An Information Criterion for Marginal Structural Models

Robert W. Platt

Department of Epidemiology, Biostatistics, and Occupational Health and of Pediatrics
McGill University

March 11, 2011

Marginal Structural Models

Model marginal expectation as a function of time-varying exposure as a function of pre-defined time-varying treatment plans

- $E[Y_{\overline{X(t)}}(t)] = f(\overline{X(t)})$
- $Y_{\overline{X(t)}}(t)$ *potential outcome* at time t
- $\overline{X(t)}$ history of exposure X to time t
- Let Z denote a vector of covariates; $Z(t)$ represents Z at time t, $\overline{Z(t)}$ history to t.
- Interpretation: expected $Y(t)$ if all subjects followed $\overline{X(t)}$.

Marginal Structural Models - Simple Example

Model marginal expectation as a function of time-varying exposure as a function of pre-defined time-varying treatment plans

- X_0, X_1 two binary treatments
- Four possible treatment histories: $(0,0),(1,0),(0,1),(1,1)$
- an MSM models expected (average) outcome for each possible treatment history if ALL subjects were to follow that history
- e.g., $E[Y_{(1,1)}]$ is the average outcome if ALL subjects (possibly contrary to fact) were to receive $X_0 = 1$, $X_1 = 1$.

Assumptions

- No unmeasured confounding

$$Y_{\overline{X(t)}}(t) \coprod X(t) | \overline{X(t-1)}, \overline{Z(t)} \tag{1}$$

  - Treatment at t is independent of potential outcomes given history of treatment and covariates;
  - each treatment change is randomized given history
- Experimental treatment assumption - $P(\overline{X})$ is nonzero for all possible treatment histories.
- Every possible treatment history must have positive probability

Estimation

- Robins 1998, 1999, Hernán and Robins 2006: $E[Y_{\overline{X(t)}}(t)]$is the unique solution to the estimating equation

$$E[q(\overline{x(t)})(Y - c(\overline{x(t)}))/w(t)] \quad (2)$$

where

$$w(t) = \prod_{i=0}^{t} P(X(i) = x(i)|\overline{X(i-1)}, \overline{Z(i)}) \quad (3)$$

ie inverse probability of treatment received given history of treatment and covariates, and q is any function.

- Requires model for $w(t)$.
  - Robins 1998: $\widehat{w}$ must converge to w at rate $n^{1/4}$.

Previous Work

Specification of model for w

- Must include confounders
- May include predictors of outcome
- Should not include predictors of treatment (instruments)
- Should account for time-modified confounders
- What about the outcome model?

Outcome Model

Specification of model for Y

- Typically some function of the exposure
- Most HIV examples have used `cum(X)` - total amount of treatment received
- Has led to misconception that this functional form is part of the MSM!
- Functional form should reflect causal question under study
- What if uncertainty exists re causal question?

Outcome Model

- Could try multiple models
- How to evaluate/compare?
- Adjusted R^2?
- Some kind of information criterion?

Simple case: two time-point MSM

Let

- $\mathcal{X}$ denote a set of treatments that can be applied at any point in time, x_1, x_2 be a sequence of treatments
- Y_{x_1,x_2} be a counterfactual outcome corresponding to a sequence of treatments, and
- $\mathcal{S} = Y_{x_1,x_2}, (x_1, x_2) \in \mathcal{X}^2$ be the set of counterfactual outcomes corresponding to all possible treatment sequences.
- Let $X(t)$ denote the observed treatment at time t,
- $\bar{L}(t)$ denote the history of all covariates up to time t,
- $V \subset L(1)$ be some baseline covariates upon we which to condition.

Two time-point MSM

- Interested in estimating the conditional expectation of the counterfactual given V: $E[Y_{x_1,x_2}|V]$.
- If for each subject, we observed all counterfactual outcomes, $\mathcal{S}$, one could fit a model $m(x_1, x_2, V)$ of $E[Y_{x_1,x_2}|V]$ directly
- For example, $m(x_1, x_2, V) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
- Given a set of competing models that have been fit to the data, $\hat{m}_i, 1 \ldots I$, can we develop an information criterion?

QIC

We assume that the weight model w is correctly specified, and that it is constant across candidate m_i.

In the full (partially unobserved) data, we propose

$$QIC(\hat{m}) = 2p - \frac{1}{n}\sum_{i=1}^{n}\sum_{x_1,x_2\in\mathcal{X}^2}(Y_{(x_1,x_2),i} - \hat{m}(x_1,x_2,V_i))^2,$$

where p is the number of free parameters in the model.

With only the observed data, we choose the model that maximizes the inverse-probability weighted quasi-likelihood information criterion:

$$QIC_W(\hat{m}) = 2p - \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \hat{m}(X(1)_i,X(2)_i,V_i))^2}{P(X_i(2)=x_i(2)|\bar{L}_i(2),X_i(1))P(X_i(1)=x_i(1)|L_i(1))} \quad (4)$$

QIC- equivalence

It is straightforward to show that

$$QIC_W(\hat{m}) = QIC(\hat{m})$$

in the two time-point setting. This extends easily to more complicated models.

Simulations - Design

- 4 time points $i = 1, \ldots, 4$
- Treatment T_i, confounder L_i generated as:
  - L_1 $N(10, 1)$
  - T_i $Bin(p_i)$ where p_i a function of L_i and $T_{I=1}$
- Y Normal, function of T_i.

Simulations - Design

- 5 scenarios (others under consideration)
- 3 sample sizes
- Fit "full", "null", and "reduced" model (including only T_1 and T_2)

Simulations - Results

- Simpler models: QIC_w selects correct or over-fit model, adj. R^2 under-fit
- More complex models: QIC_w selects correct model, adj. R^2 under-fit
  - When all coefficients nonzero, QIC_w selects correct model 85-100% of the time
  - Adj. R^2 selects reduced model most of the time
- Performance improves with sample size.

PROBIT

- Breastfeeding promotion intervention
- 17 045 subjects
- Followed at $0, 1, 2, 3, 6, 9, 12$ months
- All mothers intended to breastfeed
- We considered models for weight at 12 mos as a function of breastfeeding duration

PROBIT - MSMs

Considered four models (M = months breastfed)

- Linear$E[Y_{12}] = \beta_0 + \beta_1 * M$
- Quadratic $E[Y_{12}] = \beta_0 + \beta_1 * M + \beta_2 M^2$
- Cubic $E[Y_{12}] = \beta_0 + \beta_1 * M + \beta_2 M^2 + \beta_3 M^3$
- "saturated" model with dummy variable for each time point

Results

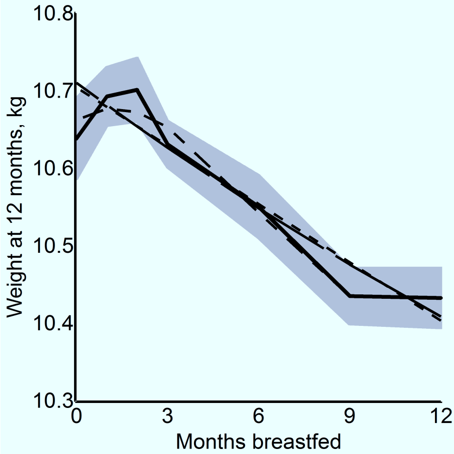

Figure: Plot of weight as function of months BF; shaded area confidence band for saturated model

Results II

Model	No. parms	QIC_w
Saturated	7	16,776
Linear exposure	2	16,784
Quadratic exposure	3	16,786
Cubic exposure	4	16,775

CD4 and HIV treatment

- Cole et al (AJE 2004) fit an MSM to CD4 count as a function of HAART treatment over time.
- Selected a model with a piecewise linear function
- linear from 0-1 year, and linear after 1 year.
- Is this best model?

Results

Model	No. parms	QIC_w
1. Intercept	1	931.77
2. Intercept and time a	5	496.94
3. Model 2 + linear exposure	6	482.11
4. Model 2 + curvilinear exposure	7	481.57
5. Model 2 + 2-part linear exposure	7	480.92
6. Model 2 + per visit (Saturated model)	25	516.58

Conclusions

- *QIC* appears to provide useful information for model selection
- Simulations: selects richer model
- Examples: chooses interesting models/provides insight

Limitations

- Proof (and simulations) assume weight model correctly specified
- No joint modeling/information criterion
- Assumes IPTW fitting of models

Future Work

- Joint modeling of weight and outcome: optimization criteria?
- Targeted Maximum Likelihood?
- Machine-learning orientation?

Thanks!

- FRSQ
- NSERC
- NIH/NICHD
- American Chemistry Council
- M. Alan Brookhart, Enrique Schisterman, Daniel Westreich, Steve Cole