CHAPTER 1

CLASSES OF FIXED-ORDER AND ADAPTIVE SMOOTH GOODNESS-OF-FIT TESTS WITH DISCRETE RIGHT-CENSORED DATA

Edsel A. Peña

Department of Statistics
University of South Carolina
Columbia, SC 29208 USA
E-mail: pena@stat.sc.edu

Classes of hazard-odds based fixed-order and adaptive smooth goodness-of-fit tests for the composite hypothesis that an unknown discrete distribution belongs to a family of distributions using right-censored observations are presented. The proposed classes of tests generalize Neyman's [33] smooth class of tests. The class of fixed-order tests is the discrete analog of the hazard-based class of tests for continuous failure times studied in Peña[35]. The class of adaptive tests employs a modified Schwartz[40] Bayesian information criterion for choosing the order of the embedding class, with the modification on the criterion accounting for the incompleteness mechanism.

1. Introduction

Statistical goodness-of-fit (gof) testing has always been an active research area as evidenced by entering the phrase "goodness of fit" in the `MathSciNet` search engine. In its simplest form a random sample $T_1, T_2, \ldots, T_n$ from an unknown distribution function F is observed, and it is desired to determine if $F = F_0$, where F_0 is a specified distribution. The most well-known gof procedure is Pearson's[34] chi-square test which utilizes the statistic

$$\chi^2 = \sum_{j=1}^{K} \frac{(O_j - E_j)^2}{E_j}, \tag{1}$$

where K is the size of the partition of the support of F_0, O_j is the number of T_i's in the jth member of the partition, and E_j is the number of T_i's expected to be in the jth member of the partition when F_0 holds. The popularity of this test is partly due to its simplicity and the fact that it requires only critical values from the family of chi-square distributions. There are other tests for the simple gof problem, such as Kolmogorov-Smirnov (KS) type tests, Neyman's[33] smooth gof tests, Cramer-von Mises (CVM) type tests, and those by Khamaladze[20,21]. A review of some of these procedures could be found in Stephens[42]. Many of these tests have extensions to the composite null hypothesis setting, where the problem is to test whether $F \in \mathcal{C}$, with $\mathcal{C}$ a specified (parametric) family of distributions, cf., Chernoff and Lehmann[6], Rao and Robson[37], D'Agostino and Stephens[9], and Greenwood and Nikulin[13]. Except for Pearson's[34] test, most of the above-mentioned procedures imposes the restriction that F is continuous, with this assumption typically made in order to facilitate the derivations of distributional results.

Though not as prevalent as the case with continuous distributions, gof tests for discrete distributions, or when data arose from grouping of continuous data, have also been considered. Kulperger and Singh[26] examined χ^2 gof tests for discrete distributions and considered the issue of random grouping. Cressie and Read[8] introduced the family of power divergence statistics for performing gof with multinomial data. Best and Rayner[4,5] proposed Neyman smooth gof tests for the null hypothesis that F is geometric and Poisson, respectively; while Eubank[10] proposed Neyman smooth-type tests for dealing with multinomial data. In Choulakian, Lockhart and Stephens[7] a test for the discrete uniform was presented; while in Spinelli and Stephens[41] CVM-type procedures were developed for testing a Poisson distribution. Kocherlakota and Kocherlakota[24], Rueda, Perez-Abreau and O'Reilly[38], Baringhaus and Henze[3], and Nakamura and Perez-Abreau[32] examined gof procedures for discrete data using the empirical probability generating function; in particular, tests for the Poisson distribution were developed. Empirical distribution-based methods were also considered for discrete models. Among papers adopting this approach were Henze[15] and Klar[23]. However, all of these papers dealing with goodness-of-fit for discrete models assume that $T_1, T_2, \ldots, T_n$ are completely observed.

In biomedical, engineering, reliability, and in other areas where the primary variable of interest is the time-to-occurrence of an event, hereon referred to as a failure time, it is typical that some of the failure times will be right-censored due to time constraints, limited resources, withdrawal from the study, loss to follow-up, etc. Numerous papers have appeared dealing with the modeling and analysis of failure times in the presence of incomplete observations. For continuous failure times, the problem of gof testing has been addressed in several papers with the aim of extending to censored data those procedures that were developed for complete data. Among these papers are those of Koziol and Green[25], Hyde[19], Hollander and Proschan[17], Nair[29,30,31], Gatsonis, Hsieh and Korwar[11], Habib and Thomas[14], Akritas[2], Hjort[16], Hollander and Peña[18], Li and Doss[28], and Kim[22]. An interesting goal in gof testing with censored data is to extend Pearson's test. The difficulty underlying such an extension is that the exact number of failures in a member of the partition is not observable. An attempt to extend Neyman's smooth gof procedure in the presence of right-censored data has also been made by Gray and Pierce[12]. Their approach parallels that of Neyman[33] where the density function is embedded in a wider class. A different extension of the smooth gof tests with continuous failure times, which adapts naturally to censored data and enables point process theory, was that in Peña[35] and Agustin and Peña[1], the latter dealing with reliability models for recurrent events.

Except for the test proposed in Hyde[19] which is a special case of the class of tests proposed in this chapter, the gof problem with right-censored discrete failure times does not seem to have been investigated extensively in the literature. The existing gof procedures for discrete and complete data mentioned earlier have not yet been extended for discrete and censored data, which is rather surprising since discrete failure times are ubiquitous in many studies. For instance, discrete failure times occur because of the intrinsic nature of the failure time process such as when failure is measured in terms of counts or the number of cycles, or due to an inherent limitation in the measurement process forcing subjects to be observed only at the end of specified intervals (e.g., weekly basis). Discrete failure times also manifest when the times are interval-censored as in biomedical studies, or when data is presented in a life-table format as is done in actuarial settings. Right-censoring occurs due to the withdrawal of subjects from the study, a fixed study period, or due to failure (death) from competing causes. In these situations, prior to performing higher-level statistical analysis such as estimation or hypothesis testing, it is desirable to know the parametric family of distributions or hazards to which F or Λ belongs since this will enable the use of more efficient inferential methods.

This chapter aims to provide a general class of gof tests for discrete failure times and in the presence of right-censoring for the composite null hypothesis. In Peña[36] a general approach for generating a class of tests for the simple null hypothesis case was presented, an approach which is a hazard-based extension of Neyman's[33] smooth goodness-of-fit tests. See

Rayner and Best[39] for an extensive discussion of the Neyman formulation of this class of smooth goodness-of-fit tests. The present chapter considers the parallel treatment of the composite null hypothesis case. The procedures presented in this chapter are discrete analogs of the intensity-based smooth goodness-of-fit tests developed in Peña[35] for continuous failure times. In this formulation, the sequence of odds associated with the hazard rates are embedded in a wider class, in contrast to the usual Neyman formulation where the sequence of probabilities are embedded, cf., Rayner and Best[39]. This intensity-based embedding facilitates the derivation of the smooth goodness-of-fit tests as score tests, thereby endowing the tests with certain local optimality properties. In contrast to the development of Pearson's test in which the vantage point is the time origin and the underlying question is: 'How many observations are expected to have values in a member of the partition of the support of F_0?' the current approach's vantage point is dynamic in that the relevant question is: 'Given that just before a certain time point there are a certain number of units at risk, how many are expected to fail at this time point?' Consequently, instead of dealing with global probabilities, the main focus are conditional probabilities, hazards, or intensities, which are the natural quantities when dealing with dynamic or time-evolving systems.

Due to space and time constraints, proofs of the propositions and theorems will not be presented in this chapter, but we focus instead on the proposed class of goodness-of-fit procedures. Results of simulation studies pertaining to the achieved levels and powers will be presented in the paper containing the proofs of the propositions and theorems. We mention that simulation studies performed for the tests associated with the simple null hypothesis case demonstrated the viability of the proposed class of tests and indicates that the proposed adaptive test using the modified Schwartz information criterion could be used as an omnibus test. Results of these simulation studies can be found in Peña[36].

2. Description of the Problem

Let $T_1, T_2, \ldots, T_n$ be independent and identically distributed (IID) random variables from an unknown discrete distribution F whose support is known to be $\mathcal{A} = \{a_1, a_2, \ldots\}$ with $a_i < a_{i+1}, i = 1, 2, \ldots$. The T_i's are not completely observed, but only the random vectors $(Z_1, \delta_1), (Z_2, \delta_2), \ldots, (Z_n, \delta_n)$ are observed with the interpretation that $\delta_i = 1$ implies $T_i = Z_i$, whereas $\delta_i = 0$ implies $T_i > Z_i$. Let $\lambda_j = \lambda_j(F), j = 1, 2, \ldots$ be the hazard of T at a_j, so $\lambda_j = \mathbf{P}(T = a_j | T \geq a_j) = \Delta F(a_j)/\bar{F}(a_j-)$, and let $\Lambda(t) = \sum_{j=1}^{\infty} \lambda_j I\{a_j \leq t\}, t \in \Re$, be the discrete hazard function associated with F. We assume in the sequel the independent censoring condition:

$$\mathbf{P}\{T = a_j | T \geq a_j\} = \lambda_j = \mathbf{P}\{T = a_j | Z \geq a_j\}, \quad j = 1, 2, \ldots. \tag{2}$$

The problem dealt with is to test the hypothesis that F belongs to a parametric class $\mathcal{F}_0$ of discrete distributions parameterized by a q-dimensional vector η taking values in Γ, an open set in $\Re^q$. Denote by $\mathcal{C}_0$ the class of hazard functions associated with $\mathcal{F}_0$ so $\mathcal{C}_0 = \{\Lambda_0(\cdot|\eta) : \eta \in \Gamma\}$, where the functional form of $\Lambda_0(\cdot|\eta)$ is known. The goodness-of-fit problem is to test the composite hypotheses

$$H_0 : \Lambda(\cdot) \in \mathcal{C}_0 \quad \text{versus} \quad H_1 : \Lambda(\cdot) \notin \mathcal{C}_0 \tag{3}$$

on the basis of the right-censored data $(Z_i, \delta_i), i = 1, 2, \ldots, n$. The simple null hypothesis case where interest is on testing

$$H_0 : \Lambda(\cdot) = \Lambda_0(\cdot) \quad \text{versus} \quad H_1 : \Lambda(\cdot) \neq \Lambda_0(\cdot) \tag{4}$$

with $\Lambda_0(\cdot)$ a fully specified discrete hazard function was dealt with in Peña[36]. The present chapter extends the results in Peña[36] to the composite case. Note that in (3), the parameter vector η is a nuisance parameter.

3. Hazard Embeddings and Likelihoods

Let $\lambda_j^0(\eta), j = 1, 2, \ldots$ be the hazards associated with $\Lambda_0(\cdot|\eta)$, so

$$\Lambda_0(t|\eta) = \sum_{\{j: a_j \leq t\}} \lambda_j^0(\eta).$$

Following Peña[36], for $\lambda_j < 1$ and $\lambda_j(\eta) < 1$, let the hazard odds be

$$\rho_j = \frac{\lambda_j}{1 - \lambda_j} \quad \text{and} \quad \rho_j^0(\eta) = \frac{\lambda_j^0(\eta)}{1 - \lambda_j^0(\eta)}.$$

For a fixed smoothing order $p \in \mathcal{Z}_+$, and for the $p \times 1$ vectors $\mathbf{\Psi}_j = \mathbf{\Psi}_j(\eta), j = 1, 2, \ldots, J$, we embed $\rho_j^0(\eta)$ into the hazard odds determined by

$$\rho_j(\theta, \eta) = \rho_j^0(\eta) \exp\{\theta^{\mathrm{t}} \mathbf{\Psi}_j(\eta)\}, \quad j = 1, 2, \ldots; \theta \in \Re^p. \tag{5}$$

This is equivalent to postulating that the logarithm of the hazard odds ratio is linear in $\mathbf{\Psi}_j(\eta)$, that is,

$$\log\left\{\frac{\rho_j(\theta, \eta)}{\rho_j^0(\eta)}\right\} = \theta^{\mathrm{t}} \mathbf{\Psi}_j(\eta), \quad j = 1, 2, \ldots.$$

Within this embedding, the partial likelihood of (θ, η) based on the observation period $(-\infty, a_J]$ for some fixed $J \in \mathcal{Z}_+$ is (cf., Peña[36])

$$L(\theta, \eta) = \prod_{j=1}^{J} \frac{\rho_j(\theta, \eta)^{O_j}}{[1 + \rho_j(\theta, \eta)]^{R_j}} \tag{6}$$

where

$$O_j = \sum_{i=1}^{n} I\{Z_i = a_j, \delta_i = 1\} \quad \text{and} \quad R_j = \sum_{i=1}^{n} I\{Z_i \geq a_j\}.$$

Furthermore, within this hazard odds embedding, the composite goodness-of-fit problem simplifies to testing $H_0 : \theta = 0, \eta \in \Gamma$ versus $H_1 : \theta \neq 0, \eta \in \Gamma$, so η is a nuisance parameter. For our notation, we shall denote by $\nabla_v = \partial/\partial v$ the gradient operator with respect to a vector v. The test is to be anchored by the estimated score statistic

$$U_\theta(0, \hat{\hat{\eta}}) = \nabla_\theta \log L(\theta, \eta)|_{\theta=0, \eta=\hat{\hat{\eta}}},$$

where $\hat{\hat{\eta}} = \hat{\eta}(\theta = 0)$ is the restricted partial likelihood maximum likelihood estimator (RPLMLE). This is the η that maximizes the restricted partial likelihood function

$$L_0(\eta) = L(0, \eta) = \prod_{j=1}^{J} \frac{\rho_j(0, \eta)^{O_j}}{[1 + \rho_j(0, \eta)]^{R_j}} = \prod_{j=1}^{J} [\lambda_j^0(\eta)]^{O_j} [1 - \lambda_j^0(\eta)]^{R_j - O_j} \tag{7}$$

with $\rho_j(0, \eta) = \rho_j^0(\eta) = \lambda_j^0(\eta)/[1 - \lambda_j^0(\eta)]$.

4. Restricted Partial Likelihood MLE

From (7), the logarithm of the partial likelihood function is

$$l_0(\eta) = \log L_0(\eta) = \sum_{j=1}^{J} \left\{ O_j \log \lambda_j^0(\eta) + (R_j - O_j) \log[1 - \lambda_j^0(\eta)] \right\}. \tag{8}$$

Consequently,

$$\nabla_\eta l_0(\eta) = \sum_{j=1}^{J} \mathbf{A}_j(\eta)[O_j - E_j^0(\eta)]$$

where, for $j = 1, 2, \ldots$,

$$E_j^0(\eta) = R_j \lambda_j^0(\eta) \quad \text{and} \quad \mathbf{A}_j(\eta) = \frac{\nabla_\eta \lambda_j^0(\eta)}{\lambda_j^0(\eta)[1 - \lambda_j^0(\eta)]}$$

are the $q \times 1$ 'standardized' gradients of $\lambda_j^0(\eta)$ with respect to η. We form the $J \times q$ matrix of standardized gradients

$$\mathbf{A}(\eta) = [\mathbf{A}_1(\eta), \mathbf{A}_2(\eta), \ldots, \mathbf{A}_J(\eta)]^{\mathrm{t}}, \tag{9}$$

and define the $J \times 1$ vectors

$$\mathbf{O} = (O_1, O_2, \ldots, O_J)^{\mathrm{t}} \quad \text{and} \quad \mathbf{E}^0(\eta) = \left(E_1^0(\eta), E_2^0(\eta), \ldots, E_J^0(\eta)\right)^{\mathrm{t}}.$$

Then, in matrix form,

$$\nabla_\eta l_0(\eta) = \mathbf{A}(\eta)^{\mathrm{t}} \left[\mathbf{O} - \mathbf{E}^0(\eta)\right].$$

The estimating equation for the RPLMLE $\hat{\hat{\eta}}$ is therefore

$$\mathbf{A}(\eta)^{\mathrm{t}} \left[\mathbf{O} - \mathbf{E}^0(\eta)\right] = \mathbf{0}. \tag{10}$$

For example, suppose that $\mathcal{C}_0$ is the class of constant hazards, which corresponds to the class of geometric distributions. Then $\mathbf{A}(\eta) = \mathbf{1}_J/[\eta(1-\eta)]$ and $\mathbf{E}^0(\eta) = \mathbf{R}\eta$, so the estimating equation becomes

$$\{\eta(1-\eta)\}^{-1}\mathbf{1}_J^{\mathrm{t}}(\mathbf{O} - \mathbf{R}\eta) = 0,$$

yielding the RPLMLE given by

$$\hat{\hat{\eta}} = \frac{\mathbf{1}_J^{\mathrm{t}}\mathbf{O}}{\mathbf{1}_J^{\mathrm{t}}\mathbf{R}} = \frac{\sum_{j=1}^J O_j}{\sum_{j=1}^J R_j}. \tag{11}$$

Clearly, in many situations, $\hat{\hat{\eta}}$ will need to be obtained iteratively or through numerical methods.

5. Asymptotics and Test Procedure

The logarithm of the partial likelihood function is given by

$$l(\theta, \eta) = \sum_{j=1}^{J} \left\{O_j \log \rho_j(\theta, \eta) - R_j \log[1 + \rho_j(\theta, \eta)]\right\}.$$

With $\mathbf{\Psi}(\eta) = [\mathbf{\Psi}_1(\eta), \mathbf{\Psi}_2(\eta), \ldots, \mathbf{\Psi}_J(\eta)]^{\mathrm{t}}$, the score function for θ, evaluated at $\theta = \mathbf{0}$, is immediately obtained to be

$$\mathbf{U}_\theta(\theta = \mathbf{0}, \eta) = \nabla_\theta l(\theta, \eta)|_{\theta=\mathbf{0}} = \mathbf{\Psi}(\eta)^{\mathrm{t}}[\mathbf{O} - \mathbf{E}^0(\eta)]. \tag{12}$$

Since η is unknown, this score function is estimated by

$$\hat{\mathbf{U}}_\theta = \mathbf{U}_\theta(\theta = \mathbf{0}, \hat{\hat{\eta}}) = \mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}}[\mathbf{O} - \mathbf{E}^0(\hat{\hat{\eta}})]. \tag{13}$$

To develop the test we need the asymptotic distribution of $\hat{\mathbf{U}}_\theta$. For this purpose, we first present the joint asymptotic distribution of the $(p+q) \times 1$ vector of scores

$$\mathbf{U}(\eta) = \begin{bmatrix} \mathbf{\Psi}(\eta)^{\rm t} \\ \mathbf{A}(\eta)^{\rm t} \end{bmatrix} [\mathbf{O} - \mathbf{E}^0(\eta)] \tag{14}$$

at $\eta = \eta_0$, the true value of η under H_0.

To achieve a more compact notation, for a vector $\mathbf{v}$, we denote by $\mathrm{Diag}(\mathbf{v})$ the diagonal matrix whose diagonal elements are those of $\mathbf{v}$. Let

$$\mathbf{D}(\eta) = \mathrm{Diag}\left(\lambda_j(\eta)[1 - \lambda_j(\eta)] : j = 1, 2, \ldots, J\right)$$

and $\lambda(\eta) = \left(\lambda_1(\eta), \lambda_2(\eta), \ldots, \lambda_J(\eta)\right)^{\rm t}$. Then, the matrix of standardized gradients could be re-expressed via

$$\mathbf{A}(\eta) = \mathbf{D}(\eta)^{-1} \nabla_{\eta^{\rm t}} \lambda(\eta). \tag{15}$$

The asymptotic distribution of $\mathbf{U}(\eta)$ can be obtained by invoking Theorem 4 in Peña[36]. To describe this asymptotic distribution, we need to introduce more notation. Let

$$\mathbf{V}(\eta) = \mathrm{Diag}(\mathbf{R})\mathbf{D}(\eta)$$

and with $\mathbf{B}(\eta) = [\mathbf{\Psi}(\eta), \mathbf{A}(\eta)]$, define the $(p+q) \times (p+q)$ matrix

$$\mathbf{\Xi}(\eta) = \mathbf{B}(\eta)^{\rm t} \mathbf{V}(\eta) \mathbf{B}(\eta).$$

Furthermore, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$, let

$$\begin{aligned} V_{ij} &= I\{Z_i = a_j, \delta_i = 1\}; \\ W_{ij} &= I\{Z_i \geq a_j\}; \\ U_{ij}(\eta) &= V_{ij} - W_{ij}\lambda_j^0(\eta), \end{aligned}$$

and

$$\mathcal{F}_j = \bigvee_{i=1}^{n} \sigma\{W_{i1}, V_{i1}, W_{i2}, V_{i2}, \ldots, W_{ij}, V_{ij}, W_{ij+1}\}.$$

From Theorem 4 in Peña[36] we obtain the following proposition.

Proposition 1: *Assume that H_0 holds and that the true value of η is η_0. Furthermore, suppose that p does not change with n and for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$, the following conditions hold:*

(i) the jth row of $\mathbf{B}(\eta_0)$, which is $\mathbf{B}_j(\eta_0) = [\mathbf{\Psi}_j(\eta_0)^{\rm t}, \mathbf{A}_j(\eta_0)^{\rm t}]$, is $\mathcal{F}_{j-1}$-measurable and $\boldsymbol{E}\{[\parallel \mathbf{B}_j(\eta_0) \parallel U_{ij}]^2\} < \infty$;

(ii) there exists a $(p+q) \times (p+q)$ *positive definite matrix* $\mathbf{\Xi}^{(0)}(\eta_0)$ *such that, as* $n \to \infty$*,*

$$n^{-1}\mathbf{\Xi}(\eta_0) \xrightarrow{\text{pr}} \mathbf{\Xi}^{(0)}(\eta_0);$$

(iii) with $V_{jj}(\eta_0) = R_j \lambda_j(\eta_0)[1-\lambda_j(\eta_0)]$*, then as* $n \to \infty$*,*

$$\max_{1 \le j \le J} trace\left\{[\mathbf{\Xi}(\eta_0)]^{-1}[\mathbf{B}_j(\eta_0)^{\text{t}} V_{jj}(\eta_0)\mathbf{B}_j(\eta_0)]\right\} \xrightarrow{\text{pr}} 0;$$

(iv) as $n \to \infty$*,* $\max_{1 \le j \le J} \parallel \mathbf{B}_j(\eta_0) \parallel^2 = O_p(1)$.

Then, as $n \to \infty$*,*

$$\frac{1}{\sqrt{n}}\mathbf{U}(\eta_0) = \frac{1}{\sqrt{n}}\mathbf{B}(\eta_0)^{\text{t}}[\mathbf{O} - \mathbf{E}^0(\eta_0)] \xrightarrow{\text{d}} N_{p+q}(\mathbf{0}, \mathbf{\Xi}^{(0)}(\eta_0)).$$

Marginalizing on the score function for θ, it follows from Proposition 1 that

$$\frac{1}{\sqrt{n}}\mathbf{\Psi}(\eta_0)^{\text{t}}[\mathbf{O} - \mathbf{E}^0(\eta_0)] \xrightarrow{\text{d}} N_p(\mathbf{0}, \mathbf{\Xi}_{11}^{(0)}(\eta_0)) \tag{16}$$

where $\mathbf{\Xi}_{11}^{(0)}(\eta_0)$ is the in-probability limit of $n^{-1}\mathbf{\Psi}(\eta_0)^{\text{t}}\mathbf{V}(\eta_0)\mathbf{\Psi}(\eta_0)$. Of course this result is not directly useful for constructing the test since η_0 is not known; however, it will become useful later when ascertaining the impact of the estimation of η_0 by $\hat{\hat{\eta}}$. For later use, we also denote by $\mathbf{\Xi}_{12}^{(0)}(\eta_0) = \mathbf{\Xi}_{21}^{(0)}(\eta_0)^{\text{t}}$ the in-probability limit of $n^{-1}\mathbf{\Psi}(\eta_0)^{\text{t}}\mathbf{V}(\eta_0)\mathbf{A}(\eta_0)$ and by $\mathbf{\Xi}_{22}^{(0)}(\eta_0)$ the in-probability limit of $n^{-1}\mathbf{A}(\eta_0)^{\text{t}}\mathbf{V}(\eta_0)\mathbf{A}(\eta_0)$.

We are now ready to present the asymptotic result which will be useful for constructing the goodness-of-fit procedure.

Theorem 1: *Assume that the conditions of Proposition 1 hold, and in addition there exists a neighborhood* Γ_0 *of* η_0 *in* Γ *such that, as* $n \to \infty$*,*

(i) for each $j = 1, 2, \ldots, J$*,* $\lambda_j(\eta)$ *is twice-differentiable with* $\eta \mapsto \nabla_\eta \lambda_j(\eta)$ *continuous at* $\eta = \eta_0$*;* $\max_{1 \le j \le J} \parallel \nabla_\eta \lambda_j(\eta) \parallel = O_p(1)$*, and for each* $l, l' \in \{1, 2, \ldots, q\}$*,*

$$\max_{1 \le j \le J} \sup_{\eta \in \Gamma_0} \left| \frac{\partial^2}{\partial \eta_l \eta_{l'}} \lambda_j(\eta) \right| = o_p(n);$$

(ii) for each $i = 1, 2, \ldots, n$ *and* $j = 1, 2, \ldots, J$*,* $\Psi_{ij}(\eta)$ *is twice-differentiable with* $\eta \mapsto \nabla_\eta \Psi_{ij}(\eta)$ *continuous at* $\eta = \eta_0$*;*

$$\max_{1 \le i \le n} \max_{1 \le j \le J} \parallel \nabla_\eta \Psi_{ij}(\eta) \parallel = O_p(1),$$

and for each $l, l' \in \{1, 2, \ldots, q\}$,

$$\max_{1 \le i \le n} \max_{1 \le j \le J} \sup_{\eta \in \Gamma_0} \left| \frac{\partial^2}{\partial \eta_l \eta_{l'}} \Psi_{ij}(\eta) \right| = o_p(n);$$

(iii) the limiting matrix $\mathbf{\Xi}_{22}^{(0)}(\eta_0)$ *is nonsingular.*

Then, under H_0 *and as* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}} [\mathbf{O} - \mathbf{E}^0(\hat{\hat{\eta}})] \xrightarrow{\mathrm{d}} N_p \left(\mathbf{0}, \mathbf{\Xi}_{11.2}^{(0)}(\eta_0) \right),$$

where $\mathbf{\Xi}_{11.2}^{(0)}(\eta_0) = \mathbf{\Xi}_{11}^{(0)}(\eta_0) - \mathbf{\Xi}_{12}^{(0)}(\eta_0) \left\{ \mathbf{\Xi}_{22}^{0)}(\eta_0) \right\}^{-1} \mathbf{\Xi}_{21}^{(0)}(\eta_0)$.

Comparing this result with that in (16), we see the effect of estimating the unknown parameter η_0 by the RPLMLE $\hat{\hat{\eta}}$ is to decrease the covariance matrix by the term $\mathbf{\Xi}_{12}^{(0)}(\eta_0) \left\{ \mathbf{\Xi}_{22}^{0)}(\eta_0) \right\}^{-1} \mathbf{\Xi}_{21}^{(0)}(\eta_0)$. Also, by recalling the definitions of the matrices $\mathbf{\Xi}_{ij}^{(0)}(\eta_0)$'s, it is immediate that the limiting covariance matrix $\mathbf{\Xi}_{11.2}^{(0)}(\eta_0)$ can be estimated consistently by

$$\begin{aligned} \hat{\mathbf{\Xi}}_{11.2}^{(0)} = \frac{1}{n} \Big\{ & \mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}} \mathbf{V}(\hat{\hat{\eta}}) \mathbf{\Psi}(\hat{\hat{\eta}}) - \\ & [\mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}} \mathbf{V}(\hat{\hat{\eta}}) \mathbf{A}(\hat{\hat{\eta}})] [\mathbf{A}(\hat{\hat{\eta}})^{\mathrm{t}} \mathbf{V}(\hat{\hat{\eta}}) \mathbf{A}(\hat{\hat{\eta}})]^{-1} [\mathbf{A}(\hat{\hat{\eta}})^{\mathrm{t}} \mathbf{V}(\hat{\hat{\eta}}) \mathbf{\Psi}(\hat{\hat{\eta}})] \Big\}. \end{aligned} \quad (17)$$

With $\mathbf{M}^-$ denoting a generalized inverse of a matrix $\mathbf{M}$, the test statistic for testing H_0 and a fixed smoothing order p is

$$\hat{S}_p^2 = \left\{ \frac{1}{\sqrt{n}} \mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}} [\mathbf{O} - \mathbf{E}^0(\hat{\hat{\eta}})] \right\}^{\mathrm{t}} \left\{ \hat{\mathbf{\Xi}}_{11.2}^{(0)} \right\}^{-} \left\{ \frac{1}{\sqrt{n}} \mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}} [\mathbf{O} - \mathbf{E}^0(\hat{\hat{\eta}})] \right\}. \quad (18)$$

Corollary 1: *Under the conditions of Theorem 1 and under* H_0*, as* $n \to \infty$, $\hat{S}_p^2 \xrightarrow{\mathrm{d}} \chi_{p^*}^2$ *with* $p^* = rank(\mathbf{\Xi}_{11.2}^{(0)}(\eta_0))$*. Therefore, an asymptotic* α*-level test of* H_0 *versus* H_1 *rejects* H_0 *whenever* $\hat{S}_p^2 > \chi_{\hat{p}^*;\alpha}^2$ *with* $\hat{p}^* = rank(\hat{\mathbf{\Xi}}_{11.2}^{(0)})$*, and where* $\chi_{p;\alpha}^2$ *is the* $100(1-\alpha)$*th percentile of a* χ_p^2 *distribution.*

To further simplify our notation, let

$$\mathbf{A}^*(\eta) = \mathbf{V}(\eta)^{\frac{1}{2}} \mathbf{A}(\eta) \quad \text{and} \quad \mathbf{\Psi}^*(\eta) = \mathbf{V}(\eta)^{\frac{1}{2}} \mathbf{\Psi}(\eta),$$

and for a full rank $J \times q$ (with $J > q$) matrix $\mathbf{X}$, let

$$P(\mathbf{X}) = \mathbf{X}(\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{t}}$$

be the projection operator (matrix) on the linear subspace $\mathcal{L}(\mathbf{X})$ generated by $\mathbf{X}$ in $\Re^J$. Also, denote by

$$P^{\perp}(\mathbf{X}) = \mathbf{I} - P(\mathbf{X})$$

the projection operator on the orthocomplement of $\mathcal{L}(\mathbf{X})$. Using these notation, the estimator $\hat{\mathbf{\Xi}}_{11.2}^{(0)}$ can be reexpressed via

$$\hat{\mathbf{\Xi}}_{11.2}^{(0)} = \frac{1}{n}\mathbf{\Psi}^*(\hat{\hat{\eta}})^{\mathrm{t}} P^{\perp}(\mathbf{A}^*(\hat{\hat{\eta}}))\mathbf{\Psi}^*(\hat{\hat{\eta}}). \tag{19}$$

Let us also define the 'standardized' observed and *dynamic* expected frequencies via

$$\mathbf{O}^* = \mathbf{V}(\hat{\hat{\eta}})^{-\frac{1}{2}}\mathbf{O} = \left(\frac{O_j}{\sqrt{R_j\lambda_j^0(\hat{\hat{\eta}})[1-\lambda_j^0(\hat{\hat{\eta}})]}} : j = 1, 2, \ldots, J\right)^{\mathrm{t}};$$

$$\mathbf{E}^*(\hat{\hat{\eta}}) = \mathbf{V}(\hat{\hat{\eta}})^{-\frac{1}{2}}\mathbf{E}^0(\hat{\hat{\eta}}) = \left(\frac{R_j\lambda_j^0(\hat{\hat{\eta}})}{\sqrt{R_j\lambda_j^0(\hat{\hat{\eta}})[1-\lambda_j^0(\hat{\hat{\eta}})]}} : j = 1, 2, \ldots, J\right)^{\mathrm{t}}$$

with the convention that $0/0 = 0$.

Using these standardized quantities, and upon further simplification, the test statistic can be expressed as

$$\hat{S}_p^2 = [\mathbf{O}^* - \mathbf{E}^*(\hat{\hat{\eta}})]^{\mathrm{t}} \left[\mathbf{\Psi}^*(\hat{\hat{\eta}}) \left\{\mathbf{\Psi}^*(\hat{\hat{\eta}})^{\mathrm{t}} P^{\perp}(\mathbf{A}^*(\hat{\hat{\eta}}))\mathbf{\Psi}^*(\hat{\hat{\eta}})\right\}^{-} \mathbf{\Psi}^*(\hat{\hat{\eta}})^{\mathrm{t}}\right] [\mathbf{O}^* - \mathbf{E}^*(\hat{\hat{\eta}})]. \tag{20}$$

Under an orthogonality condition between $\mathbf{A}^*(\hat{\hat{\eta}})$ and $\mathbf{\Psi}^*(\hat{\hat{\eta}})$, we further obtain the more compact and norm-like nature of the statistic given in the following corollary. This corollary also implies that under the orthogonality condition, the estimation of η_0 by $\hat{\hat{\eta}}$ does not require any adjustments in the limiting covariance matrix relative to the case when η_0 is known, an 'adaptiveness' property.

Corollary 2: *If $\mathbf{\Psi}^*(\hat{\hat{\eta}})$ lies in $\mathcal{L}(\mathbf{A}^*(\hat{\hat{\eta}}))^{\perp}$, the orthocomplement of $\mathcal{L}(\mathbf{A}^*(\hat{\hat{\eta}}))$, then*

$$\hat{S}_p^2 = \| P(\mathbf{\Psi}^*(\hat{\hat{\eta}}))[\mathbf{O}^* - \mathbf{E}^*(\hat{\hat{\eta}})] \|^2 .$$

For purposes of studying the asymptotic local power properties of the test, Theorem 1 could be generalized to cover the behavior under local alternatives. This generalization is contained in the following theorem.

Theorem 2: *If the conditions of Theorem 1 hold, then under the sequence of local alternatives* $H_1^{(n)} : \theta^{(n)} = n^{-\frac{1}{2}}\gamma(1+o(1))$ *for* $\gamma \in \Re^p$ *and as* $n \to \infty$,

$$\frac{1}{\sqrt{n}}\mathbf{\Psi}(\hat{\hat{\eta}})^{\mathrm{t}}[\mathbf{O} - \mathbf{E}^0(\hat{\hat{\eta}})] \xrightarrow{\mathrm{d}} N_p\left(\mathbf{\Xi}_{11.2}^{(0)}(\eta_0)\gamma, \mathbf{\Xi}_{11.2}^{(0)}(\eta_0)\right).$$

As a consequence, the asymptotic local power of the test described above for the sequence of local alternatives specified in Theorem 2 is

$$\mathrm{ALP}(\gamma) = \mathbf{P}\left\{\chi^2_{p^*}(\delta^2(\gamma)) > \chi^2_{p^*;\alpha}\right\}, \tag{21}$$

where the noncentrality parameter is

$$\delta^2(\gamma) = \gamma^{\mathrm{t}}\mathbf{\Xi}_{11.2}^{(0)}(\eta_0)\gamma,$$

which could be consistently estimated by

$$\hat{\delta}^2 = \frac{1}{n}[\mathbf{\Psi}^*(\hat{\hat{\eta}})\gamma]^{\mathrm{t}}P^{\perp}(\mathbf{A}^*(\hat{\hat{\eta}}))[\mathbf{\Psi}^*(\hat{\hat{\eta}})\gamma].$$

Under the orthogonality condition of Corollary 2 this simplifies to

$$\hat{\delta}^2 = \frac{1}{n} \parallel \mathbf{\Psi}^*(\hat{\hat{\eta}})\gamma \parallel^2 \xrightarrow{\mathrm{pr}} \gamma^{\mathrm{t}}\mathbf{\Xi}_{11}(\eta_0)\gamma = \delta^2.$$

6. Some Choices of $\mathbf{\Psi}$

For a fixed smoothing order p, three particular choices of the $J \times p$ matrix $\mathbf{\Psi}(\eta)$ are provided below. The first specification is given by

$$\mathbf{\Psi}_1 = \left(\left(\frac{\mathbf{R}}{n}\right)^0, \left(\frac{\mathbf{R}}{n}\right)^1, \ldots, \left(\frac{\mathbf{R}}{n}\right)^{p-1}\right), \tag{22}$$

where

$$(\mathbf{R}/n)^k = ((R_1/n)^k, (R_2/n)^k, \ldots, (R_J/n)^k)^{\mathrm{t}}.$$

Note that this choice does not depend functionally on η, but its distribution depends on η. This choice has proven effective in goodness-of-fit testing for the simple null hypothesis setting for this discrete failure time setting[36], and as such we expect that this will also perform satisfactorily in this composite null hypothesis setting.

The second specification, which depends functionally on η, is

$$\mathbf{\Psi}_2(\eta) = \left([\lambda_0(\eta)]^0, [\lambda_0(\eta)]^1, \ldots, [\lambda_0(\eta)]^{p-1}\right), \tag{23}$$

where

$$[\lambda_0(\eta)]^k = ([\lambda_1^0(\eta)]^k, [\lambda_2^0(\eta)]^k, \ldots, [\lambda_J^0(\eta)]^k)^{\mathrm{t}}.$$

The analogous choice for the continuous failure time situation was quite effective in generating tests with commendable powers (cf., Peña[35]).

The third specification produces a test statistic which generalizes Pearson's statistic. Let $C_1, C_2, \ldots, C_p$ be a (disjoint) partition of $\mathcal{J} = \{1, 2, \ldots, J\}$. Define

$$\boldsymbol{\Psi}_3 = \left(\mathbf{1}_{C_1}, \mathbf{1}_{C_2}, \ldots, \mathbf{1}_{C_p}\right)^{\mathrm{t}} \tag{24}$$

where for $C \subseteq \mathcal{J}$, $\mathbf{1}_C$ is a $J \times 1$ vector whose jth element is $I\{j \in C\}$. Furthermore, define

$$O_\bullet(C) = \sum_{j \in C} O_j \quad \text{and} \quad \hat{E}^0_\bullet(C) = \sum_{j \in C} E^0_j(\hat{\eta}).$$

Also, with

$$\hat{V}^*_\bullet(C) = \mathbf{1}^{\mathrm{t}}_C \mathbf{V}(\hat{\eta})^{1/2} P^\perp(\mathbf{A}^*(\hat{\eta})\mathbf{V}(\hat{\eta})^{1/2}\mathbf{1}_C$$

the resulting test statistic for the specification (24) is given by

$$\hat{S}^2_p = \sum_{i=1}^{p} \frac{\left[O_\bullet(C_i) - \hat{E}^0_\bullet(C_i)\right]^2}{\hat{V}^*_\bullet(C_i)}, \tag{25}$$

which is a Pearson-type test statistic.

However, these choices do not satisfy the orthogonality condition in Corollary 2, so the correction term for the covariance matrix will be required. It is possible to start with these choices to arrive at a $\boldsymbol{\Psi}'$ that satisfies the orthogonality condition using a Gram-Schmidt type of orthogonalization. But, as pointed out in Peña[35] in the continuous failure time setting, the benefits of such a programme may not outweigh the effort and difficulty in performing the orthogonalization.

7. Adaptive Choice of Smoothing Order

The testing procedure described in the preceding section requires that the smoothing order p be fixed. This introduces an arbitrariness in the procedure, and without a good prior knowledge of the class of hazards that holds if the class under the null hypothesis does not hold, there is a great potential of choosing a p that is far from optimal. Of course, repeated testing with different smoothing orders is unwise since it will inflate the Type I error rates. It is therefore imperative and important to have a data-driven or adaptive approach for determining the smoothing order p.

We propose a procedure that uses a modified Schwartz information criterion (Schwartz[40]) to decide on the smoothing order p. We mention that for the classical Neyman's smooth goodness-of-fit test, Ledwina[27] proposed the use of the Schwartz information criterion for adaptively determining the smoothing order. Let $L_p(\theta_p, \eta)$ denote the partial likelihood of (θ_p, η) when the smoothing order p is given in (6), and let $l_p(\theta_p, \eta) = \log L_p(\theta_p, \eta)$ be the associated log-partial likelihood function. Denote by $(\hat{\theta}_p, \hat{\eta})$ the partial likelihood maximum likelihood estimator (PLMLE), so that

$$L_p(\hat{\theta}_p, \hat{\eta}) = \sup_{\theta_p \in \Re^p;\ \eta \in \Gamma} L_p(\theta_p, \eta).$$

Clearly, as in the computation of the RPLMLE $\hat{\hat{\eta}}$, numerical techniques will be needed to compute the PLMLE. Let $\mathbf{U}_p(\theta_p, \eta)$ and $\mathbf{I}_p(\theta_p, \eta)$ be the score function vector and observed Fisher information matrix associated with $L_p(\theta_p, \eta)$, respectively. Thus,

$$\mathbf{U}_p(\theta_p, \eta) = \begin{bmatrix} \nabla_\theta l_p(\theta_p, \eta) \\ \nabla_\eta l_p(\theta_p, \eta) \end{bmatrix};$$

and

$$\mathbf{I}_p(\theta_p, \eta) = -\begin{bmatrix} \frac{\partial^2}{\partial\theta_p \partial\theta_p^{\rm t}} l_p(\theta_p, \eta) & \frac{\partial^2}{\partial\theta_p \partial\eta^{\rm t}} l_p(\theta_p, \eta) \\ \frac{\partial^2}{\partial\theta_p^{\rm t} \partial\eta} l_p(\theta_p, \eta) & \frac{\partial^2}{\partial\eta \partial\eta^{\rm t}} l_p(\theta_p, \eta) \end{bmatrix}.$$

A possible approach to iteratively computing the PLMLE $(\hat{\theta}_p, \hat{\eta})$ is via the Newton-Raphson updating given by

$$\begin{bmatrix} \hat{\theta}_p \\ \hat{\eta} \end{bmatrix} \leftarrow \begin{bmatrix} \hat{\theta}_p \\ \hat{\eta} \end{bmatrix} + [\mathbf{I}_p(\hat{\theta}_p, \hat{\eta})]^{-1} \mathbf{U}_p(\hat{\theta}_p, \hat{\eta}). \tag{26}$$

Denote by $\hat{\lambda}_{\max}$ the largest eigenvalue of $\mathbf{I}_p(\hat{\theta}_p, \hat{\eta})$. The modified Schwartz information criterion is defined to be

$$\mathrm{MSIC}(p) = l_p(\hat{\theta}_p, \hat{\eta}) - \frac{p}{2}\left[\log(n) + \log(\hat{\lambda}_{\max})\right]. \tag{27}$$

The first two terms in the right-hand side of (27) is the usual Schwartz information criterion for complete data. The last term in (27) represents the correction arising from the right-censoring. The justification for this modification will be provided in more a general framework involving incomplete data in a forthcoming paper.

The order selection procedure and the associated goodness-of-fit test proceeds as follows: First, a value of $P_{\max}$, which represents the upper bound of the smoothing order is specified. We propose to set the value of $P_{\max}$ to 10, though it could be changed to some other value. Second, the smoothing order to be used in the test statistic, denoted by p^*, is the value of p that maximizes MSIC(p) for $p = 1, 2, \ldots, P_{\max}$, that is,

$$p^* = \arg\max_{1 \le p \le P_{\max}} \text{MSIC}(p). \tag{28}$$

Of course, note that this p^* is also a function of $P_{\max}$, although we suppress writing this explicitly. Finally, the asymptotic adaptive α-level test of H_0 versus H_1 rejects H_0 in favor of H_1 whenever $\hat{S}^2_{p^*} > \chi^2_{1;\alpha}$. The fact that the critical value is that associated with a one degree-of-freedom chi-square distribution follows from the following asymptotic result, whose proof will be presented in a forthcoming paper.

Theorem 3: *If the conditions of Theorem 1 hold, then under H_0 and as $n \to \infty$, $p^* \xrightarrow{\text{pr}} 1$ and $\hat{S}^2_{p^*} \xrightarrow{\text{d}} \chi^2_1$.*

For practical purposes, instead of using the asymptotic critical value of $\chi^2_{1;\alpha}$, for small to moderate sample sizes, we recommend the use of the test which rejects H_0 in favor of H_1 whenever $\hat{S}^2_{p^*} > \chi^2_{p^*;\alpha}$. Another possibility, as yet unexplored, is to approximate the appropriate critical value using a bootstrap procedure.

Acknowledgements

The author wishes to acknowledge the US National Science Foundation DMS Grant 0102870 which partially supported this research. Furthermore, he wishes to thank the Organizing Committee of MMR2002, Professors Bo Lindqvist and Kjell Doksum, for inviting him to present a talk in the Conference and to contribute this chapter in the Conference Proceedings.

References

1. Z. Agustin and E. Peña, Goodness-of-Fit of the Distribution of Time-to-First-Occurrence in Recurrent Event Models, *Lifetime Data Analysis*, **7**, 287-304 (2001).
2. M. Akritas, Pearson-type goodness-of-fit tests: the univariate case, *Journal of the American Statistical Association*, **83**, 222-230 (1988).
3. L. Baringhaus and N. Henze, A goodness of fit test for the Poisson distribution based on the empirical generating function, *Statistics and Probability Letters*, **13**, 269–274 (1992).
4. D. Best and J. Rayner, Goodness of fit for the geometric distribution, *Biometrical Journal*, **31**, 307–311 (1989).
5. D. Best and J. Rayner, Goodness of fit for the Poisson distribution, *Statistics and Probability Letters*, **44**, 259–265 (1999).

6. H. Chernoff and E. Lehmann, The use of maximum likelihood estimates in χ^2 tests for goodness of fit, *Annals of Mathematical Statistics*, **25**, 579–586 (1954).
7. V. Choulakian, R. Lockhart, and M. Stephens, Cramer-von Mises statistics for discrete distributions, *Canadian Journal of Statistics*, **22**, 125–137 (1994).
8. N. Cressie and T. Read, Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society, B*, **46**, 440–464 (1984).
9. R. D'Agostino and M. Stephens, *Goodness-of-Fit Techniques* (Marcel Dekker, Inc., New York, 1986).
10. R. Eubank, Testing goodness of fit with multinomial data, *Journal of the American Statistical Association*, **92**, 1084–1093 (1997).
11. C. Gatsonis, H. Hsieh and R. Korwar, Simple nonparametric tests for a known standard survival based on censored data, *Communications in Statistics — Theory & Methods A*, **14**, 2137–2162 (1985).
12. R. Gray and D. Pierce, Goodness-of-fit for censored survival data, *The Annals of Statistics*, **13**, 552–563 (1985).
13. P. Greenwood and M. Nikulin, *A Guide to Chi-Squared Testing* (John Wiley & Sons, Inc., New York, 1996).
14. M. Habib and D. Thomas, Chi-squared goodness-of-fit tests for randomly censored data, *The Annals of Statistics*, **14**, 759–765 (1986).
15. N. Henze, Empirical-distribution-function goodness-of-fit tests for discrete models, *Canadian Journal of Statistics*, **24**, 81–93 (1996).
16. N. Hjort, Goodness-of-fit tests in models for life history data based on cumulative hazard rates, *The Annals of Statistics*, **18**, 1221–1258 (1990).
17. M. Hollander and F. Proschan, Testing to determine the underlying distribution using randomly censored data, *Biometrics*, **35**, 193–401 (1979).
18. M. Hollander and E. Peña, A chi-squared goodness-of-fit test for randomly censored data, *Journal of the American Statistical Association*, **87**, 458–463 (1992).
19. J. Hyde, Testing survival under right censoring and left truncation, *Biometrika*, **64**, 225–230 (1977).
20. E. Khamaladze, A martingale approach in the theory of goodness-of-fit tests, *Teor. Veroyatnost. i Primenen*, **26**, 246–265 (1981) (English translation in *Theory Probab. Appl.*, **26**, 240–257).
21. E. Khamaladze, Goodness of fit problem and scanning innovation martingales, *The Annals of Statistics*, **21**, 798–829 (1993).
22. J. Kim, Chi-square goodness-of-fit tests for randomly censored data, *The Annals of Statistics*, **21**, 1621–1639 (1993).
23. B. Klar, Goodness-of-fit tests for discrete models based on the integrated distribution function, *Metrika*, **49**, 53–69 (1999).
24. S. Kocherlakota and K. Kocherlakota, Goodness of fit tests for discrete distributions, *Communications in Statistics — Theory & Methods A*, **15**, 815–829 (1986).
25. J. Koziol and S. Green, A Cramer-von Mises statistic for randomly censored data, *Biometrika*, **63**, 139-156 (1976).
26. R. Kulperger and A. Singh, On random grouping in goodness of fit tests

of discrete distributions, *Journal of Statistical Planning and Inference*, **7**, 109–115 (1982).

27. T. Ledwina, Data driven version of Neyman's smooth test of fit, *Journal of the American Statistical Association*, **89**, 1000–1005 (1994).
28. G. Li and H. Doss, Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data, *The Annals of Statistics*, **21**, 772–797 (1993).
29. V. Nair, Plots and tests for goodness of fit with randomly censored data, *Biometrika*, **68**, 99–103 (1982).
30. V. Nair, Goodness of fit tests for multiply right censored data, in *Nonparametric Statistical Inference, Vol. I, II* (Budapest, 1980), 653–666, Colloq. Math. Soc. János Bolyai, **32**(North-Holland, Amsterdam-New York, 1982).
31. V. Nair, Confidence Bands for Survival Functions with Censored Data: A Comparative Study, *Technometrics*, **26**, 265–275 (1984).
32. M. Nakamura and V. Perez-Abreau, Use of an empirical probability generating function for testing a Poisson model, *Canadian Journal of Statistics*, **21**, 149–156 (1993).
33. J. Neyman, "Smooth" test for goodness of fit, *Skand. Aktuarietidskr.*, **20**, 150–199 (1937).
34. K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling, *Philos. Mag.*, 5th ser., **50**, 157–175 (1900).
35. E. Peña, Smooth Goodness-of-Fit Tests for Composite Hypothesis in Hazard-Based Models, *The Annals of Statistics*, **26**, 1935–1971 (1998).
36. E. Peña, Intensity-Based Approach to Goodness-of-Fit Testing with Discrete Right-Censored Data. Technical report, Department of Statistics, University of South Carolina (2002).
37. K. Rao and D. Robson, A chi-square statistic for goodness-of-fit tests within the exponential family, *Communications in Statistics*, **3**, 1139–1153 (1974).
38. R. Rueda, V. Perez-Abreu and F. O'Reilly, Goodness of fit for the Poisson distribution based on the probability generating function, *Communications in Statistics — Theory & Methods A*, **20**, 3093–3110 (1991).
39. J. Rayner and D. Best, *Smooth Tests of Goodness of Fit* (Oxford University Press, New York, 1989).
40. G. Schwartz, Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461–464 (1978).
41. J. Spinelli and M. Stephens, Cramer-von Mises tests of fit for the Poisson distribution, *Canadian Journal of Statistics*, **25**, 257–268 (1997).
42. M. Stephens, Introduction to Kolmogorov (1933) On the empirical determination of a distribution, in *Breakthroughs in Statistics 2: Methodology and Distribution*, Eds. S. Kotz and N. Johnson (Springer: New York, 1992) pp.93–105.