L_2(F_\theta)$. It is clear that

$$v_F(\varphi) = \int \varphi(x)\,\mathrm{d}v_F(x)$$

is linear in φ, that is, if $\varphi_1, \varphi_2 \in L_2(F_\theta)$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, then

$$v_F(\alpha_1\varphi_1 + \alpha_2\varphi_2) = \alpha_1 v_F(\varphi_1) + \alpha_2 v_F(\varphi_2).$$

This implies that $\widehat{v}_F(\varphi)$ is also linear in φ, and (13) can be rewritten as

$$\widehat{v}_F(\varphi) = v_F(\varphi) - \langle \varphi, q_\theta\rangle_F^\top\, v_F(q_\theta) = v_F\big(\varphi - \langle \varphi, q_\theta\rangle_F^\top\, q_\theta\big), \tag{14}$$

where $\langle \varphi, q_\theta\rangle_F$ denotes the vector of inner products in $L_2(F)$ of φ and the components of q_θ:

$$\langle \varphi, q_\theta\rangle_F := \int \varphi(x) q_\theta(x)\,\mathrm{d}F_\theta(x).$$

Thus we have the following reformulation of Proposition 2.1. To formulate its (ii) part, we extend the m-dimensional vector of score functions q_θ to the $(m+1)$-dimensional vector having q_0 as the first coordinate:

$$q = \begin{pmatrix} q_0 \\ q_\theta \end{pmatrix} = \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_m \end{pmatrix}.$$

Here q_0 is the function, which is constant and equals 1 for all x. Note that the extended q will still be a vector with orthonormal coordinates, because our regularity assumption (a_3) implies

$$\int [\dot f/f]_\theta(x)\,\mathrm{d}F_\theta(x) = \langle [\dot f/f]_\theta, q_0\rangle_F = 0, \quad \text{or} \quad \langle q_\theta, q_0\rangle_F = 0,$$

Proposition 3.1 ([4]). *For the limiting processes of $v_{n,\theta}$ and $v_{n,\hat\theta}$ we have*

(i)

$$\widehat{v}_F(\varphi) = v_F\big(\varphi - \langle \varphi, q_\theta\rangle_F^\top\, q_\theta\big),$$

which represents $\widehat{v}_F$ as a projection of function parametric Brownian bridge v_F, and

(ii)

$$\widehat{v}_F(\varphi) = w_F\big(\varphi - \langle \varphi, q\rangle_F^\top\, q\big), \tag{15}$$

which represents $\widehat{v}_F$ as a projection of function parametric Brownian motion w_F.

Proof. To see that (i) is true, note that the form of the argument of v_F in (i) follows from (14), and since q_θ is orthonormal, this is the orthogonal projection of φ, parallel to q_θ,

$$\pi\varphi = \varphi - \langle \varphi, q_\theta\rangle_F^\top\, q_\theta.$$

To see that (ii) is true, we use the projection structure behind the function parametric form of Brownian bridge $v_F(\varphi)$, $\varphi \in L_2(F)$, as well. According to (12),

$$\begin{aligned} v_F(\varphi) &= \int \varphi(x)\,\mathrm{d}v_F(x) = \int \varphi(x)\,\mathrm{d}w_F(x) - \int \varphi(x)\,\mathrm{d}F_\theta(x) w_F(\infty) \\ &= w_F(\varphi) - \langle \varphi, q_0\rangle_F\, w_F(q_0), \end{aligned} \tag{16}$$

where we recall that q_0 denotes the function, identically equal to 1, $q_0(\cdot) \equiv 1$. Again, since $w_F(\varphi)$ is linear in φ, we can rewrite the last expression as

$$v_F(\varphi) = w_F\big(\varphi - \langle\varphi, q_0\rangle_F\, q_0\big), \tag{17}$$

and the argument of w_F here is the orthogonal projection of φ, parallel to q_0. Now we substitute (16) into (14). This will represent $\widehat{v}_F$ as a projection of w_F:

$$\widehat{v}_F(\varphi) = w_F(\varphi) - \langle\varphi, q_0\rangle_F\, w_F(q_0) - \langle\varphi, q_\theta\rangle_F^{\mathsf{T}}\, w_F(q_\theta). \tag{18}$$

We replaced the term $v_F(q_\theta)$ in (14) by the term $w_F(q_\theta)$, and we can indeed do this: as it follows from (17),

$$v_F(q_\theta) = w_F(q_\theta - \langle q_\theta, q_0\rangle_F\, q_0),$$

while $\langle q_\theta, q_0\rangle_F = 0$, and therefore $v_F(q_\theta) = w_F(q_\theta)$. □

Now let us consider two different regular parametric families $\{F_\theta : \theta \in \Theta\}$ and $\{G_\theta : \theta \in \Theta\}$, with two different score functions q and r (extended and orthonormal, as above). We assume, however, that the vector functions q and r are of equal dimensions. In this notation we use the same letters θ and Θ, but we do not mean to say that these are in any sense the "same" parameters, say shift and scale parameters in both cases. They may be parameters of entirely different nature in these two different families. They only should be of the same dimension and they should lead to linearly independent, and therefore eventually orthonormal, score functions.

Consider two limiting Gaussian processes

$$\widehat{v}_F(\varphi) = w_F\Big(\varphi - \sum_{i=0}^{m}\langle\varphi, q_i\rangle_F\, q_i\Big), \quad \varphi \in L_2(F_\theta), \quad \langle q_i, q_j\rangle_F = \delta_{ij},$$

$$and \widehat{v}_G(\psi) = w_G\Big(\psi - \sum_{i=0}^{m}\langle\psi, r_i\rangle_G\, r_i\Big), \quad \psi \in L_2(G_\theta), \quad \langle r_i, r_j\rangle_G = \delta_{ij}.$$

What we will show now is that, under the additional assumption of equivalence (mutual absolute continuity) between the distributions F_θ and G_θ, we can map $\widehat{v}_F$ to $\widehat{v}_G$ in a one-to-one way, and the mapping has a practically convenient form. More specifically, we will construct a unitary operator $K = K_{q,r}$ mapping $L_2(G_\theta)$ onto $L_2(F_\theta)$, so that

$$\widehat{v}_F(K\psi) \stackrel{d}{=} \widehat{v}_G(\psi), \quad \psi \in L_2(G_\theta).$$

Because this K is a unitary operator, we will also have

$$\widehat{v}_G(K^{-1}\varphi) \stackrel{d}{=} \widehat{v}_F(\varphi), \quad \varphi \in L_2(F_\theta).$$

Allowing ourselves some freedom of speech, we will say that $\widehat{v}_F$ is "rotated" into $\widehat{v}_G$.

For the sake of better transparency, let us construct K in a sequence of three problems. In the first, or "zero problem," let us map w_F into w_G isometrically. Here, the dependence on the parameter θ will play no role, and it can be skipped from the notations. Consider the square root of the density of G with respect to F:

$$\ell(x) = \left(\frac{\mathrm{d}G(x)}{\mathrm{d}F(x)}\right)^{1/2}.$$

Since F and G are equivalent measures, we have

$$\ell \in L_2(F), \text{ with } \|\ell\|_F^2 = \int \ell^2(x)\,\mathrm{d}F(x) = 1,$$

$$1/\ell \in L_2(G), \text{ with } \|1/\ell\|_G^2 = \int \frac{1}{\ell^2(x)}\,\mathrm{d}G(x) = 1.$$

Let $\ell\psi(\cdot) = \ell(\cdot)\psi(\cdot)$ denote the operator of multiplication by the function ℓ, acting on functions $\psi \in L_2(G)$.

Lemma 3.2. *The operator ℓ is an isometry from $L_2(G)$ to $L_2(F)$, and we have $w_F(\ell\psi) = w_G(\psi)$.*

Proof. Indeed,

$$\int \big[l(x)\psi(x)\big]^2\,\mathrm{d}F(x) = \int \psi^2(x)\,\mathrm{d}G(x),$$

and therefore,

$$\mathbb{E}w_F^2(\ell\psi) = \|\ell\psi\|_F^2 = \|\psi\|_G^2 = \mathbb{E}w_G^2(\psi). \qquad \square$$

The next problem is to rotate the Brownian bridge v_F into the Brownian bridge v_G. Now we have one-dimensional functions $q_0(\cdot) = 1$ and $r_0(\cdot) = 1$. Note that the latter function is identically equal to 1 in $L_2(G)$, but its image under the operator ℓ will not be identically equal to 1 in $L_2(F)$. We know that

$$v_G(\psi) = \begin{cases} w_G(\psi) & \text{if } \psi \perp r_0 \\ 0 & \text{if } \psi = r_0 \end{cases} \quad \text{and} \quad v_F(\varphi) = \begin{cases} w_F(\varphi) & \text{if } \varphi \perp q_0 \\ 0 & \text{if } \varphi = q_0 \end{cases}.$$

Therefore, in order to rotate v_F into v_G we need a unitary operator from $L_2(G)$ to $L_2(F)$ which will map the linear subspace $\mathcal{L}_G(r_0) = \{cr_0(\cdot) : c \in \mathbb{R}\}$ into the linear subspace $\mathcal{L}_F(q_0)$, and which, therefore, will map the orthogonal complement of $\mathcal{L}_G(r_0)$ in $L_2(G)$ (denoted by $\mathcal{L}_{G,\perp}(r_0)$) into the orthogonal complement of $\mathcal{L}_F(q_0)$ in $L_2(F)$ (denoted by $\mathcal{L}_{F,\perp}(q_0)$). In order to do this, consider first the operator $K_{a,b}$ mapping $L_2(F)$ to $L_2(F)$ via

$$K_{a,b}(\cdot) = I - 2\frac{\langle a-b,\cdot\rangle_F}{\|a-b\|_F^2}(a-b), \tag{19}$$

where $a, b \in L_2(F)$ are two fixed functions of unit norm, and I is the identity operator. It is easy to check that this operator has the following properties:

1. $K_{a,b}$ is unitary, i.e., $\|K_{a,b}\varphi\|_F = \|\varphi\|_F$,
2. $K_{a,b} = K_{a,b}^{-1}$, i.e., $K_{a,b}K_{a,b} = I$,
3. $K_{a,b}$ is self-adjoint, i.e., $\langle K_{a,b}\varphi, \gamma\rangle_F = \langle \varphi, K_{a,b}\gamma\rangle_F$,
4. $K_{a,b}\,a = b$ and $K_{a,b}\,b = a$.

Now let us choose $a = q_0$ and $b = \ell r_0$, and consider the process $v_F(K_{q_0,\ell r_0}\ell\psi)$ for $\psi \in L_2(G)$. We claim that this process has the same distribution as $v_G(\psi)$. Together with the statement on Brownian motions, above, we obtain

Proposition 3.3. *If distributions F and G are equivalent, then*

$$w_F(\ell\psi) = w_G(\psi) \quad \textit{and} \quad v_F(K_{q_0,\ell r_0}\ell\psi) \overset{d}{=} v_G(\psi).$$

Proof. Indeed, if $\psi = r_0$, then $K_{q_0,\ell r_0}\ell\psi = q_0$, and so

$$v_F(K_{q_0,\ell r_0}\ell r_0) = v_F(q_0) = 0 = v_G(r_0).$$

On the other hand, if $\psi \perp r_0$, then $\ell\psi \perp \ell r_0$, and therefore $K_{q_0,\ell r_0}\ell\psi \perp q_0$. From the equality $v_F(\varphi) = w_F(\varphi)$ when $\varphi \perp q_0$, it follows that

$$v_F(K_{q_0,\ell r_0}\ell\psi) = w_F(K_{q_0,\ell r_0}\ell\psi),$$

and the variance of the right-hand side for any such ψ is

$$\mathbb{E}w_F^2(K_{q_0,\ell r_0}\ell\psi) = \langle K_{q_0,\ell r_0}\ell\psi, K_{q_0,\ell r_0}\ell\psi\rangle_F = \langle \ell\psi, \ell\psi\rangle_F = \langle \psi,\psi\rangle_G = \mathbb{E}w_G^2(\psi).$$

Therefore, indeed, v_F was "rotated" into v_G. $\square$

We are now ready to tackle the third problem in our sequence, the rotation of $\widehat{v}_F$ into $\widehat{v}_G$. Let us first consider the case of regular parametric families with one-dimensional parameter, leading to two-dimensional extended score functions $(q_0, q_1)^\mathsf{T}$ for one family and $(r_0, r_1)^\mathsf{T}$ for the other. Now we have

$$\widehat{v}_F(\varphi) = w_F\big(\varphi - \langle\varphi, q_0\rangle_F q_0 - \langle\varphi, q_1\rangle_F q_1\big), \quad \varphi \in L_2(F_\theta),$$
$$\widehat{v}_G(\psi) = w_G\big(\psi - \langle\psi, r_0\rangle_G r_0 - \langle\psi, r_1\rangle_G r_1\big), \quad \psi \in L_2(G_\theta).$$

Consider the operator $K_{q_0,\ell r_0}$ used for the previous problem above and apply it to ℓr_1, thus creating

$$\widetilde{\ell r_1} := K_{q_0,\ell r_0}\ell r_1.$$

The operator $K_{q_0,\ell r_0}$ correctly rotates the function ℓr_0 into q_0, but it does not necessarily rotate ℓr_1 into q_1, but only into $\widetilde{\ell r_1}$. Since it is a unitary operator, it preserves angles, and therefore $\widetilde{\ell r_1} \perp q_0$. Now we can rotate $\widetilde{\ell r_1}$ further into q_1 using the operator $K_{q_1,\widetilde{\ell r_1}}$. Note that this operator leaves all functions orthogonal to q_1 and $\widetilde{\ell r_1}$ unchanged, so it will leave q_0 unchanged. Now consider the operator $K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0}$. We have

Proposition 3.4. *If distributions F_θ and G_θ are equivalent and if $q_0, q_1 \in L_2(F_\theta)$ are orthonormal as well as $r_0, r_1 \in L_2(G_\theta)$, then*

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi) \stackrel{d}{=} \widehat{v}_G(\psi), \quad \psi \in L_2(G_\theta).$$

Proof. Indeed, if $\psi = r_0$, then

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell r_0) = \widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} q_0) = \widehat{v}_F(q_0) = 0 = \widehat{v}_G(r_0),$$

and similarly, if $\psi = r_1$,

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell r_1) = \widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} \widetilde{\ell r_1}) = \widehat{v}_F(q_1) = 0 = \widehat{v}_G(r_1).$$

Moreover, $K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0}$ is a product of unitary operators, hence it is itself a unitary operator. As we have just seen, it maps ℓr_0 into q_0 and ℓr_1 into q_1. Therefore, it will map $\ell\psi$, for any $\psi \perp r_0, r_1$, into a function, orthogonal to q_0 and q_1. It follows that for any such ψ,

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi) = w_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi),$$

and the variance of the right-hand side is

$$\langle K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi, K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi \rangle_F = \langle \ell\psi, \ell\psi \rangle_F = \langle \psi, \psi \rangle_G.$$

This means that for $\psi \perp r_0, r_1$,

$$\widehat{v}_F(K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0} \ell\psi) \stackrel{d}{=} w_G(\psi) = \widehat{v}_G(\psi),$$

as claimed. □

Finally, for parametric families with an m-dimensional parameter, we use induction. Given $j \in \{0, 1, \dots, m-1\}$, suppose we have a unitary operator $U_{q,\ell r}(j)$ that maps ℓr_i to q_i for $0 \le i \le j$. For example, we have constructed above $U_{q,\ell r}(0) = K_{q_0,\ell r_0}$ and $U_{q,\ell r}(1) = K_{q_1,\widetilde{\ell r_1}} K_{q_0,\ell r_0}$. Now define the function

$$\widetilde{\ell r_{j+1}} := U_{q,\ell r}(j) \ell r_{j+1}$$

and introduce

$$U_{q,\ell r}(j+1) = K_{q_{j+1},\widetilde{\ell r_{j+1}}} U_{q,\ell r}(j).$$

Then $U_{q,\ell r}(j+1)$ is a unitary operator that maps ℓr_i to q_i for $0 \le i \le j+1$. Continuing in this fashion, we see that $U_{q,\ell r}(m)$ is a unitary operator that maps ℓr_i to q_i for $0 \le i \le m$. Therefore we obtain our final statement.

Proposition 3.5 ([7])**.** *If distributions F_θ and G_θ are equivalent and if q and r are orthonormal systems of $m+1$ functions (as described above) from $L_2(F_\theta)$ and $L_2(G_\theta)$, respectively, then*

$$\widehat{v}_F(U_{q,\ell r}(m) \ell\psi) \stackrel{d}{=} \widehat{v}_G(\psi). \tag{20}$$

It can be proved by an argument analogous to the previous case.

4. The Case of Point Processes. Unitary Transformations Again

In this section we describe similarities in rotation between the situation with parametric families of distribution and parametric models for intensities of point process. Let us consider a sequence of point processes N_n with (random) intensity functions $\lambda_{n,\theta}$ and compensators

$$\Lambda_{n,\theta}(t) = \int_0^t \lambda_{n,\theta}(s)\,\mathrm{d}s.$$

One of the key facts for us is that, if $\Lambda_{n,\theta}(t)/n$ converges to a deterministic function, say $B(t)$, as $n \to \infty$, then the normalized martingale

$$w_{n,\theta} = \frac{1}{\sqrt{n}}[N_n(t) - \Lambda_{n,\theta}(t)]$$

converges to the Brownian motion (see, e.g., [2,3]), while the same process with the estimated parameter $w_{n,\widehat{\theta}}$, can be approximated by a projection of $w_{n,\theta}$:

$$w_{n,\widehat{\theta}}(t) = w_{n,\theta}(t) - \frac{1}{n}\int_0^t [\dot{\lambda}/\lambda]^{\mathsf{T}}_{n,\theta}(s)\lambda_{n,\theta}(s)\,\mathrm{d}s\, R^{-1}_{n,\theta}\int_0^T [\dot{\lambda}/\lambda]_{n,\theta}(s)\,\mathrm{d}w_{n,\theta}(s) + o_P(1).$$

The key regularity assumptions ate such that

$$[\dot{\lambda}/\lambda]_{n,\theta}(t) \to \alpha(t), \quad \frac{1}{n}\lambda_{n,\theta}(t) \to \beta(t), \quad n \to \infty, \tag{21}$$

for some deterministic functions α and β. As a consequence, we expect that

$$w_{n,\theta} \xrightarrow{d} w_B, \quad w_{n,\widehat{\theta}} \xrightarrow{d} \widehat{w}_B,$$

with $B(t) = \int_0^t \beta(s)\,\mathrm{d}s$, and

$$\widehat{w}_B(t) = w_B(t) - \int_0^t \alpha^{\mathsf{T}}(s)\beta(s)\,\mathrm{d}s\, R^{-1}_{\theta}\int_0^T \alpha(s)\,\mathrm{d}w_B(s). \tag{22}$$

If we have another parametric model with the same regularity assumptions, then we will end up with another Brownian motion $w_{\widetilde{B}}(t)$, in time $\widetilde{B}$, and another projection $\widehat{w}_{\widetilde{B}}(t)$, parallel to a different score function $\widetilde{\alpha}$. If the parameters in the two cases are of the same dimension, then it again becomes possible to "rotate" $\widehat{w}_B(t)$ into $\widehat{w}_{\widetilde{B}}(t)$, and back if we wish. The form of the unitary operator needed for this task will be exactly the same as the one we have obtained for the case of i.i.d. samples.

There is, however, one difference that for the case of empirical processes the first coordinate of the score function always is the function $q_0(\cdot) = 1$, while this is not the case for the point processes: the first coordinate of the vector α may be any function, square-integrable with respect to the limiting "time" B.

As in the i.i.d. case, there is a question how to choose the "standard" problem in which to rotate the other problems. Indeed, one has here multiplicity of choices. As a simple choice, we suggest below to use Poisson processes with a variable intensity. At the first glance this looks somewhat strange, because then the function $\lambda_{n,\theta}$ will be a deterministic function from the very beginning and the regularity assumptions (21) will be easily satisfied. This is surprisingly simple, but, on the other hand, it is convenient.

Specifically, let us start with the space $L_2(\omega)$ of square-integrable functions on $[0,T], T \leq \infty$, with a weight function ω and choose orthonormal functions $p_0, \ldots, p_{m-1}$ from $L_2(\omega)$, i.e., such that

$$\int_0^T p_j(s)p_k(s)\omega(s)ds = \delta_{j,k}.$$

One example can be given by the Laguerre polynomials with $\omega(t) = e^{-t}$, $t \geq 0$. If we agree to consider a finite time horizon $T < \infty$, then it would be natural to use the constant weight function ω. Define now the intensity function

$$\mu_{n,\theta}(t) = n \exp\bigg(\sum_{j=0}^{m-1} \theta_j p_j(t)\bigg)\omega(t), \quad 0 \leq t \leq T.$$

Another possibility is to choose

$$\mu_{n,\theta}(t) = n \exp\bigg(\sum_{j=0}^{m-1} \theta_j p_j(t)\omega^{1/2}(t)\bigg), \quad 0 \leq t \leq T. \tag{23}$$

In this latter case one can choose as a true "target" distribution the distribution of the time-homogeneous Poisson process with intensity n. This distribution is a part of the parametric family above with the vector $\theta = 0$. As the target parametric family we choose distributions of Poisson processes with parameter $\theta = (\theta_0, \theta_1, \ldots, \theta_{m-1})^{\mathsf{T}}$, which takes values in a small open neighbourhood of 0. We need an open neighbourhood such that differentiation with respect to θ will not meet with difficulties, and it suffices to have this neighbourhood small. For this neighbourhood, at $\theta^* = (0, 0, \ldots, 0)^{\mathsf{T}}$ and $t \in [0, T]$ we have

$$[\dot{\mu}/\mu]_{n,\theta^*}(t) = \Big(p_j(t)\omega^{1/2}(t)\Big)_{j=0}^{m-1},$$

while

$$\frac{1}{n}\mu_{n,\theta^*}(t) = 1,$$

and one can easily see the consequence of the assumption of orthonormality of the functions $(p_j)_{j=0}^{m-1}$: the matrix

$$R_{\theta^*} = \bigg[\int_0^T p_j(t)p_k(t)\omega(t)dt\bigg]_{j,k=0}^{m-1} = I.$$

Therefore the coordinates of $[\dot{\mu}/\mu]_{n,\theta^*}$ are already ortho-normal.

Now we can show, in more or less explicit form, the rotation of $w_{n,\widehat{\theta}}(t)$ into the process, $\tilde{w}_{n,\widehat{\theta}}(t)$, which would arise from our Poisson model above. As in Section 3, it is notationally convenient to introduce orthonormal version of the vector-functions $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ and of these functions for Poisson process. For the intensity $\lambda_{n,\theta}(t)$ we could have done it already after Proposition 2.1. Namely, denote

$$q_{n,\theta}(t) = R_{n,\theta}^{-1/2}[\dot{\lambda}/\lambda]_{n,\theta}(t).$$

This is the vector-function with orthonormal coordinates in $L_2(\Lambda_{n,\theta}/n)$:

$$\frac{1}{n}\int q_{n,\theta}(t)q_{n,\theta}^{\mathsf{T}}(t)\lambda_{n,\theta}(t)\mathrm{d}t = I.$$

The limits of $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ and $\frac{1}{n}\lambda_{n,\theta}(t)$ in (21) suggest the limiting form of this vector-function:

$$q(t) = R_\theta^{-1/2}\alpha(t) = (q_0(t), q_1(t), \ldots, q_{m-1}(t))^{\mathsf{T}}$$

with the orthonormality property

$$\int q(t)q(t)^{\mathsf{T}}\beta(t)\mathrm{d}t = I.$$

For our Poisson process we have already the vector-function $[\dot{\mu}/\mu]_{n,\theta^*}$, whose coordinates are orthonormal on $[0, T]$. It does not change with n.

Now the procedure will look literary the same as the rotation of Brownian bridges v_F and v_G. Adopting (23) as the target parametric model, denote

$$M_{n,\theta}(t) = \int_0^t \mu_{n,\theta}(s)ds, \quad \text{so that} \quad M_{n,\theta^*}(t) = nt.$$

Now choose a function ℓ (the Hellinger function) as

$$\ell_{n,\theta}(t) = \left(\frac{\mathrm{d}M_{n,\theta^*}}{\mathrm{d}\Lambda_{n,\theta}}(t)\right)^{1/2} = \left(\frac{\mu_{n,\theta^*}}{\lambda_{n,\theta}}(t)\right)^{1/2}$$

or

$$\ell_{n,\theta}(t) = \left(\frac{n}{\lambda_{n,\theta}(t)}\right)^{1/2}.$$

Thus, if $\psi \in L_2(M_{n,\theta^*}/n)$, then $\ell\psi \in L_2(\Lambda_{n,\theta}/n)$.

In limiting form, this expression becomes

$$\ell_\theta(t) = \left(\frac{1}{\beta(t)}\right)^{1/2},$$

and if $\psi \in L_2(M)$, then $\ell\psi \in L_2(B)$, where, as above, $B(t) = \int_0^t \beta(s)\mathrm{d}s$. Thus, for the possibility to rotate to the Poisson model we need to require that ℓ be well defined, that is, $\lambda_{n,\theta}(t) > 0$ and $\beta(t) > 0$ for all $t > 0$.

In the expression of $K_{a,b}$ (see (19)), we first will go straight to the limiting expressions, that is, we will prepare for the case of large n. If it happens that the result of our rotation behaves close to what is expected in the Poisson case, then we will fond out that the values of n, which we have used in our simulations, are "large enough". We can use expressions for finite n and compare the outcomes later.

Let us use $a = q_0(t)$ and $b = \ell p_0(t) = \ell(t)$. This leads to the transformation

$$\hat{w}_B(K_{q_0,\ell}\ell\psi) \stackrel{d}{=} \hat{w}_M(\psi),$$

and again, if we choose $\psi = p_0$, then $K_{q_0,\ell}\ell p_0 = q_0$, and therefore

$$\hat{w}_B(K_{q_0,\ell}\ell p_0) = \hat{w}_M(p_0) = 0,$$

which is, certainly, correct.

As the next step, we create the function $K_{q_0,\ell}\ell p_1 = \widetilde{\ell p_1}$ and use it to construct our next operator, $K_{q_1,\widetilde{\ell p_1}}$. The product $K_{q_1,\widetilde{\ell p_1}} K_{q_0,\ell}$ will map ℓp_0 and ℓp_1 into q_0 and q_1, respectively. Now we have, again,

$$\hat{w}_B(K_{q_1,\widetilde{\ell p_1}} K_{q_0,\ell}\psi) \stackrel{d}{=} \hat{w}_M(\psi),$$

and if the parameter θ is two-dimensional, then this equality is the final result. For a general dimension m we proceed as in the previous section: for

$$U_{q,p}(0) = K_{q_0,\ell} \quad \text{and} \quad U_{q,p}(1) = K_{q_1,\widetilde{\ell p_1}} U_{q,p}(0)$$

we continue with

$$\widetilde{\ell p_j} = U_{q,p}(j-1)\ell p_j$$

and then define

$$U_{q,p}(j) = K_{q_1,\widetilde{\ell p_j}} U_{q,p}(j-1).$$

It is the final operator $U_{q,p}(m)$ which will be needed in the sequel: a unitary operator which will map $p_0, \dots, p_{m-1}$ into $q_0, \dots, q_{m-1}$, and, therefore, will map all functions, orthogonal to $p_0, \dots, p_{m-1}$, to functions, orthogonal to $q_0, \dots, q_{m-1}$.

The situation should be clearer described in terms of subspaces. Decompose $L_2(M)$ into the subspace $\mathcal{L}(p)$ spanned by $p_0, \dots, p_{m-1}$ and its orthogonal complement $\mathcal{L}_\perp(p)$. Similarly, decompose $L_2(B)$ into the subspace $\mathcal{L}(q)$ spanned by $q_0, \dots, q_{m-1}$ and its orthogonal complement $\mathcal{L}_\perp(q)$. Then considering multiplication by ℓ as an isometry from $L_2(M)$ to $L_2(B)$,

$$\psi \in L_2(M) \implies \ell\psi \in L_2(B), \quad \|\psi\| = \|\ell\psi\|,$$

then this operator will map $\mathcal{L}_\perp(p)$ into some subspace in $L_2(B)$,

$$\ell\mathcal{L}_\perp(p) \subset L_2(B).$$

Then it is the operator $U_{q,p}(m)$, as an operator in $L_2(B)$, acting on $\ell\psi$, which will map $\ell\mathcal{L}_\perp(p)$ into $\mathcal{L}(q)$:

$$U_{q,p}(m)\ell p_j = q_j, \quad j = 1, \dots, m.$$

It would be better to consider why the mapping of any testing problem for intensities of the point process, with only usual regularity assumptions, is basically the same problem always. This is true because in any model with these regularity assumptions we will end up with a Brownian motion in some time B – it will be specific for the model, and with a projection of this Brownian motion, parallel to the functions $q_0, \dots, q_{m-1}$ – also specific for the model. While the method of unitary mapping remains applicable and the same.

We will need to apply the operator $U_{q,p}(m)$ to empirical processes with estimated parameters, that is, to the situation with finite n. Then we need to be sure that the transformed process $w_{n,\hat\theta}(U_{q,p}\ell\psi)$, $\psi \in \Psi$, where $\Psi \subset L_2(M)$ is a class of functions of our choice, does converge in distribution to the limiting process $\hat w_M(\psi)$, $\psi \in \Psi$. The most natural choice will be the set of indicator functions $\psi_t(s) = \mathbb{1}_{\{s\le t\}})$ indexed by $t \ge 0$. It is obvious that as the function-parametric process, $\hat w_M(\psi_t)$ coincides with its point-parametric version $\hat w_M(t)$, and therefore the transformed empirical process $w_{n,\hat\theta}(U_{q,p}\ell\psi_t)$ should asymptotically behave as the point-parametric projected Brownian motion $\hat w_M(t)$.

It is very interesting to see what will be the graph of "rotated" ψ_t, that is, the graph of $U_{q,p}\ell\psi_t$. A sample of three graphs is shown in Figures 5.1 and 5.2 in the case of a point process model described in Example A of the next section. There the parameter is two-dimensional and we wished to transform the process into the projected Poisson process described above. The graphs have been calculated by S. Umut Can.

5. Some Specific Examples

Before we turn to specific examples, let us have a look on the expression of the limiting process (22) in the situation when the parameter of the intensity $\lambda_{n,\theta}$ of the point process is one-dimensional. In this situation α is a scalar function and $R_\theta = \int_0^T \alpha^2(s)\beta(s)\mathrm{d}s$ is a number. Then, considering the integral from α with respect to $\widehat w_B$:

$$\int_0^t \alpha(s)\mathrm{d}\widehat w_B(s) = \int_0^t \alpha(s)\mathrm{d}w_B(s) - \frac{\int_0^t \alpha^2(s)\beta(s)\,\mathrm{d}s}{\int_0^T \alpha^2(s)\beta(s)\mathrm{d}s}\int_0^T \alpha(s)\,\mathrm{d}w_B(s),$$

we see that the right-hand side is just the Brownian bridge in time

$$\tau = \frac{\int_0^t \alpha^2(s)\beta(s)\,\mathrm{d}s}{\int_0^T \alpha^2(s)\beta(s)\mathrm{d}s}, \quad t \in [0,T].$$

Therefore, all classical goodness of fit statistics from the Brownian bridge will be distribution free as statistics from the process $\int_0^t \alpha(s)\mathrm{d}\widehat w_B(s)$. The projection argument behind $\widehat w_B$ was used here, but to achieve distribution freeness no "rotation" was necessary. Full details are given in [9].

Example A. Consider a sequence of point processes $N_n(t)$ with compensated form

$$N_n(t) - \int_0^t c_\theta(t')[n - N_n(t')]dt';$$

in other words, the difference above is a martingale. Here we choose c_θ as the failure rate of Weibull distribution

$$c_\theta(t) = \frac{f_\theta(t)}{1 - F_\theta(t)} = \frac{\theta_1}{\theta_0}\left(\frac{t}{\theta_0}\right)^{\theta_1 - 1},$$

with parameters such that the corresponding Weibull's distribution behaves close to the distribution of life-times of, say, New Zealand population. These values are $\theta_0 = 86$ and $\theta_1 = 9$.

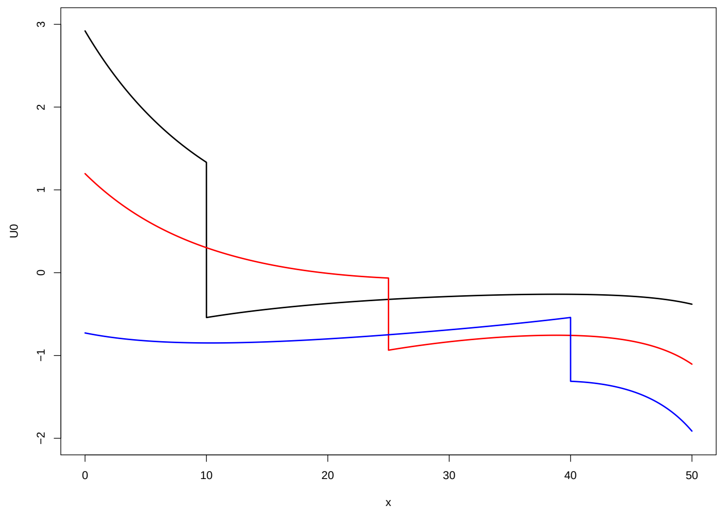

FIGURE 5.1. These are images of indicator functions $\mathbb{1}_{\{s \le t\}}$ for $t = 10, 25$ and 40 after first rotation by the operator $K_{q_0, \ell p_0}\ell$.

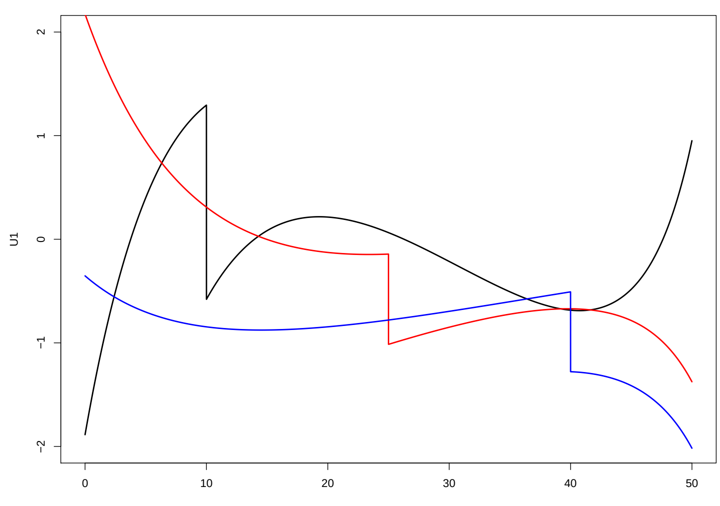

FIGURE 5.2. These are images of indicator functions $\mathbb{1}_{\{s \le t\}}$ for $t = 10, 25$ and 40 after two rotations, i.e., by the operator $U_{q,p}\ell$. Who would think that if you integrate these three functions with respect to $dw_{\hat{\theta},n}(s)$ the resulting three integrals will asymptotically jointly behave as $\hat{w}_M(t), t = 10, 25,$ and 40?

We know that our process is, actually, a binomial process based on n i.i.d. observations from the distribution with the failure rate c_θ, i.e., from Weibull's distribution. If we would center N_n by $nF_\theta(t)$ and normalize by $\sqrt{n}$, we would obtain an empirical process, of which the limiting process will be the F_θ-Brownian bridge. Centered as in the above display, and again normalized by $\sqrt{n}$, we obtain a basic martingale (cf. [1, 5, 10]), and its weak limit will be the F_θ-Brownian motion.

The vector-function $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ is now two-dimensional,

$$\frac{\dot{\lambda}_{n,\theta}}{\lambda_{n,\theta}}(t) = \frac{\dot{c}_{n,\theta}}{c_{n,\theta}}(t) = \left(-\frac{\theta_1}{\theta_0}, \frac{1}{\theta_1} + \ln\frac{t}{\theta_0}\right)^{\mathsf{T}}.$$

The function $\lambda_{n,\theta}/n$ and its limit is

$$\frac{1}{n}\lambda_{n,\theta}(t) = c_\theta(t)\frac{n - N_n(t)}{n} \to \beta_{\theta_0,\theta_1}(t),$$

where

$$\beta_{\theta_0,\theta_1}(t) = \frac{\theta_1}{\theta_0}\left(\frac{t}{\theta_0}\right)^{\theta_1 - 1}\exp\left(-(\frac{t}{\theta_0})^{\theta_1}\right)$$

is density of the Weibull distribution. The distribution function itself, in the current parametrisation, is $F_\theta(t) = 1 - \exp\left(-(\frac{t}{\theta_0})^{\theta_1}\right)$.

The covariance matrix R in its limiting form becomes

$$R_\theta = \int \begin{bmatrix} (\theta_1/\theta_0)^2, & -(1 + \theta_1\ln(t/\theta_0))/\theta_0 \\ -(1 + \theta_1\ln(t/\theta_0))/\theta_0, & (1 + \theta_1\ln(t/\theta_0))^2/\theta_1^2 \end{bmatrix} \beta_{\theta_0,\theta_1}(t)dt$$

or, changing the variable t to $\tau = t/\theta_0$ and separating the constant terms, we obtain a slightly simpler expression

$$R_\theta = \int \begin{bmatrix} (\theta_1/\theta_0)^2, & -(1 + \theta_1\ln\tau)/\theta_0 \\ -(1 + \theta_1\ln\tau)/\theta_0, & (1 + \theta_1\ln\tau)^2/\theta_1^2 \end{bmatrix} \beta_{1,\theta_1}(\tau)d\tau.$$

We note, as a side remark, that the matrix under the integral sign is, certainly, degenerate for every t, but the matrix R_θ is non-degenerate, it is invertible.

Now consider the integral on the anti-diagonal of this matrix. Since

$$\frac{d}{d\vartheta}\vartheta t^{\vartheta-1} = (1 + \vartheta\ln t)t^{\vartheta-1},$$

one can write

$$\int (1 + \vartheta\ln t)t^{\vartheta-1}\theta_1\exp\left(-t^{\theta_1}\right)dt = \frac{d}{d\vartheta}\int \vartheta t^{\vartheta-1}\theta_1\exp\left(-t^{\theta_1}\right)dt.$$

In the last integral we change the variable $t^{\theta_1} = z$ so that $t = z^{1/\theta_1}, dt = (1/\theta_1)z^{1/\theta_1 - 1}$. This leads to

$$\begin{aligned}\frac{d}{d\vartheta}\int \vartheta t^{\vartheta-1}\theta_1\exp\left(-t^{\theta_1}\right)dt &= \frac{d}{d\vartheta}\int \vartheta z^{(\vartheta-1)/\theta_1}\exp(-z)\,z^{1/\theta_1 - 1}dz \\ &= \frac{d}{d\vartheta}\int \vartheta z^{(\vartheta/\theta_1 - 1)}\exp(-z)\,dz = \frac{d}{d\vartheta}\vartheta\Gamma(\frac{\vartheta}{\theta_1})\end{aligned}$$

which at $\vartheta = \theta_1$ becomes $\dot{\Gamma}(1)$. This implies that we know explicitly the elements of the matrix R_θ, except one integral

$$R_\theta = \begin{bmatrix} \left(\frac{\theta_1}{\theta_0}\right)^2, & -\frac{1}{\theta_0}\dot{\Gamma}(1) \\ -\frac{1}{\theta_0}\dot{\Gamma}(1), & \frac{1}{\theta_1^2}\int(1 + \theta_1\ln\tau)^2\beta_{1,\theta_1}(\tau)d\tau \end{bmatrix}.$$

Example B. One real life situation where this process appears is, as we said, the analysis of life times in human populations. However, in general human populations the huge bulk of life times belongs to the interval of 50-100 years. For example, according to New Zealand life tables for 2012-14 for general populations 50 (years) is only 4%-point and 100 (years) is about 99%-point. Therefore, it makes sense to analyse the life times only after age of fifty. If X_i is a life time of an i-th individual, then we consider $X_{i,x_0} = \max(0, X_i - x_0)$ and the point process N_{n,x_0} based on these "over the threshold" values, and then we can choose x_0 equal to 50, or to any other value of interest. We can also assume that we know how many people of age over x_0 we have in the population under study. Thus, for

$$N_{n,x_0}(t) = \sum_{i=1}^{n} \mathbb{1}(X_{i,x_0} \leq t),$$

we have the representation

$$N_{n,x_0}(t) - \int_0^t c_\theta(x_0 + t')[n - N_{n,x_0}(t')]dt',$$

where the difference is a martingale. The functions $[\dot{\lambda}/\lambda]_{n,\theta}(t)$ and $\lambda_{n,\theta}/n$ now take the form

$$\frac{\dot{\lambda}_{n,\theta}}{\lambda_{n,\theta}}(t) = \frac{\dot{c}_{n,\theta}}{c_{n,\theta}}(x_0 + t) = \left(-\frac{\theta_1}{\theta_0}, \frac{1}{\theta_1} + \ln(\frac{t + x_0}{\theta_0})\right)^{\mathsf{T}}$$

and

$$\frac{1}{n}\lambda_{n,\theta}(t) = c_\theta(x_0 + t)\frac{n - N_{n,x_0}(t)}{n} \to \frac{\beta_{\theta_0,\theta_1}(x_0 + t)}{1 - F_{\theta_0,\theta_1}(x_0)}.$$

The matrix R_θ will also change, but in an obvious way. It is more interesting to note that we will need, in applications, to consider life-times not exceeding some value x_1, say, $x_1 = 100$, so that $N_{n,x_0}(t)$ will be stopped at some duration $x_1 - x_0$, equal, say, to 50 years (of life over age 50).

Example C. Now consider the same situation, but with $n - N_n(t)$ replaced by the process of "those at risk" (see, e.g., [1]). More specifically, consider a sequence of pairs $(X_i, C_i)_{i=1}^n$, where X_i is the survival time of i-th individual, and C_i is a censoring random variable of this survival time. Our main interest is in these survival times, however, one can only observe $\tilde{X}_i = \min(X_i, C_i)$ together with the indicator function $\delta_i = \mathbb{I}(X_i = \tilde{X}_i) = \mathbb{I}(X_i < C_i)$. The point process of interest is given as

$$N_n^c(t) = \sum_{i=1}^{n} \mathbb{1}(\tilde{X}_i \leq t)\delta_i$$

which counts the number of "genuine" survival times observed no later than t. Another point process, of those at risk at time t is given as

$$Y_n(t) = \sum_{i=1}^{n} \mathbb{1}(\tilde{X}_i \geq t).$$

With the help of this process, the process N_n^c can be compensated to the martingale as follows:

$$N_n^c(t) - \int_0^t c_\theta(t')Y_n(t')dt', \tag{24}$$

and the difference is a martingale (see [1]). Here c_θ, as in Example A, is the failure rate (or the force of mortality in demographic applications) of the hypothetical distribution F_θ, depending on parameter θ.

If one is interested in computer simulation of N_n^c, one should somehow choose not only parametric family F_θ, of interests for practitioner, but also a distribution G of truncating variables C_i's. Evolution in time of Y_n will strongly depend on this choice. However, this evolution is not looked at too much and evolution of N_n^c is studied, as it is implied by (24).

With $\lambda_{n,\theta}(t) = c_\theta(t)Y_n(t)$, let us clarify now the limit behaviour of the functions $\lambda_{n,\theta}(t)/n$ and $[\dot{\lambda}/\lambda]_{n,\theta}$. From the definition of $Y_n(t)$ and the Law of Large Numbers, it follows that, as $n \to \infty$,

$$c_\theta(t)\frac{Y_n(t)}{n} \to c_\theta(t)[1 - F_\theta(t)][1 - G(t)] = f_\theta(t)[1 - G(t)],$$

while

$$[\dot{\lambda}/\lambda]_{n,\theta}(t) = [\dot{c}/c]_\theta(t).$$

Therefore, if we choose $c_\theta(t)$ the same as in Example A, i.e., the failure rate of Weibull's distribution, then all will not differ from what we said in that example. However, this time the limit of $\lambda_{n,\theta}(t)/n$ is not a probability density.

Example D. Marked point processes. The example is interesting and has many applications, but is not treated here. We think it will be another example of regular models which permit the treatment as described in Section 3.

Acknowledgement

I would like to thank Prof. Robert Mnatsakanov, WVU, for exceptionally careful reading of the text at the final stages and for useful and fruitful discussions that followed.

References

1. P. K. Andersen, O. Borgan, R. D. Gill, N. Keiding, *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.
2. D. J. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes. vol. I. Elementary Theory and Methods*. Second edition. Probability and its Applications (New York). Springer-Verlag, New York, 2003; *vol. II. General Theory and Structure*. Second edition. Probability and its Applications (New York). Springer, New York, 2008.
3. A. F. Karr, *Point Processes and their Statistical Inference*. Second edition. Probability: Pure and Applied, 7. Marcel Dekker, Inc., New York, 1991.
4. E. Khmaladze, The use of Omega-square Tests for Testing Parametric Hypotheses. *Theory Probab. Appl.* **24** (1979), no. 2, 283–302.
5. E. Khmaladze, Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability and Its Applications* **26** (1982), no. 2, 240–257.
6. E. Khmaladze, Note on distribution free testing for discrete distributions. *Ann. Statist.* **41** (2013), no. 6, 2979–2993.
7. E. Khmaladze, Unitary transformations, empirical processes and distribution free testing. *Bernoulli* **22** (2016), no. 1, 563–588.
8. E. Khmaladze, Distribution free testing for conditional distributions given covariates. *Statist. Probab. Lett.* **129** (2017), 348–354.
9. L. Roberts, Distribution free testing of goodness of fit in a one dimensional parameter space. *Statist. Probab. Lett.* **99** (2015), 215–222.
10. G. R. Shorack, J. A. Wellner, *Empirical Processes with Applications to Statistics*. SIAM, Philadelphia, PA, 2009.

(Received 17.02.2020)

Victoria University of Wellington, Wellington, New Zealand
E-mail address: Estate.Khmaladze@vuw.ac.nz