Identification, Sensitivity Analysis, Prior Information

Joe Hogan
Brown University

May 19, 2008

Outline of Topics

Basic notation

$$\begin{aligned} Y_{\text{obs}} &= \text{observed response} \\ Y_{\text{mis}} &= \text{missing response} \\ X, V &= \text{covariates} \\ R &= \text{missing data indicators} \\ \theta &= \text{parameter of interest} \end{aligned}$$

Full data model

Underlying data-generating model

- Full-data model (everything conditioned on X)

$$f(Y_{\text{obs}}, Y_{\text{mis}}, R \mid \theta) = f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)\, f(Y_{\text{obs}}, R \mid \theta)$$

- For missing data and causal inference:
  - Have information on $f(Y_{\text{obs}}, R \mid \theta)$ – can model it
  - No information on $f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)$ – must extrapolate it
  - Only partial information about θ
  - Inference about θ depends on extrapolation

Full data model

Underlying data-generating model

- Full-data model (everything conditioned on X)

$$f(Y_{\text{obs}}, Y_{\text{mis}}, R \mid \theta) = f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)\, f(Y_{\text{obs}}, R \mid \theta)$$

- For missing data and causal inference:
  - Have information on $f(Y_{\text{obs}}, R \mid \theta)$ – can model it
  - No information on $f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)$ – must extrapolate it
  - Only partial information about θ
  - Inference about θ depends on extrapolation

Full data model

Underlying data-generating model

- Full-data model (everything conditioned on X)

$$f(Y_{\text{obs}}, Y_{\text{mis}}, R \mid \theta) = f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)\, f(Y_{\text{obs}}, R \mid \theta)$$

- For missing data and causal inference:
  - Have information on $f(Y_{\text{obs}}, R \mid \theta)$ – can model it
  - No information on $f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)$ – must extrapolate it
  - Only partial information about θ
  - Inference about θ depends on extrapolation

How to extrapolate

Must constrain or structure $f(Y_{\text{mis}} \mid Y_{\text{obs}}, R, \theta)$ using *subjective* assumptions

Assumptions are subjective because data have no information to verify them

- missing at random
- ignorable treatment assignment (conditional on X, V)
- no unmeasured confounding
- missing *not* at random

Focus should be on *model parameterization*

Full-data model should be indexed by one or more parameters that characterize non-identified parts of the distribution

Heuristic: For a working full-data model $f(Y_{\text{obs}}, Y_{\text{mis}}, R \mid \theta)$, let $\theta = g(\phi, \Delta)$.

- ϕ identified by (Y_{obs}, R)
- Δ *not* identified — ‘sensitivity parameter’

See Robins (1997 Stat Med), Rubin (1977 JASA), Vansteelandt et al (2006 Stat Sinica), Daniels and Hogan (2008).

Properties of effective model parameterization

- Δ must have coherent interpretation so we can argue about its most plausible values
- Model should be centered at familiar set of assumptions (MAR, no unmeasured confounding, etc.)
- Function $\theta = g(\phi, \Delta)$ should make clear what aspects of the model are driving inference about θ
  - proportion missing information
  - parametric assumptions about $f(Y_{\text{mis}} \mid Y_{\text{obs}}, R)$
  - departures from MAR

Effective parameterization in a simple case

Data

- Full-data response: (Y_1, Y_2)
- Y_2 missing on some individuals
- $R = 1$ if Y_2 observed; $R = 0$ if missing

Objective: Estimate $\theta = E(Y_2)$

Effective parameterization in a simple case

Model

$$
\begin{aligned}
E(Y_1) &= \mu_1 \\
E(R) &= \pi \\
E(Y_2 \mid Y_1, R = 1) &= \beta_0 + \beta_1 Y_1 \\
\\
E(Y_2 \mid Y_1, R = 0) &= (\beta_0 + \Delta_0) + (\beta_1 + \Delta_1) Y_1
\end{aligned}
$$

Effective parameterization in a simple case

What is Δ?

$$\begin{aligned} \Delta_0 + \Delta_1 Y_1 &= E(Y_2 \mid Y_1, R = 1) - E(Y_2 \mid Y_1, R = 0) \\ &= \text{difference in } E(Y_2 \mid Y_1) \text{ between those} \\ &\quad\ \text{with observed vs. missing } Y_2. \end{aligned}$$

- Identified and non-identified parameters well separated
- Model centered at MAR ($\Delta_0 = \Delta_1 = 0$)

Effective parameterization in a simple case

- Can convey influence of modeling assumptions on estimate of θ

$$\theta = E(Y_2) = \beta_0 + \beta_1\mu_1 + \pi(\Delta_0 + \Delta_1\mu_1)$$

- For 'sensitivity analysis', can plot

$$\theta = g(\phi, \Delta_0, \Delta_1)$$

as function of (Δ_0, Δ_1).

Warning to practitioners

Not all models admit this sort of parameterization!

- Parametric selection models
- Parametric versions of IV estimators

Must be aware of where identification is coming from

- Distributional assumptions (e.g. full data are normal)
- Modeling assumptions (e.g. linear trajectory over time)

Scharfstein approach

- Uses ‘effective parameterization’
- Incorporates lots of auxiliary information (a must)
  (Design: think carefully about auxiliary information)
- Keeps parametric assumptions mainly confined to observed data (can check these)
- Efficiency

Goetghebeur approach

- Also uses ‘effective parameterization’
- Intervals: reflect both sampling error and model uncertainty (should use even under ignorability?)
- Clear that the *causal* model potentially has several dimensions for sensitivity analysis
- Representation of sensitivity to ‘unmeasured confounding’
- Double robustness

How should inferences be reported?

- Interval estimate only
  - Conveys limitations of available information
  - Must account for sampling variation *and* lack of information
  - Intervals can be very wide in practice
- Sensitivity analysis
  - Plot $\theta(\Delta)$ as function of Δ
  - Conveys range of possible conclusions
  - How to get around multiple comparison problem?
  - Consumers gravitate to their favorite conclusion?
- Single summary
  - What is the most appropriate point estimate?
  - Posterior distribution?

Are we being Bayesian without realizing it?

Inference about incomplete data *requires* subjectivity

Why not formalize it?

- Prior formulation

$$p(\theta) = p(\phi, \Delta) = p(\Delta \mid \phi)\ p(\phi)$$

- Can use flat or vague priors for $p(\phi)$
- Subjectivity represented by $p(\Delta \mid \phi)$

Examples: Scharfstein et al (2003 Biostatistics); Daniels and Hogan (2008, Chapter 9).

Summary

- Inference from incomplete data requires *subjective* assumptions
- *Subjective* = cannot be verified by data even when $n = \infty$
- Key modeling objective: separate *identified* and *nonidentified* parts of the full data model
- DS and EG: outstanding examples; conveys complexity of the issues
- Role for Bayesian formalism?

Summary

Thanks to Liz Stuart!!