A Broad Framework for Joint Modeling And Some Tales From the Unexpected

Geert Molenberghs

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

Universiteit Hasselt & Katholieke Universiteit Leuven, Belgium

geert.molenberghs@uhasselt.be & geert.molenberghs@med.kuleuven.be

www.ibiostat.be

Raleigh, North Carolina State University

with: M. Aerts, M. Davidian, M.G. Kenward, D. Rizopoulos, A.A. Tsiatis, G. Verbeke

Common Ground for Various Settings

Sequential trials	outcome & sample size
Incomplete data	outcome(s) & missigness process
Completely random sample size	outcome & sample size
'Informative' cluster size	outcome & cluster size
Classical survival	time to event & censorship
(Narrow) joint modeling	time to event(s) & longitudinal process
Random observation times	outcome & measurement schedule

A Classic: Joint Models with Missing Data

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})}$

MCAR $\longrightarrow$ MAR $\longrightarrow$ MNAR

$$f(\boldsymbol{r}_i | X_i, \boldsymbol{\psi}) \quad f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{\psi}) \quad f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})$$

Pattern-mixture Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{r}_i, \boldsymbol{\theta}) \, f(\boldsymbol{r}_i | X_i, \boldsymbol{\psi})$

Shared-parameter Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{b}_i, \boldsymbol{\theta}) \, f(\boldsymbol{r}_i | X_i, \boldsymbol{b}_i, \boldsymbol{\psi})$

Generic Setting

$$f(\boldsymbol{y}_i, \boldsymbol{c}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) \quad = \quad f(\boldsymbol{y}_i|\boldsymbol{\theta}) \cdot f(\boldsymbol{c}_i|\boldsymbol{y}_i, \boldsymbol{\psi}) \quad = \quad f(\boldsymbol{y}_i|\boldsymbol{c}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) \cdot f(\boldsymbol{c}_i|\boldsymbol{\theta}, \boldsymbol{\psi})$$

Setting	$\boldsymbol{Y}_i$	$\boldsymbol{C}_i$
Sequential trials	Y_i	N
Incomplete data	$\boldsymbol{Y}_i$	$\boldsymbol{R}_i$
Completely random sample size	Y_i	N
'Informative' cluster size	$\boldsymbol{Y}_i$	T_i
Classical survival	T_i	C_i
(Narrow) joint modeling	$\boldsymbol{Y}_i$	$T_i, \ldots$
Random observation times	$\boldsymbol{Y}_i$	$\boldsymbol{T}_i$

Fundamental Concepts

Ignorability:

$$f(\boldsymbol{c}_i|\boldsymbol{y}_i,\boldsymbol{\psi}) = f(\boldsymbol{c}_i|\boldsymbol{y}_i^o,\boldsymbol{\psi})$$

combined with separability

Ancillarity:

▷ S minimally sufficient statistic for θ

▷ $T|S$ does not contain information about θ

Completeness:

- ▷ $g(\cdot)$ a measurable function of S
- ▷ $E[g(s)] = 0$ for all θ $\Longrightarrow$ $g(s) = 0$ a.e.

Weights:

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{\theta}^*, \boldsymbol{\psi}^*) &= f(\boldsymbol{y}_i|\boldsymbol{c}_i, \boldsymbol{\theta}^*) \cdot \frac{f(\boldsymbol{c}_i|\boldsymbol{\psi}^*)}{f(\boldsymbol{c}_i|\boldsymbol{y}_i, \boldsymbol{\theta}^*, \boldsymbol{\psi}^*)} \\ &= f(\boldsymbol{y}_i|\boldsymbol{c}_i, \boldsymbol{\theta}^*) \cdot w_i(\boldsymbol{y}_i, \boldsymbol{c}_i, \boldsymbol{\theta}^*, \boldsymbol{\psi}^*) \end{aligned}$$

MAR: $w_i(\boldsymbol{y}_i, \boldsymbol{c}_i) = w_i(\boldsymbol{y}_i^o, \boldsymbol{c}_i)$

Deterministic rule:

$$w_i(\boldsymbol{y}_i^o, \boldsymbol{c}_i) \in \{0, 1\}$$

Simple Setting

- Observe

$$Y_i \sim N(\mu, 1) \qquad i = 1, \ldots, n$$

- Construct sum statistics

$$K = \sum_{i=1}^{n} Y_i$$

- Apply a stopping rule, e.g.,

$$F\left(\alpha + \frac{\beta}{n} k\right)$$

- If continuation applies, observe $Y_i, \qquad i = n + 1, \ldots, 2n$

- Data:

$$(Y_1, \ldots, Y_N, N)$$

Stopping Rule

$$F\left(\alpha + \frac{\beta}{n}k\right) = \Phi\left(\alpha + \frac{\beta}{n}k\right)$$

- Three important cases:

Purely random sample size	$\beta = 0$
Probabilistic stopping	$\beta \neq 0$
Purely random sample size	$\beta \to +\infty$

Four Relevant Distributions

$\boldsymbol{Y}$	$\prod_{i=1}^{N} \phi(y_i;\mu)$
$N\|\boldsymbol{Y}$	$\Phi\left(\alpha+\frac{\beta}{n}k\right)$
$\boldsymbol{Y}\|N$	$\dfrac{\prod_{i=1}^{N} \phi(y_i;\mu)\Phi\left(\alpha+\frac{\beta}{n}k\right)^{z}\left[1-\Phi\left(\alpha+\frac{\beta}{n}k\right)\right]^{1-z}}{\Phi\left(\frac{\alpha+\beta\mu}{\sqrt{1+\beta^2/n}}\right)^{z}\left[1-\Phi\left(\frac{\alpha+\beta\mu}{\sqrt{1+\beta^2/n}}\right)\right]^{1-z}}$
N	$\Phi\left(\frac{\alpha+\beta\mu}{\sqrt{1+\beta^2/n}}\right)$

Joint Likelihood

$$L(\mu) = \prod_{i=1}^{N} \phi(y_i; \mu) \cdot \Phi\left(\alpha + \frac{\beta}{n} k\right)^{z} \cdot \left\{1 - \Phi\left(\alpha + \frac{\beta}{n} k\right)\right\}^{1-z}$$

$$\ell(\mu) = -\frac{1}{2} \sum_{i=1}^{N} (y_i - \mu)^2$$

$$S(\mu) = \sum_{i=1}^{N} (y_i - \mu)$$

$$H(\mu) = -N$$

Conditional Likelihood

$$L_c(\mu) = \frac{\Pi_{i=1}^{N} \phi(y_i;\mu)\Phi\left(\alpha + \frac{\beta}{n}k\right)^z \left[1 - \Phi\left(\alpha + \frac{\beta}{n}k\right)\right]^{1-z}}{\Phi\left(\frac{\alpha+\beta\mu}{\sqrt{1+\beta^2/n}}\right)^z \left[1 - \Phi\left(\frac{\alpha+\beta\mu}{\sqrt{1+\beta^2/n}}\right)\right]^{1-z}}$$

$$N = n \quad : \quad \begin{cases} \ell(\mu) = -\frac{1}{2}\Sigma_{i=1}^{N}(y_i-\mu)^2 - \ln\Phi(\nu) \\ S(\mu) = \Sigma_{i=1}^{N}(y_i-\mu) - \widetilde{\beta}\cdot\frac{\phi(\nu)}{\Phi(\nu)} \\ H(\mu) = -N + \widetilde{\beta}^2\cdot[\nu\cdot\Phi(\nu) + \phi(\nu)]\cdot\frac{\phi(\nu)}{\Phi(\nu)^2} \end{cases}$$

$$N = 2n \quad : \quad \begin{cases} \ell(\mu) = -\frac{1}{2}\Sigma_{i=1}^{N}(y_i-\mu)^2 - \ln[1-\Phi(\nu)] \\ S(\mu) = \Sigma_{i=1}^{N}(y_i-\mu) + \widetilde{\beta}\cdot\frac{\phi(\nu)}{1-\Phi(\nu)} \\ H(\mu) = -N - \widetilde{\beta}^2\cdot\{\nu\cdot[1-\Phi(\nu)] - \phi(\nu)\}\cdot\frac{\phi(\nu)}{[1-\Phi(\nu)]^2} \end{cases}$$

Information, Bias, MSE

- Notation:

$$\widetilde{\alpha} = \alpha/\sqrt{1+\beta^2/n}$$

$$\widetilde{\beta} = \beta/\sqrt{1+\beta^2/n}$$

$$\nu = (\alpha+\beta\mu)/\sqrt{1+\beta^2/n}$$

- **Information:**

$$I(\mu) = n[2-\Phi(\widetilde{\alpha}+\widetilde{\beta}\mu)]$$

$$I_c(\mu) = n[2-\Phi(\widetilde{\alpha}+\widetilde{\beta}\mu)] - \frac{\widetilde{\beta}^2\phi(\widetilde{\alpha}+\widetilde{\beta}\mu)^2}{\Phi(\widetilde{\alpha}+\widetilde{\beta}\mu)[1-\Phi(\widetilde{\alpha}+\widetilde{\beta}\mu)]}$$

- **Counterintuitive?** N is not ancillary

- Some limits:

$$\beta = 0 \quad \Longrightarrow \quad I(\mu) \equiv I_c(\mu)$$

$$n \longrightarrow \infty \quad \Longrightarrow \quad I(\mu) = I_c(\mu)$$

- Both estimators asymptotically unbiased

- The conditional score $E[S_c(\mu)] = 0$:

  conditional likelihood estimator unbiased in finite samples

- **Mean squared error**

  - $\beta < +\infty$:

$$\mathsf{MSE}(\widehat{\mu}) = \frac{1}{n[2-\Phi(\nu)]} + \frac{1}{4n^2}\widetilde{\beta}^2\phi(\nu)^2$$

$$\mathsf{MSE}(\widehat{\mu}_c) \simeq \frac{1}{n[2-\Phi(\nu)]} + \frac{1}{[2-\Phi(\nu)]^2\Phi(\nu)[1-\Phi(\nu)]n^2}\widetilde{\beta}^2\phi(\nu)^2$$

  - $\beta \to +\infty$:

$$\mathsf{MSE}(\widehat{\mu}) = \frac{1}{n[2-\Phi(\sqrt{n}\mu)]} + \frac{1}{4n}\phi(\sqrt{n}\mu)^2$$

$$\mathsf{MSE}(\widehat{\mu}_c) \simeq \frac{1}{n[2-\Phi(\sqrt{n}\mu)]} + \frac{1}{[2-\Phi(\sqrt{n}\mu)]^2\Phi(\sqrt{n}\mu)[1-\Phi(\sqrt{n}\mu)]n}\phi(\sqrt{n}\mu)^2$$

Likelihood Estimators: Conditional Bias

- Conditional likelihood estimator: conditionally unbiased
- Joint likelihood estimator ($\beta < +\infty$):

$$E(\overline{\mu}|N=n) \;=\; \mu + \frac{\beta}{\sqrt{1+\beta^2/n}} \cdot \frac{\phi(\nu)}{n\Phi(\nu)}$$

$$E(\overline{\mu}|N=2n) \;=\; \mu - \frac{\beta}{\sqrt{1+\beta^2/n}} \cdot \frac{\phi(\nu)}{2n[1-\Phi(\nu)]}$$

- Joint likelihood estimator ($\beta \to +\infty$):

$$E(\overline{\mu}|N=n) \;=\; \mu + \frac{\phi(\sqrt{n}\mu)}{\sqrt{n}\Phi(\sqrt{n}\mu)}$$

$$E(\overline{\mu}|N=2n) \;=\; \mu - \frac{\phi(\sqrt{n}\mu)}{2\sqrt{n}[1-\Phi(\sqrt{n}\mu)]}$$

- $\beta \to +\infty \quad \& \quad n \to +\infty$:

  ▷ $\mu < 0$

$$E(\overline{\mu}|N = n) \to 0$$

$$E(\overline{\mu}|N = 2n) \to \mu$$

  ▷ $\mu > 0$

$$E(\overline{\mu}|N = n) \to \mu$$

$$E(\overline{\mu}|N = 2n) \to \frac{\mu}{2}$$

  ▷ **Yet: asymptotically unbiased**

Lack of Incompleteness $\longrightarrow$ Counterintuitive Results

- N fixed $\Longrightarrow$ K complete

- Nice results apply

- **Lehmann-Scheffé theorem:**

$$\left.\begin{array}{l}\text{unbiased}\\ \text{complete}\\ \text{sufficient}\end{array}\right\} \Longrightarrow \text{best mean-unbiased}$$

- **Basu's theorem:**

$$\left.\begin{array}{l}\text{complete}\\[1em]\text{sufficient}\end{array}\right\} \quad \Longrightarrow \quad \text{independent of any ancillary statistic}$$

- **Not** so in our case!

- Sufficient statistic is (N, K)

- Joint distribution:

$$p_\mu(N,k) \;=\; p_0(N,k)\cdot \exp\left(k\mu - \frac{1}{2}n\mu^2\right)$$

$$p_0(n,k) \;=\; \phi_n(k)\cdot \Phi\left(\alpha + \frac{\beta}{n}k\right)$$

$$p_0(2n,k) \;=\; \phi_{2n}(k)\cdot \left[1 - \Phi\left(\frac{\alpha + \frac{\beta k}{2n}}{\sqrt{\frac{2n+\beta^2}{2n}}}\right)\right]$$

- To establish incompleteness, we must find a function $g(k,N)$ satisfying:

$$g(k,2n)\cdot p_0(2n,k) = -\int \phi_n(k-z)\cdot g(z,n)\cdot \phi_n(z)\cdot \Phi\left(\alpha + \frac{\beta}{n}z\right)\, dz$$

- Example:

$$g(k,n) \;=\; \frac{\ell}{\Phi\left(\alpha + \frac{\beta}{n}k\right)}$$

$$g(k,2n) \;=\; -\frac{\ell}{1 - \Phi\left(\frac{\alpha+\frac{\beta k}{2n}}{\sqrt{\frac{2n+\beta^2}{2n}}},\right)}$$

Ordinary Sample Average is Biased and Not Optimal

- **Generalized sample average:**

$$\overline{\mu} = \frac{K}{N} \cdot [c \cdot I(N = n) + d \cdot I(N = 2n)]$$

- Expectation:

$$E(\overline{\mu}) = d\mu + (c - d)\mu\Phi(\nu) + \frac{2c - d}{2n}\frac{\beta}{\sqrt{1 + \beta^2/n}} \cdot \phi(\nu)$$

Generalized Sample Size

$\beta = 0$ — completely random sample size

- A family of unbiased estimators:

$$d = \frac{1 - c\Phi}{1 - \Phi} \qquad \Longrightarrow \qquad E(\overline{\mu}) = \mu$$

- Minimum variance:

$$c_{\text{opt}} = 1 \; - \; \frac{1}{2n} \cdot \frac{1 - \Phi}{\mu^2 + \frac{2-\Phi}{2n}}, \qquad d_{\text{opt}} = 1 \; + \; \frac{1}{2n} \cdot \frac{\Phi}{\mu^2 + \frac{2-\Phi}{2n}}.$$

- **No uniform optimum**

- **Ordinary sample average is not optimal**

Generalized Sample Size
$\beta \neq 0$ — stopping rule

- All generalized sample averages **biased**

- This includes ordinary sample average!

- Asymptotically unbiased (as we known from likelihood)

Generalization 1: Arbitrary Number of Looks

- Our stopping rule

$$\Phi\left(\alpha + \frac{\beta}{n}k\right)$$

for $\beta \to +\infty$ can be generalized to arbitrary numbers of looks

$$n_1 < n_2 < \ldots < n_L$$

by applying

$$\Phi\left(\alpha_j + \frac{\beta_j}{n_j}k\right)$$

- Results carry over

Generalization 2: Completely Random Sample Size

- Allow for sample sizes with probabilities

n	0	1	$\cdots$	m
π_n	π_0	π_1	$\cdots$	π_m

- Incomplete sufficient statistic:

$$g(z,n) = b_n$$

$$\sum_{n=0}^{m} \pi_n \cdot b_n = 0$$

- Generalized sample average:

$$\overline{\mu} = \sum_{n=0}^{m} \frac{K}{n} \cdot a_n \cdot I(N = n)$$

- **Unbiased** generalized sample average:

$$a_n = 1 + b_n$$

- Optimal estimator:

$$a_n = \frac{\dfrac{1}{\mu^2 + \sigma^2/n}}{\sum_{k=0}^{m} \dfrac{\pi_k}{\mu^2 + \sigma^2/k}}$$

Generalization 3: Missing Data

- Our stopping rule

$$\Phi\left(\alpha + \frac{\beta}{n}k\right)$$

for $0 < \beta < +\infty$ corresponds to MAR missingness **in independent data**

- Can be generalized to longitudinal sequences

Joint likelihood	$\equiv$	selection model representation
Conditional likelihood	$\equiv$	pattern-mixture model

- Assume, for example

$$\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}, \boldsymbol{Y}_{i2})$$

  - ▷ first block always observed
  - ▷ second one possibly missing

- Sample average is **not** an option:

$$\widehat{\boldsymbol{\mu}}_1 \ = \ \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_{i1}$$

$$\widehat{\boldsymbol{\mu}}_2 \ = \ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_{i2} - \Sigma_{21}\Sigma_{11}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_{i1} - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_{i1}\right)$$

Generalization 4: Informative Cluster Sizes

- **Either:** Model for two random variables:
  - ▷ Outcomes $\boldsymbol{Y}_i$
  - ▷ Cluster size t_i
- **Or:** model for three random variables:
  - ▷ Outcomes $\boldsymbol{Y}_i$
  - ▷ Observed cluster size t_i
  - ▷ Theoretical cluster size $n_i \geq t_i$
- Entirely similar to missing-data setting

Generalization 5: Joint Model for Longitudinal and Survival Data

- Model for two, three, or four random variables:

  ▷ Longitudinal process $\boldsymbol{Y}_i$

  ▷ Time-to-event outcome T_i

  ▷ Censoring time C_i

  ▷ Missing data process $\boldsymbol{R}_i$

- Formulate an appropriate "shared-parameter model"

- Results carry over

Conclusions

- Commonality across seemingly unrelated settings

- Ordinary sample average loses its 'perfect status' $\longleftarrow$ lack of **completeness**
  - ▷ Already with completely random sample size
  - ▷ More so with deterministic or probabilistic stopping rule

- It nevertheless is a **valid estimator:**
  - ▷ Asymptotically unbiased (unbiased for CRSS)
  - ▷ Almost optimal
  - ▷ Asymptotically optimal
  - ▷ Consistent with the likelihood