Adaptive Goodness-of-Fit with Incomplete Data

Edsel A. Peña*
Department of Statistics
University of South Carolina

JSM, NY, August 14, 2002

*Research support from NIH and NSF.

The Problem

- $T_1, T_2, \ldots, T_n$ are IID rvs from an unknown discrete distribution F.
- F has support $\mathcal{A} = \{a_1, a_2, \ldots\}$ with $a_i < a_{i+1}, i = 1, 2, \ldots$.
- 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:

$$\delta_i = 1 \Rightarrow T_i = Z_i;$$

$$\delta_i = 0 \Rightarrow T_i > Z_i.$$

- Let $\lambda_j = \lambda_j(F), j = 1, 2, \ldots$ be the hazard of T at a_j, so

$$\lambda_j = \mathbf{P}(T = a_j | T \geq a_j) = \frac{\Delta F(a_j)}{\bar{F}(a_j -)}.$$

- Assumption: Independent censoring condition:

$$\begin{aligned} \mathbf{P}\{T = a_j | T \geq a_j\} &= \lambda_j \\ &= \mathbf{P}\{T = a_j | Z \geq a_j\}, j = 1, 2, \ldots. \end{aligned}$$

- General problem is to determine if $F \in \mathcal{F}_0$, a class of discrete distributions parameterized by a q-dimensional vector η taking values in $\Gamma \subseteq \Re^q$.

- Let $\mathcal{C}_0$ be the class of hazard functions associated with $\mathcal{F}_0$ so

$$\mathcal{C}_0 = \{\Lambda_0(\cdot|\eta) : \eta \in \Gamma\};$$

the functional form of $\Lambda_0(\cdot|\eta)$ being known.

- The specific composite GOF problem considered in this talk is to test the composite hypotheses

$$H_0 : \Lambda(\cdot) \in \mathcal{C}_0$$
$$H_1 : \Lambda(\cdot) \notin \mathcal{C}_0$$

on the basis of the right-censored data

$$(Z_i, \delta_i), i = 1, 2, \ldots, n.$$

- Note that in the composite GOF problem, the parameter vector η is a nuisance parameter.

Relevance and Importance

- Discrete failure times manifest in a variety of fields.

- Limitations in measurement proccess; nature of failure time (e.g., in cycles); quantum theory.

- To reminisce about D. Basu: 'Everything is discrete!'

- Right-censoring is prevalent in reliability and engineering applications, medical and public health situations, in economic settings, and in other areas.

- Desirable to know the parametric family of distributions or hazards to which F or Λ belongs.

- Such knowledge enables the use of more efficient inferential methods such as in estimating important parameters or performing group comparisons.

Hazard Embeddings and Likelihoods

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

- Following Peña (2002), 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 $\Psi_j = \Psi_j(\eta), j = 1, 2, \ldots, J$, we embed $\rho_j^0(\eta)$ into the hazard odds determined by

$$\rho_j(\theta, \eta) = \rho_j^0(\eta) \exp\{\theta^t \Psi_j(\eta)\}.$$

- This is equivalent to postulating that the logarithm of the hazard odds ratio is linear in $\Psi_j(\eta)$, that is,

$$\log\left\{\frac{\rho_j(\theta, \eta)}{\rho_j^0(\eta)}\right\} = \theta^t \Psi_j(\eta), j = 1, 2, \ldots.$$

- 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

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

$$O_j = \sum_{i=1}^{n} I\{Z_i = a_j, \delta_i = 1\};$$

$$R_j = \sum_{i=1}^{n} I\{Z_i \geq a_j\}.$$

- Furthermore, within this hazard odds embedding, the composite GOF problem simplifies to testing

$$H_0 : \theta = 0, \eta \in \Gamma \quad \text{vs.} \quad H_1 : \theta \neq 0, \eta \in \Gamma.$$

- Estimated score statistic:

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

$\hat{\hat{\eta}} = \hat{\eta}(\theta = 0)$ is the restricted partial likelihood MLE (RPLMLE).

Restricted Partial Likelihood MLE

- $\hat{\hat{\eta}}$ is the η that maximizes the restricted partial likelihood function

$$L_0(\eta) = \prod_{j=1}^{J} [\lambda_j^0(\eta)]^{O_j}[1-\lambda_j^0(\eta)]^{R_j-O_j}$$

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

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

"dynamic expected freq." "standardized gradients"

- Form the $J \times q$ matrix of standardized gradients

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

  and the $J \times 1$ vectors

$$\begin{aligned}\mathbf{O} &= (O_1, O_2, \ldots, O_J)^{\mathsf{t}};\\ \mathbf{E}^0(\eta) &= \Big(E_1^0(\eta), E_2^0(\eta), \ldots, E_J^0(\eta)\Big)^{\mathsf{t}}.\end{aligned}$$

- Matrix form: $\nabla_\eta l_0(\eta) = \mathbf{A}(\eta)^{\mathsf{t}} \left[\mathbf{O} - \mathbf{E}^0(\eta)\right]$

- Estimating equation for the RPLMLE $\widehat{\widehat{\eta}}$:

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

Asymptotics and Test

- With $\Psi(\eta) = [\Psi_1(\eta), \Psi_2(\eta), \ldots, \Psi_J(\eta)]^t$, the score function for θ at $\theta = \mathbf{0}$ is

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

- Estimated Score Function:

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

- Needed: Asymptotic distribution of $\hat{\mathrm{U}}_\theta$.

- Entails obtaining the asymptotic distribution of the $(p+q) \times 1$ vector of scores:

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

- Needed notations:

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

$$\lambda(\eta) = (\lambda_1(\eta), \lambda_2(\eta), \ldots, \lambda_J(\eta))^{\mathsf{t}}$$

$$\mathbf{A}(\eta) = \mathbf{D}(\eta)^{-1} \nabla_{\eta^{\mathsf{t}}} \lambda(\eta)$$

$$\mathbf{V}(\eta) = \mathsf{Dg}(\mathbf{R})\mathbf{D}(\eta); \quad \mathbf{B}(\eta) = [\mathbf{\Psi}(\eta), \mathbf{A}(\eta)]$$

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

- **Proposition 1** *Spse* H_0 *holds with* $\eta = \eta_0$ *and* p *does not change with* n. *Under regu, larity conditions, in particular if, as* $n \to \infty$, $\exists (p+q) \times (p+q)$ *pos def matrix* $\Xi^{(0)}(\eta_0)$ *with*

$$\frac{1}{n}\Xi(\eta_0) \xrightarrow{\text{pr}} \Xi^{(0)}(\eta_0),$$

then

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

- **Corollary 1**

$$\frac{1}{\sqrt{n}}\mathbf{\Psi}(\eta_0)^{\mathsf{t}}[\mathrm{O} - \mathrm{E}^0(\eta_0)] \xrightarrow{\mathsf{d}} N_p(\mathbf{0}, \Xi_{11}^{(0)}(\eta_0));$$

$$\frac{1}{n}\mathbf{\Psi}(\eta_0)^{\mathsf{t}}\mathrm{V}(\eta_0)\mathbf{\Psi}(\eta_0) \xrightarrow{\mathsf{pr}} \Xi_{11}^{(0)}(\eta_0).$$

- Result not directly useful since η_0 is unknown. This however leads to the desired asymptotic result.

- **Theorem 1** *Under H_0 and regularity con, ditions,*

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

where

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

- Effect of estimating the unknown parameter η_0 by the RPLMLE $\hat{\hat{\eta}}$ is to decrease the covariance matrix by the term

$$\Xi_{12}^{(0)}(\eta_0)\left\{\Xi_{22}^{0)}(\eta_0)\right\}^{-1}\Xi_{21}^{(0)}(\eta_0).$$

- Substituting the estimator $\hat{\hat{\eta}}$ for η_0 does not have an effect on the limiting variance provided $\Xi_{12}^{(0)}(\eta_0) = 0$, which is an orthogonality condition between $\boldsymbol{\Psi}$ and $\mathbf{A}(\eta_0)$.

- **Test Statistic:**

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

- **Test Procedure:** An asymptotic α-level test rejects H_0 whenever

$$\widehat{S}_p^2 > \chi^2_{\widehat{p}^*;\alpha}$$

with $\widehat{p}^* = \mathrm{r}(\widehat{\Xi}_{11.2}^{(0)})$.

Two Choices of Ψ

- $A_1, A_2, \ldots, A_p$ a partition of $\{a_1, a_2, \ldots, a_J\}$. Define

$$\Psi_1 = \left[1_{A_1}, 1_{A_1}, \ldots, 1_{A_p}\right]'$$

where $1_A = (I\{a_j \in A\}, j = 1, 2, \ldots, J)'$.

- This choice leads to a generalization of Pearson's chi-square test. The test statistic for the simple null case is:

$$S_p^2(\Psi_1) = \sum_{i=1}^{p} \frac{[O_\bullet(A_i) - E_\bullet^0(A_i)]^2}{V_\bullet^0(A_i)}.$$

- Another choice, which has proven effective in the simple null case, is provided by

$$\Psi_2 = \left(\left(\frac{\mathbf{R}}{n}\right)^0, \left(\frac{\mathbf{R}}{n}\right)^1, \ldots, \left(\frac{\mathbf{R}}{n}\right)^{p-1} \right)'.$$

- When $p = 1$ and in the simple null setting, the test statistic is

$$S^2(\psi_1) = \frac{[\sum_{j=1}^{J}(O_j - E_j^0)]^2}{\sum_{j=1}^{J} R_j \lambda_j^0 (1 - \lambda_j^0)}.$$

This coincides with Hyde's ('77, Bmka) statistic.

Adaptive Choice of Smoothing Order

- Test requires that the smoothing order p be fixed.
  - Arbitrary.
  - Potential of choosing a p that is far from optimal.
- Repeated testing with different smoothing orders? Unwise since Type I error rates will become inflated.
- Imperative and Important! A data-driven or adaptive approach for determining p.

- **Proposal:** Use a modified Schwarz information criterion. Modified to accommodate right-censoring.

- For a given p:

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

- Modified Schwarz information criterion:

$$\mathsf{MSIC}(p) = \log L_p(\widehat{\theta}_p, \widehat{\eta}) - \frac{p}{2}\left[\log(n) + \log(\hat{\lambda}_{\mathsf{max}})\right]$$

with $\hat{\lambda}_{\mathsf{max}}$ being the largest eigenvalue of $\mathbf{I}_p(\widehat{\theta}_p, \widehat{\eta})$.

- Adaptively-chosen smoothing order:

$$p^* = \arg \max_{1 \le p \le P_{\max}} \{\mathrm{MSIC}(p)\},$$

$P_{\max}$ a pre-specified maximum order e.g., 10.

- **Adaptive Test Procedure:** Rejects H_0 whenever

$$S^2_{p^*} \ge \chi^2_{\hat{k}^*;\alpha},$$

where $\hat{k}^* = \mathrm{r}(\widehat{\Xi}^{(0)}_{11.2})$.

Simulation Results: Simple Null

- Simple Null Hypothesis: failure times are geometrically distributed.

- Simulation studies to determine achieved levels and powers of the tests with fixed order ($p = 1, 2, 3, 4$) and the adaptive test with $P_{\text{max}} = 10$ associated with Ψ_2.

- Table: presents performance of tests under 25% censoring for $n = 100$ and $J = 100$. Hypothesized null mean was 30. Based on 1000 replications.

- Empirical levels and powers (in percents) of the 5% asymptotic level fixed-order and adaptive tests for testing the geometric distribution.

Test Stat	Geo. (Null)	Geo. (Alt)	Neg. Bin.	'Poly' Haz	'Trig' Haz
S_1^2	4.6	52.5	2.0	11.7	8.8
S_2^2	5.1	45.8	92.8	58.2	33.9
S_3^2	5.9	41.5	90.6	53.7	83.6
S_4^2	7.6	40.3	87.9	54.3	92.1
$S_{p^*}^2$	6.9	54.5	94.2	58.4	91.5

Concluding Remarks

- A general approach to construct GOF tests in the presence of right-censored discrete data.
- Approach can be described as ‘functional’ in nature.
- Adaptive approach utilizing Schwarz Bayesian information criterion for determining the smoothing order.
- Simulation studies for simple null case indicates that the adaptive test serve as an omnibus test.