Introduction

• Joined the department in 2001.
• Teach two courses:
  – *Elements of Stochastic Processes* (Biostat 2040).
  – *Analysis of Incomplete Data* (Biostat 2065).
• Some research interests:
  – *Statistical Analysis of Missing Data*.
  – *Analysis of Correlated Outcomes*.
  – *Semi-Parametric Statistics*.
• Statistician of the Biostatistical Center of NSABP
  ([www.nsabp.pitt.edu](http://www.nsabp.pitt.edu)).
  – Work on cancer clinical trials.
  – Recently involved in studies to investigate the association between genomic data and clinical outcomes (with Drs. John Bryant and Joe Costantino).
An Interesting Phenomenon

- Pseudolikelihood method
  - Likelihood function of \( \theta : L(\theta; \alpha) = L(\theta | \alpha) \).
  - \( \theta \): parameter of interest; \( \alpha \): nuisance parameter.
  - \( \hat{\alpha} \) is a consistent estimator of \( \alpha \); \( \alpha_0 \) is the true value.
  - Consider \( \hat{\theta} = \arg \max_{\theta} L(\theta; \alpha_0) \) and
    \( \tilde{\theta} = \arg \max_{\theta} L(\theta; \hat{\alpha}) \) \(<\text{pseudolikelihood estimate}>\).

- In some circumstances, \( \text{var}(\hat{\theta}) < \text{var}(\tilde{\theta}) \).
- Demonstrate through a missing-data problem.
- Question: how to improve the efficiency when \( \alpha_0 \) is actually known, from other source?
Bivariate data with outcome-dependent nonresponse

- Model assumptions:
  
  \( a \) \( X \sim f(x; \alpha), \ [Y | X] \sim g(y | x, \theta) \).

  \( b \) \( \Pr[ R = 1 | X, Y ] = w(y; \psi) \).

- A conditional likelihood method

\[
L(\theta; F) = \prod_{i=1}^{m} p(x_i | y_i; \theta, F) \quad (\text{for } x_i \perp R_i, \text{ given } y_i)
\]

\[
\propto \prod_{i=1}^{m} \int g(y_i | x_i, \theta) dF(x) \quad m = \# \text{ of c.c.}
\]

where \( F(x) = F(x; \alpha_0) \) represents the true CDF of \( X \).

- \( \hat{\theta} = \arg \max_{\theta} L(\theta; F) \)

- \( \tilde{\theta} = \arg \max_{\theta} L(\theta; F(x; \hat{\alpha})), \) where \( \hat{\alpha} = \arg \max_{\alpha} \prod_{i=1}^{n} f(x; \alpha) \).
Full likelihood for 2-pattern data

- Model assumptions:
  
  \( X \sim f(x \mid \alpha), \quad [Y \mid X] \sim g(y \mid x, \theta). \)

  \( (b) \quad \Pr[ R = 1 \mid X, Y ] = w(y \mid \psi). \)

- Full likelihood:

\[
L_{\text{Full}}(\alpha, \theta, \psi) = \prod_{i=1}^{m} f(x_i \mid \alpha) g(y_i \mid x_i, \theta) w(y_i \mid \psi)
\]

\[
\cdot \prod_{i=m+1}^{n} f(x_i \mid \alpha) \int g(y \mid x_i, \theta) \{1 - w(y \mid \psi)\} dy
\]

\[
= \{\prod_{i=1}^{n} f(x_i \mid \alpha) \} \cdot \{\prod_{i=1}^{m} g(y_i \mid x_i, \theta) w(y_i \mid \psi)\} \prod_{i=m+1}^{n} \int g(y \mid x_i, \theta) \{1 - w(y \mid \psi)\} dy
\]

\[
= L(\alpha) L(\theta, \psi)
\]
Asymptotic Variance of $\hat{\theta}$

- Let $l(\theta) = \log L(\theta; \alpha)$, from $\hat{\theta} = \arg\max_{\theta} L(\theta; \alpha_0)$,

\[0 = l_\theta(\hat{\theta}; \alpha_0) \approx l_\theta(\theta_0; \alpha_0) + l_{\theta\theta}(\theta_0; \alpha_0)(\hat{\theta} - \theta_0)\]

then

\[\sqrt{n}(\hat{\theta} - \theta_0) \approx -\{l_{\theta\theta}(\theta_0; \alpha_0)\}^{-1} \sqrt{n} \frac{l_\theta(\theta_0; \alpha_0)}{n}\]

\[\approx -E(l_{\theta\theta,1})^{-1} \sqrt{n} \frac{1}{n} \sum_{i=1}^n l_{\theta,i}(\theta_0; \alpha_0).\]

Therefore,

\[\text{Var}(\sqrt{n}(\hat{\theta} - \theta_0)) \approx E(l_{\theta\theta,1})^{-1}\text{Var}(l_{\theta,1}(\theta_0; \alpha_0))E(l_{\theta\theta,1})^{-1}\]

\[= E(l_{\theta\theta,1})^{-1} E(l_{\theta,1}l_{\theta,1}^T) E(l_{\theta\theta,1})^{-1},\]

(A sandwich-type estimator).
Asymptotic Variance of $\tilde{\theta}$ (I)

- Let $S(\alpha) = \log f(x; \alpha)$, $\hat{\alpha} = \arg \max_{\alpha} \prod_{i=1}^{n} f(x_i; \alpha)$ is MLE of $\alpha$,

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \approx -E(S_{\alpha\alpha})^{-1} \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} S_{\alpha,i} \to N(0,-E(S_{\alpha\alpha})^{-1}).$$

- From $\tilde{\theta} = \arg \max_{\theta} L(\theta; \hat{\alpha}) = \arg \max_{\theta} l(\theta; \hat{\alpha})$,

$$0 = l_\theta(\tilde{\theta}; \hat{\alpha}) \approx l_\theta(\theta_0; \hat{\alpha}) + l_{\theta\theta}(\theta_0; \hat{\alpha})(\tilde{\theta} - \theta_0),$$

then

$$\sqrt{n}(\tilde{\theta} - \theta_0) \approx -\left\{ \frac{l_{\theta\theta}(\theta_0; \hat{\alpha})}{n} \right\}^{-1} \sqrt{n} \frac{l_\theta(\theta_0; \hat{\alpha})}{n}$$

$$\approx -E(l_{\theta\theta,1})^{-1} \sqrt{n} \frac{l_\theta(\theta_0; \alpha_0) + l_{\theta\alpha}(\hat{\alpha} - \alpha_0)}{n}$$

$$\approx -E(l_{\theta\theta,1})^{-1} \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} \{l_{\theta,i} - E(l_{\theta\alpha,1})E(S_{\alpha\alpha})^{-1}S_{\alpha,i}\}. $$
Asymptotic Variance of $\hat{\theta}$ (II)

- From $\sqrt{n}(\tilde{\theta} - \theta_0) \approx -E(l_{\theta,1})^{-1}\sqrt{n}\sum_{i=1}^{n}\{l_{\theta,i} - E(l_{\theta,1})E(S_{\alpha\alpha})^{-1}S_{\alpha,i}\}$,

$$Var(\sqrt{n}(\tilde{\theta} - \theta_0)) \approx E(l_{\theta,1})^{-1}\{E(l_{\theta,1}l_{\theta,1}^T) - E(l_{\theta,1}S_{\alpha,1}^T)E(S_{\alpha\alpha})^{-1}E(l_{\theta,1})^T - E(l_{\theta,1})E(S_{\alpha\alpha})^{-1}E(S_{\alpha,1}l_{\theta,1}^T) - E(l_{\theta,1}S_{\alpha,1}^T)E(S_{\alpha\alpha})^{-1}E(l_{\theta,1})^T\}E(l_{\theta,1})^{-1}.$$

- On the other hand,

$$0 = E(l_{\theta,1}) = \int l_{\theta,1}(\alpha, \theta)f(x_1; \alpha)g(y_1 \mid x_1; \theta)p(r_1 \mid x_1, y_1; \psi)dx_1dy_1dr_1$$

$$= \int l_{\theta,1}(\alpha, \theta)L(\alpha)L(\theta, \psi) \ d \mu$$

So, $0 = \frac{\partial}{\partial \alpha} E(l_{\theta,1})^T = E(l_{\alpha,\theta,1}) + \int S_{\alpha,1} l_{\theta,1}^T L(\alpha)L(\theta, \psi) \ d \mu$

$$= E(l_{\alpha,\theta,1}) + E(S_{\alpha,1}l_{\theta,1}^T) = E(l_{\theta,1})^T + E(S_{\alpha,1}l_{\theta,1}^T).$$

- Therefore,

$$Var(\sqrt{n}(\tilde{\theta} - \theta_0)) \approx E(l_{\theta,1})^{-1}\{E(l_{\theta,1}l_{\theta,1}^T) - E(l_{\theta,1})E(-S_{\alpha\alpha})^{-1}E(l_{\theta,1})^T\}E(l_{\theta,1})^{-1}$$
Another Pseudolikelihood Estimator

• When the functional form of $F(x)$ is unknown, let
  \[ \tilde{\theta} = \arg \max_\theta L(\theta; F_n) \]
  \[ = \arg \max_\theta \prod_{i=1}^m \int g(y_i | x_i, \theta) dF_n(x) \]
  \[ = \arg \max_\theta \prod_{i=1}^m \frac{g(y_i | x_i, \theta)}{\frac{1}{n} \sum_{j=1}^n g(y_i | x_j, \theta)} \] (PL2)

where, $F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$ is the empirical distribution of $X$.

• Simulation studies suggest that $\tilde{\theta}$ is even more efficient even though it has no assumption on $F(x)$.  

\[ X \quad Y \]

\[ R = 1 \]
\[ R = 0 \]
Auxiliary Information

- If $F(x) = F(x; \alpha_0)$ is known, for example, in some survey studies, can we get more efficient estimator?
- Answer: yes, with empirical likelihood.
Empirical Likelihood

• "Empirical likelihood is a nonparametric method of inference based on a data-driven likelihood ratio function" (Art Owen).

• Suppose \( \{x_i\}_{i=1}^n \) is a random sample of \( X \sim F(x) \), then the non-parametric estimator for \( F(x) \) is \( F_n(x) = \frac{1}{n} \sum_{i=1}^n I\{x_i \leq x\} \),

or \( \prod_{i=1}^n p_i \) is maximized subject to \( \sum_{i=1}^n p_i = 1 \), where \( p_i = pr\{X = x_i\} \).

• With auxiliary information \( E(w(x)) = 0 \) available, the empirical likelihood estimator maximizes \( \prod_{i=1}^n p_i \) subject to constraints

\[
\sum_{i=1}^n p_i = 1 \text{ and } \sum_{i=1}^n p_i w(x_i) = 1.
\]
Incorporate Auxiliary Information

- For example, $E(X) = \mu_0$ is known.
- The PL2 estimator solves the estimating equations
  \[
  0 = l_\theta(\theta; F_n) = \sum_{i=1}^{n} l_{\theta,i}(\theta; F_n).
  \]
- New estimator:
  (i) Maximizes $\prod_{i=1}^{n} p_i$ subject to constraints
      \[
      \sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i (x_i - \mu_0) = 1.
      \]
      Get $\{\hat{p}_i\}_{i=1}^{n}$.
  (ii) Let $\bar{\theta}$ solves $0 = \sum_{i=1}^{n} \hat{p}_i l_{\theta,i}(\theta; F_n)$, then
      \[
      Var(\bar{\theta}) = (El_{\theta \theta})^{-1} \{Var(l_\theta(\theta; F_n)) - E(l_\theta x)Var(x)^{-1} E(xl_\theta)\}(El_{\theta \theta})^{-1}.
      \]
A simulation study

- Complete data:
  
  (1) \([x] \sim N(0,1),\)
  
  (2) \([y \mid x] \sim N(1 + x, 1).\)

- Missing-data mechanism:

\[
pr[R = 1 \mid x, y] = \Phi(y - 1).
\]

- Compare the performance of the PL and EL estimates for the regression model (2), where

\[
\theta = (\beta_0, \beta_1, \sigma^2) = (1, 1, 1).
\]

Note: sample size \(n = 300,\) average \# of c.c. \(= 150.\)
## Simulation results

**Table:** empirical biases and standard deviations of two estimators over 1000 replicates

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>0.012</td>
<td>0.019</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
<td>(0.115)</td>
<td>(0.197)</td>
</tr>
<tr>
<td>EL</td>
<td>0.012</td>
<td>0.017</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.11)</td>
<td>(0.186)</td>
</tr>
</tbody>
</table>
Reference


• An unpublished manuscript.