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 $A = \{a_1, a_2, \ldots\}$ 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

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

• Assumption: Independent censoring condition:

$$\mathbb{P}\{T = a_j | T \geq a_j\} = \lambda_j$$

$$= \mathbb{P}\{T = a_j | Z \geq a_j\}, j = 1, 2, \ldots.$$
• General problem is to determine if $F \in \mathcal{F}_0$, a class of discrete distributions parameterized by a $q$-dimensional vector $\eta$ taking values in $\Gamma \subseteq \mathbb{R}^q$.

• 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 considers composite hypotheses

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

right-censored data \((Z_i, \delta_i), i = 1, 2, \ldots, n.\)

• Note that in the composite GOF problem, the parameter vector \(\eta\) is a nuisance parameter.
Relevance and Importance

- Discrete failure times manifest in a variety of fields.

- Limitations in measurement process; 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 $\Lambda$ 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 $\lambda < 1$ 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 $p \in \mathbb{Z}_+$, 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 \((\theta, \eta)\) based on the observation period \((-\infty, a_J]\) for some fixed \(J \in \mathbb{Z}_+\) 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

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

• Estimated score statistic:

\[ U_\theta(0, \tilde{\eta}) = \nabla_\theta \log L(\theta, \eta)|_{\theta = 0, \eta = \tilde{\eta}}; \]

\[ \tilde{\eta} = \hat{\eta}(\theta = 0) \text{ is the restricted partial likelihood MLE (RPLMLE).} \]
Restricted Partial Likelihood MLE

- \( \hat{\eta} \) is the \( \eta \) that maximizes the restricted partial likelihood function

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

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

\[
E_j^0(\eta) = R_j \lambda_j^0(\eta); \quad A_j(\eta) = \frac{\nabla_\eta \lambda_j^0(\eta)}{\lambda_j^0(\eta)[1 - \lambda_j^0(\eta)]}
\]

"dynamic expected freq." "standardized gradients"
• Form the $J \times q$ matrix of standardized gradients

$$A(\eta) = [A_1(\eta), A_2(\eta), \ldots, A_J(\eta)]^t,$$

and the $J \times 1$ vectors

$$O = (O_1, O_2, \ldots, O_J)^t;$$
$$E^0(\eta) = (E^0_1(\eta), E^0_2(\eta), \ldots, E^0_J(\eta))^t.$$

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

• Estimating equation for the RPLMLE $\hat{\eta}$:

$$A(\eta)^t [O - E^0(\eta)] = 0.$$
Asymptotics and Test

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

$$U_\theta(\theta = 0, \eta) = \Psi(\eta)^t[O - E^0(\eta)].$$

- Estimated Score Function:

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

- Needed: Asymptotic distribution of $\hat{U}_\theta$. 
• Entails obtaining the asymptotic distribution of the $(p + q) \times 1$ vector of scores:

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

• Needed notations:

\[
D(\eta) = Dg \left( \lambda_j(\eta)[1 - \lambda_j(\eta)] : j = 1, \ldots, J \right)
\]

\[
\lambda(\eta) = (\lambda_1(\eta), \lambda_2(\eta), \ldots, \lambda_J(\eta))^t 
\]

\[
A(\eta) = D(\eta)^{-1} \nabla_{\eta^t} \lambda(\eta) 
\]

\[
V(\eta) = Dg(R)D(\eta) ; \quad B(\eta) = [\Psi(\eta), A(\eta)] 
\]

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

then

$$
\frac{1}{\sqrt{n}} U(\eta_0) = \frac{1}{\sqrt{n}} B(\eta_0)^\top [O - E^0(\eta_0)]
$$

$\xrightarrow{d}$

$$
N_{p+q}(0, \Xi^{(0)}(\eta_0)).
$$
• Corollary 1

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

\[ \frac{1}{n} \Psi(\eta_0)^t V(\eta_0) \Psi(\eta_0) \xrightarrow{pr} \Xi^{(0)}_{11}(\eta_0). \]

• Result not directly useful since \( \eta_0 \) is unknown. This however leads to the desired asymptotic result.
• **Theorem 1** Under $H_0$ and regularity conditions,

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

where

\[
\Xi^{(0)}_{11.2}(\eta_0) = \Xi^{(0)}_{11}(\eta_0) - \Xi^{(0)}_{12}(\eta_0) \left\{ \Xi^{(0)}_{22}(\eta_0) \right\}^{-1} \Xi^{(0)}_{21}(\eta_0).
\]
• Effect of estimating the unknown parameter $\eta_0$ by the RPLMLE $\hat{\eta}$ 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 $\hat{\eta}$ for $\eta_0$ does not have an effect on the limiting variance provided $\Xi_{12}^{(0)}(\eta_0) = 0$, which is an orthogonality condition between $\Psi$ and $A(\eta_0)$. 
• Test Statistic:

\[
\hat{s}_p^2 = \left\{ \frac{1}{\sqrt{n}} \Psi(\hat{\eta})^t [O - E^0(\hat{\eta})] \right\}^t \left\{ \hat{\Sigma}_{11.2}^{(0)} \right\}^- \times \left\{ \frac{1}{\sqrt{n}} \Psi(\hat{\eta})^t [O - E^0(\hat{\eta})] \right\}.
\]

• Test Procedure: An asymptotic \(\alpha\)-level test rejects \(H_0\) whenever

\[
\hat{s}_p^2 > \chi^2_{\tilde{p}^*};\alpha
\]

with \(\tilde{p}^* = r(\hat{\Sigma}_{11.2}^{(0)})\).
Two Choices of $\Psi$

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

$$\Psi_1 = \begin{bmatrix} 1_{A_1}, 1_{A_1}, \ldots, 1_{A_p} \end{bmatrix}'$$

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

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

• When \( p = 1 \) and in the simple setting, the test statistic is

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

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 $p$.  

23
• **Proposal:** Use a modified Schwarz information criterion. Modified to accommodate right-censoring.

• For a given $p$:

\[
L_p(\hat{\theta}_p, \hat{\eta}) = \sup_{\theta_p \in \mathcal{R}^p; \eta \in \Gamma} L_p(\theta_p, \eta).
\]

• Modified Schwarz information criterion:

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

with $\hat{\lambda}_{\text{max}}$ being the largest eigenvalue of $I_p(\hat{\theta}_p, \hat{\eta})$. 
\begin{itemize}
\item Adaptively-chosen smoothing order:
\[ p^* = \arg \max_{1 \leq p \leq P_{\text{max}}} \{ \text{MSIC}(p) \}, \]
\[ P_{\text{max}} \text{ a pre-specified maximum order e.g., 10.} \]

\item **Adaptive Test Procedure:** Rejects $H_0$ whenever
\[ S_{p^*}^2 \geq \chi_{\hat{k}^*; \alpha}^2, \]
where $\hat{k}^* = r(\hat{\Omega}^{(0)}_{11.2})$. 
\end{itemize}
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 \( \Psi_2 \).

- 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.**

<table>
<thead>
<tr>
<th>Test Stat</th>
<th>Geo. (Null)</th>
<th>Geo. (Alt)</th>
<th>Neg. Bin.</th>
<th>'Poly' Haz</th>
<th>'Trig' Haz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^2_1$</td>
<td>4.6</td>
<td>52.5</td>
<td>2.0</td>
<td>11.7</td>
<td>8.8</td>
</tr>
<tr>
<td>$S^2_2$</td>
<td>5.1</td>
<td>45.8</td>
<td>92.8</td>
<td>58.2</td>
<td>33.9</td>
</tr>
<tr>
<td>$S^2_3$</td>
<td>5.9</td>
<td>41.5</td>
<td>90.6</td>
<td>53.7</td>
<td>83.6</td>
</tr>
<tr>
<td>$S^2_4$</td>
<td>7.6</td>
<td>40.3</td>
<td>87.9</td>
<td>54.3</td>
<td>92.1</td>
</tr>
<tr>
<td>$S^2_{p^*}$</td>
<td>6.9</td>
<td>54.5</td>
<td>94.2</td>
<td>58.4</td>
<td>91.5</td>
</tr>
</tbody>
</table>
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.