Goodness-of-Fit Tests with Right-Censored Discrete Data

Edsel A. Peña
Department of Statistics
University of South Carolina
Columbia, SC 29208 USA
pena@stat.sc.edu

Dawn Garrison
Department of Statistics
University of South Carolina
Columbia, SC 29208 USA
dgarriso@stat.sc.edu

Abstract

A class of goodness-of-fit tests for right-censored discrete data is described. Finite and asymptotic properties of the test statistics are presented. Specific members of this class of tests are illustrated, and in particular, a class of tests for the null hypothesis that the discrete failure times are geometrically distributed is described. Under the geometric failure time setting, simulation studies indicate that an adaptive procedure based on a modified Schwarz Bayesian information criterion for choosing the smoothing order have excellent powers against a wide variety of alternatives, and thus could serve as an omnibus test.

1 Introduction

Let $T_1, T_2, \ldots, T_n$ be independent and identically distributed (IID) random variables representing the failure times of n reliability systems. In general, these could be the times to event occurrence for n units in a reliability/engineering study, a biomedical setting, an economic situation, etc. We assume that the common, but unknown, distribution function F of the T_i's is a discrete distribution with support $\mathcal{A} = \{a_1, a_2, a_3, \ldots\}$ where $a_j < a_{j+1}$. This may result for instance from situations where the failure time is measured in terms of counts, or it could be that the underlying failure time is continuous, but the observable is a discretized version of this failure time such as when the systems are monitored on a daily or weekly basis, or when the data comes in the form of a life-table.

In reliability, engineering, biomedical, and in other areas, however, it is typical to have some of the failure times to become right-censored due to time constraints, limited resources, competing causes of failure, loss to follow-up, etc. The observable data from the study therefore comes in the form

$$(Z_1, \delta_1), (Z_2, \delta_2), \ldots, (Z_n, \delta_n) \tag{1}$$

where $\delta_i \in \{0, 1\}$, and with the interpretation that $\delta_i = 1$ implies that $T_i = Z_i$, whereas $\delta_i = 0$ implies that $T_i > Z_i$. The problem considered in this paper is to test the null hypothesis that $F = F_0$ versus the alternative hypothesis that $F \neq F_0$, where F_0 is a known distribution function, based on a realization of the data in (1). This is called the simple hypothesis goodness-of-fit problem. The composite goodness-of-fit problem on the other hand tests the null hypothes that $F \in \mathcal{C}$ versus the alternative hypothesis that $F \notin \mathcal{C}$, where $\mathcal{C} = \{F(\cdot; \eta) : \eta \in \Gamma\}$ is a specified parametric family of distributions, e.g., the family of geometric distributions.

When the failure time distribution is continuous, several papers have appeared dealing with the goodness-of-fit problem in the presence of right-censoring. Among these papers are those by Koziol and Green (1976), Hyde (1977), Hollander and Proschan (1979), Nair (1981), Gatsonis, Hsieh and Korwar (1985), Habib and Thomas (1986), Akritas (1988), Hjort (1990), Hollander and Peña (1992), Li and Doss (1993), Kim (1993) and Peña (1998). Some of these papers extended the classical Pearson-type test to the censored setting. When the data is complete, the goodness-of-fit problem has also been dealt with for the case with discrete data. Some papers that addressed this issue are those by Kulperger and Singh (1982), Cressie and Read (1984), Best and Rayner (1989, 1999), Eubank (1997), Choulakian, Lockhart and Stephens (1994), Spinelli and Stephens (1997), Kocherlakota and Kocherlakota (1986), Rueda, Perez-Abreau and O'Reilly (1991), Baringhaus and Henze (1992), Nakamura and Perez-Abreau (1993), Henze (1996) and Klar (1999). Unfortunately, except for Hyde (1977), the goodness-of-fit problem with right-censored discrete data does not seem to have been investigated extensively. This is rather surprising, since discrete failure times are clearly ubiquitous in applied work in reliability, engineering, biomedical, and economic settings.

The present paper partially tackles this goodness-of-fit problem for right-censored discrete failure data by providing a class of procedures for performing the test for the simple hypothesis. The composite goodness-of-fit problem will be dealt with in another paper. The class of tests presented belongs to the class of procedures which are the discrete analogs of the intensity-based smooth goodness-of-fit tests in Peña (1998). The members of this class of tests are partially determined by specifying a smoothing order, which limits the utility of the procedures. To alleviate this problem, a modified Schwarz (1978) Bayesian information criterion based procedure for determining the smoothing order is described. This produces a test that have excellent sensitivity against a wide range of alternatives, rendering it to be a viable candidate as an omnibus test.

2 Tests with Fixed Smoothing Orders

For $j \in \{1, 2, \ldots\}$, let $\lambda_j = \Delta F(a_j)/[1 - F(a_j-)]$, where $\Delta F(a_j) = \Pr\{T_i = a_j | T_i \sim F\}$, so $\lambda_j = \Pr\{T_i = a_j | T_i \geq a_j\}$ is the hazard at a_j under F. Let λ_j^0 be the hazard at a_j under the null hypothesis distribution F_0. Define the hazard odds to be $\rho_j = \lambda_j/(1 - \lambda_j)$ and $\rho_j^0 = \lambda_j^0/(1 - \lambda_j^0)$. Also, denote by a_J the upper limit of the observation period for the n units, with J possibly depending on n. Fix a positive integer p, called the smoothing order, and let $\Psi = (\Psi_1, \ldots, \Psi_J)$ be a $p \times J$ matrix, which could be random. Denote by $\psi_i, i = 1, 2, \ldots, p$, the ith row vector of Ψ, and assume that these p $1 \times J$ vectors are linearly independent. The idea behind the development of the test procedure is to embed the hypothesized hazard vector $(\rho_1^0, \rho_2^0, \ldots, \rho_J^0)$ in the parameterized class $\mathcal{C}_p = \{(\rho_1(\theta), \rho_2(\theta), \ldots, \rho_J(\theta)) : \theta = (\theta_1, \ldots, \theta_p)' \in \Re^p\}$, where

$$\rho_j(\theta) = \rho_j^0 \exp\{\theta' \Psi_j\}, j = 1, 2, \ldots, J. \tag{2}$$

With this embedding, the class of goodness-of-fit tests for the equivalent hypothesis $H_0 : \theta = \mathbf{0}$ versus $H_1 : \theta \neq \mathbf{0}$ corresponds to the score test arising from the resulting model. The class of test statistics so derived is the discrete analog of the intensity-based smooth goodness-of-fit tests in Peña (1998) for continuous failure time data, which in turn is the generalization of the class of smooth tests by Neyman (1937) (see Rayner and Best (1989)).

For $j = 1, 2, \ldots$, let $O_j = \sum_{i=1}^n I\{Z_i = a_j, \delta_i = 1\}$ and $R_j = \sum_{i=1}^n I\{Z_i \geq a_j\}$. The relevant partial likelihood under the embedding model is given by $L_1^*(\theta) = \prod_{j=1}^J \rho_j(\theta)^{O_j}/[1 + \rho_j(\theta)]^{R_j}$. The associated score vector and observed information matrix are given, respectively, by

$$\mathbf{U}^*(\theta) = \sum_{j=1}^J \Psi_j\{O_j - R_j \lambda_j(\theta)\} \quad \text{and} \quad \mathbf{I}^*(\theta) = \sum_{j=1}^J \Psi_j \Psi_j' \lambda_j(\theta)[1 - \lambda_j(\theta)]. \tag{3}$$

Evaluating at $\theta = \mathbf{0}$, we obtain $\mathbf{U}_0^* = \sum_{j=1}^J \Psi_j\{O_j - R_j \lambda_j^0\}$ and $\mathbf{I}_0^* = \sum_{j=1}^J \Psi_j \Psi_j' \lambda_j^0 [1 - \lambda_j^0]$. With $\mathbf{O} = (O_1, O_2, \ldots, O_J)'$ and $\mathbf{E}_0 = (E_1^0, E_2^0, \ldots, E_J^0)'$ with $E_j^0 = R_j \lambda_j^0$, and $\mathbf{V}_0$ being the $J \times J$ diagonal matrix with diagonal elements $V_{jj}^0 = R_j \lambda_j^0 (1 - \lambda_j^0)$, then in matrix notation, $\mathbf{U}_0^* = \Psi(\mathbf{O} - \mathbf{E}_0)$ and $\mathbf{I}_0^* = \Psi \mathbf{V}_0 \Psi'$. The score test statistic for testing $H_0 : \theta = \mathbf{0}$ is the quadratic form

$$S_p^2 = (\mathbf{O} - \mathbf{E}_0)' \Psi' (\Psi \mathbf{V}_0 \Psi')^- \Psi (\mathbf{O} - \mathbf{E}_0). \tag{4}$$

By virtue of Theorem 1 which is stated below, under $H_0 : \theta = \mathbf{0}$, S_p^2 converges to a central chi-squared distribution with degrees-of-freedom k^*, which is the rank of Ξ, the in-probability limit of $\mathbf{I}_0^*/n = \Psi \mathbf{V}_0 \Psi'/n$ as $n \to \infty$. An asymptotic α-level test of $H_0 : \theta = \mathbf{0}$ versus $H_1 : \theta \neq \mathbf{0}$ is therefore of form: "Reject

H_0 whenever $S_p^2 \geq \chi^2_{\hat{k}^*;\alpha}$," where $\chi^2_{k;\alpha}$ denotes the $100(1-\alpha)$the percentile of χ^2_k, and $\hat{k}^*$ is the rank of $(\Psi \mathbf{V}_0 \Psi')/n$. The following theorem describes the null asymptotic distribution of the normalized score vector. Note that J may increase with n and Ψ may be random in this result. The proof and the exact technical conditions of this result can be found in Peña (2002).

Theorem 1 *Assume that $H_0 : \theta = \mathbf{0}$ holds. If Ψ is such that for each $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, J$, Ψ_j is measurable with respect to the σ-field induced by all information just before a_j and $E\{||\Psi||^2\} < \infty$; p does not change with n; and there exists a $p \times p$ positive definite matrix $\mathbf{\Xi}_0$ such that, as $n \to \infty$, $n^{-1}\mathbf{I}_0^* = n^{-1}\Psi \mathbf{V}_0 \Psi' \xrightarrow{pr} \mathbf{\Xi}_0$; $\max_{1 \leq j \leq J} \mathrm{trace}\{(\Psi \mathbf{V}_0 \Psi')^{-1}(\Psi_j V_{jj}^0 \Psi_j')\} \xrightarrow{pr} 0$; and $\max_{1 \leq j \leq J} ||\Psi_j||^2 = O_p(1)$, then $n^{-\frac{1}{2}} U_0^* = n^{-\frac{1}{2}} \Psi(\mathbf{O} - \mathbf{E}_0) \xrightarrow{d} N_p(\mathbf{0}, \mathbf{\Xi}_0)$. As a consequence, $S_p^2 \xrightarrow{d} \chi^2_{k^*}$.*

3 An Adaptive Test

An obvious deficiency of the test procedure described above is it requires the user to specify the smoothing order p. We describe an approach to determine adaptively the smoothing order. As before, let $L_p^*(\theta_p)$ be the (partial) likelihood of $\theta_p = (\theta_1, \ldots, \theta_p)$ and $\mathbf{I}_p^*(\theta_p)$ the observed information matrix. Denote by $\hat{\theta}_p$ the maximum (partial) likelihood estimator of θ_p, which will generally be obtained iteratively, e.g., via the Newton-Raphson iteration. For a fixed positive integer $P_m \leq J$, define

$$p^* \equiv p_{BICAdj} = \mathrm{argmax}_{1 \leq p \leq P_m} \left\{ \log L_p^*(\hat{\theta}_p) - \frac{p}{2}\left[\log(n) + \hat{\zeta}_1\right]\right\}, \tag{5}$$

where $\hat{\zeta}_1$ is the largest eigenvalue of $\mathbf{I}_p^*(\hat{\theta}_p)$. This is a modified Schwarz (1978) Bayesian information criterion (BIC) with the adjustment arising from the right-censoring. The theoretical justification for this adjustment will be provided in a more general setting in another paper. The adaptive test statistic is $S_{p^*}^2$, and the adaptive test rejects H_0 whenever $S_{p^*}^2$ exceeds $\chi^2_{\hat{k}^*;\alpha}$, where $\hat{k}^*$ is the rank of $\mathbf{I}_{p^*}^*(\mathbf{0})$.

4 Specific Choices of Smoothing Matrix

Let $\mathcal{J} = \{1, 2, \ldots, J\}$, and for a subset $A \subseteq \mathcal{J}$, denote by $\mathbf{1}_A$ the $J \times 1$ vector whose jth element is $I\{j \in A\}$. We adopt the convention that $0/0 = 0$. We describe two specific choices of the $\mathbf{\Psi}$ matrix for a fixed smoothing parameter p.

Let $A_1, A_2, \ldots, A_p$ with $A_i \neq \emptyset, i = 1, 2, \ldots, p$, be a (disjoint) partition of $\mathcal{J}$. Define $\mathbf{\Psi}_{indi}$ to be the (nonrandom) $p \times J$ matrix

$$\mathbf{\Psi}_{indi} = \left[\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \ldots, \mathbf{1}_{A_p}\right]'. \tag{6}$$

For $A \subseteq \mathcal{J}$, let $O_\bullet(A) = \mathbf{1}_A' \mathbf{O} = \sum_{j \in A} O_j$, $E_\bullet^0(A) = \mathbf{1}_A' \mathbf{E}_0 = \sum_{j \in A} E_j^0$, and $V_\bullet^0(A) = \mathbf{1}_A' \mathbf{V}_0 \mathbf{1}_A = \sum_{j \in A} V_{jj}^0 = \sum_{j \in A} E_j^0(1 - \lambda_j^0)$. Then $\mathbf{\Psi}_{indi}$ produces the statistic

$$S_p^2(\mathbf{\Psi}_{indi}) = \sum_{i=1}^p \frac{[O_\bullet(A_i) - E_\bullet^0(A_i)]^2}{V_\bullet^0(A_i)}. \tag{7}$$

Thus, through our formulation, we are able to obtain Pearson-type test statistics except for the divisors being $V_\bullet^0(A_i)$'s instead of $E_\bullet^0(A_i)$'s. Furthermore, note that the expected frequencies $E_\bullet^0(A_i)$'s are *dynamic* in the sense of being sums of the expected frequencies at each $a_j, j \in A_i$, with the expected frequency at a_j being a *conditional* expected frequency given information just prior to a_j.

Let $\mathbf{R} = (R_1, R_2, \ldots, R_J)'$ and, for $k \in \mathcal{Z}_+$, let $\mathbf{R}^k = (R_1^k, R_2^k, \ldots, R_J^k)'$. Given a fixed $p \in \mathcal{Z}_+$ with $p \leq J$, define the $p \times J$ random matrix

$$\mathbf{\Psi}_{poly}^p = \left[\left(\frac{\mathbf{R}}{n}\right)^0, \left(\frac{\mathbf{R}}{n}\right)^1, \ldots, \left(\frac{\mathbf{R}}{n}\right)^{p-1}\right]'. \tag{8}$$

This specification has intuitive appeal since the collection of vectors $\{(\frac{\mathbf{R}}{n})^0, (\frac{\mathbf{R}}{n})^1, \ldots, (\frac{\mathbf{R}}{n})^{J-1}\}$ could serve as a random basis set for J-dimensional Euclidean space, and $\mathbf{\Psi}^p_{poly}$ is a subset of this collection. The jth column of $\mathbf{\Psi}^p_{poly}$ is therefore $\mathbf{\Psi}^p_j = (1, R_j/n, (R_j/n)^2, \ldots, (R_j/n)^{p-1})'$. For $i, i_1, i_2 = 1, 2, \ldots, p$, let $U^*_i(\mathbf{\Psi}^p_{poly}) = \sum_{j=1}^J (R_j/n)^{i-1}(O_j - E^0_j)$ and $I^*_{i_1 i_2}(\mathbf{\Psi}^p_{poly}) = \sum_{j=1}^J (R_j/n)^{i_1+i_2-2} V^0_{jj}$. The test statistic generated by $\mathbf{\Psi}^p_{poly}$ is

$$S^2(\mathbf{\Psi}^p_{poly}) = \left[\left(U^*_i(\mathbf{\Psi}^p_{poly})\right)_{i=1,\ldots,p}\right]' \left[\left(I^*_{i_1 i_2}(\mathbf{\Psi}^p_{poly})\right)_{i_1,i_2=1,\ldots,p}\right]^- \left[\left(U^*_i(\mathbf{\Psi}^p_{poly})\right)_{i=1,\ldots,p}\right]. \qquad (9)$$

A special case of $S^2(\mathbf{\Psi}^p_{poly})$ is obtained by taking $p = 1$, which coincides with setting $\psi_1 = \mathbf{1}'_{\mathcal{J}}$. The resulting statistic is $S^2(\psi_1) = [\sum_{j=1}^J (O_j - E^0_j)]^2 / \sum_{j=1}^J R_j \lambda^0_j (1 - \lambda^0_j) = [O_\bullet - E^0_\bullet / \sqrt{V^0_\bullet}]^2$, where $O_\bullet = \mathbf{1}'_{\mathcal{J}} \mathbf{O} = \sum_{j=1}^J O_j$, $E^0_\bullet = \mathbf{1}'_{\mathcal{J}} \mathbf{E}_0 = \sum_{j=1}^J E^0_j$, and $V^0_\bullet = \mathbf{1}'_{\mathcal{J}} \mathbf{V}_0 \mathbf{1}_{\mathcal{J}} = \sum_{j=1}^J V^0_{jj} = \sum_{j=1}^J R_j \lambda^0_j (1 - \lambda^0_j)$. This statistic compares the total number of observed failures ($O_\bullet$) and the total conditionally expected number of failures ($E_\bullet$) over the whole study duration, and this coincides with Hyde's (1977) test statistic (cf., Peña (2002)). The class of tests for testing that the failure time distribution is geometric with mean $1/\eta_0$ is easily obtained from (7) and (9) by replacing λ^0_j's in these formulas by η_0. Based on our simulation results, we recommend the choice of $\mathbf{\Psi}$ in (8) instead of the choice in (6).

5 Simulation Results

For testing the null hypothesis that the failure times are geometrically distributed, simulation studies were performed to determine the achieved levels and powers of the tests with fixed order ($p = 1, 2, 3, 4$) and the adaptive test with $P_m = 10$ associated with the Ψ-specification given by (8). Detailed description, results, and discussions of these simulations are in Peña (2002). The table below presents a snapshot of the performance of these tests under 25% censoring proportion with $n = 100$ and $J = 100$. The hypothesized mean value under the null was 30. The level and power runs were based on 1000 replications. From this table, observe the performance of the adaptive test.

Table 1: Empirical levels and powers (in percents) of the 5% asymptotic level fixed-order and adaptive tests for testing the geometric distribution. The second column contains the achieved levels, while the other columns contain the achieved powers for different hazard alternatives.

Test Statistic	Geometric (Null)	Geometric (Different Mean)	Negative Binomial	'Polynomial' Hazards	'Trigonometric' Hazards
S^2_1	4.6	52.5	2.0	11.7	8.8
S^2_2	5.1	45.8	92.8	58.2	33.9
S^2_3	5.9	41.5	90.6	53.7	83.6
S^2_4	7.6	40.3	87.9	54.3	92.1
$S^2_{p^*}$ (Adaptive)	6.9	54.5	94.2	58.4	91.5

Acknowledgements

E. Peña wishes to acknowledge the support provided by NIH/NIGMS Grant 1 R01 GM56182 which partially supported his research and NSF Grant DMS 0102870 which supports his research, and also thanks the organizers of MMR2002 for inviting him to participate.

References

Akritas, M. (1988). Pearson-type goodness-of-fit tests: the univariate case. *J. Amer. Statist. Assoc.*, **83**, 222–230.

Baringhaus, L. and Henze, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function. *Statist. Probab. Lett.*, **13**, 269–274.

Best, D. and Rayner, J. (1989). Goodness of fit for the geometric distribution. *Biom. J.*, **31**, 307–311.

Best, D. and Rayner, J. (1999). Goodness of fit for the Poisson distribution. *Statist. Probab. Lett.*, **44**, 259–265.

Choulakian, V., Lockhart, R. and Stephens, M. (1994). Cramer-von Mises statistics for discrete distributions. *Canad. J. Statist.*, **22**, 125–137.

Cressie, N. and Read, T. (1984). Multinomial goodness-of-fit tests. *J. Royal Statist. Soc. Ser. B*, **46**, 440–464.

Eubank, R. (1997). Testing goodness of fit with multinomial data. *J. Amer. Statist. Assoc.*, **92**, 1084–1093.

Gatsonis, C., Hsieh, H. and Korwar, R. (1985). Simple nonparametric tests for a known standard survival based on censored data. *Comm. Statist. Theory Methods*, **14**, 2137–2162.

Habib, M. and Thomas, D. (1986). Chi-squared goodness-of-fit tests for randomly censored data. *Ann. Statist.*, **14**, 759–765.

Henze, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. *Canad. J. Statist.*, **24**, 81–93.

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

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. *Biom. J.*, **35**, 193–401.

Hollander, M. and Peña, E. (1992). A chi-squared goodness-of-fit test for randomly censored data. *J. Amer. Statist. Assoc.*, **87**, 458–463.

Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika*, **64**, 225–230.

Kim, J. (1993). Chi-square goodness-of-fit tests for randomly censored data. *Ann. Statist.*, **21**, 1621–1639.

Klar, B. (1999). Goodness-of-fit tests for discrete models based on the integrated distribution function. *Metrika*, **49**, 53–69.

Kocherlakota, S. and Kocherlakota, K. (1986). Goodness of fit tests for discrete distributions. *Comm. Statist. A — Theory Methods*, **15**, 815–829.

Koziol, J. and Green, S. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika*, **63**, 139-156.

Kulperger, R. and Singh, A. (1982). On random grouping in goodness of fit tests of discrete distributions. *J. Statist. Plann. Inference*, **7**, 109–115.

Li, G. and Doss, H. (1993). Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data. *Ann. Statist.*, **21**, 772–797.

Nair, V. (1981). Plots and tests for goodness of fit with randomly censored data. *Biometrika*, **68**, 99–103.

Nakamura, M. and Perez-Abreau, V. (1993). Use of an empirical probability generating function for testing a Poisson model. *Canad. J. Statist.*, **21**, 149–156.

Neyman, J. (1937). "Smooth" test for goodness of fit. *Skand. Aktuarietidskr.*, **20**, 150–199.

Peña, E. (1998). Smooth Goodness-of-Fit Tests for Composite Hypothesis in Hazard-Based Models. *Ann. Statist.*, **26**, 1935–1971.

Peña, E. (2002). Goodness-of-Fit with Right-Censored Discrete Data. Submitted for publication.

Rueda, R., Perez-Abreu, V. and O'Reilly, F. (1991). Goodness of fit for the Poisson distribution based on the probability generating function. *Comm. Statist. Theory Methods A*,**20**, 3093–3110.

Rayner, J. and Best, D. (1989). *Smooth Tests of Goodness of Fit.* Oxford University Press: New York.

Schwarz. G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461–464.

Spinelli, J. and Stephens, M. (1997). Cramer-von Mises tests of fit for the Poisson distribution. *Canad. J. Statist.*, **25**, 257–268.