The Partitioned Monte Carlo Expectation Maximization (PMCEM) algorithm

Michael D. Regier, Erica E. Moodie
McGill University,
Department of Epidemiology, Biostatistics,
and Occupational Health

CANNectIN

17 June 2011
Outline

Background

Model

The EM algorithm

PMCEM

Simulation study

Conclusions
The context: Marginal structural model (MSM)

The methodological context is that we are interested in obtaining the marginal effect of treatment on an outcome when

1. A set of covariates, $X$, confound the treatment, $A$, and the outcome, $Y$ and
2. Not all treatments are observable (i.e. there is a censoring mechanism).
Simplifying the MSM context

• We will consider the point-treatment context where at least one of the confounders is measured with error.

• It is a first step towards
  • Understanding the inherent complexities of the problem,
  • Identifying possible methodologies for unbiased estimation in the presence of measurement error, and
  • Extending the methods to more complex systems.
Under the counterfactual framework,

- Let \( a \) be a possible value of \( A \), and
- \( Y_a \) denote the potential response we would expect to observe if the subject followed treatment \( a \).
  - \( Y_{a=1} \) denote a subject’s outcome if treated, and
  - \( Y_{a=0} \) if untreated.
- For a continuous outcome, we want to estimate the marginal effect of treatment
- We use the marginal structural model \( \mathbb{E}[Y_a] = g(A : \beta) \) where \( \beta \) parameterizes the model.
- For the point treatment scenario we consider \( \mathbb{E}[Y_a] = \beta_0 + \beta_1 A \).
Point-treatment details

- Under the assumptions of consistency, exchangeability (Robins 1999), positivity (Hernan and Robins 2006), and time ordering where exposure precedes outcome (Mortimer 2005), we can obtain unbiased estimates of $\beta$.

- The estimates are obtained using a weighted M-estimator

$$\sum_{i=1}^{n} W_i(1, A)^{-1}(Y_i - \beta_0 - \beta_1 A_i)$$

- Through the creation of a pseudo-population, the weighting breaks the confounding relations.
  - $Y_a \perp\!\perp A|X$
Point-treatment stabilized weight

For the point-treatment scenario with censored treatments (Hernan et al. 2001), the stabilized weight for the $i^{th}$ individual is

$$sw_i = \frac{p(A_i = a_i|C_i = 0)p(C_i = 0)}{p(A_i = a_i|C_i = 0, X_i = x_i)p(C_i = 0|X_i = x_i)}$$
The focus of the investigation: The denominator

The key problem lies in the denominator,

\[ p(A_i = a_i|C_i = 0, X_i = x_i)p(C_i = 0|X_i = x_i) \]

- If a confounder \( X \) is unobservable but a proxy is used \( X^* \) then we are using

\[ p(A_i = a_i|C_i = 0, X_i^* = x_i^*)p(C_i = 0|X_i^* = x_i^*) \]
Measurement error

Two general types
- Classical measurement error models
  - The conditional distribution of $X^*$ given $X$ is modelled
- Regression calibration models
  - The conditional distribution of $X$ given $X^*$ is modelled
  - Berkson error models
Focus: Classical additive error model

The classical unbiased additive error model for the $i^{th}$ subjects is

$$X_i^* = X_i + \epsilon$$

where

- $X_i$ is the unobserved variable,
- $E(\epsilon|X_i) = 0$, and
- $Var(\epsilon|X_i) = \tau^2$.
- $\tau$ is directly interpreted as proportion of imprecision in the measurement of $X$ (Gustafson 2004)
Implications of measurement error

- Attenuation occurs in linear regression models when covariates are mismeasured.
- Gustafson shows that attenuation can be expected for parameters associated with mismeasured covariates in logistic regression.
- For both ordinary least squares and logistic regression attenuation is enhanced as the correlation amongst covariates strengthens (Gustafson 2004).
Effect of using proxy confounders

There are four effects of using proxy confounders in the denominator of the MSM weights on the parameter of interest, $\beta_1$:

1. Little effect,
2. Attenuation,
3. Augmentation, and
4. Sign reversal.
Little effect

- Background
- Model
- The EM algorithm
- PMCEM
- Simulation study
- Conclusions

Graphs showing little effect.
Attenuation
Augmentation
Sign reversal

Background

Model

The EM algorithm

PMCEM

Simulation study

Conclusions

-10 −5 0 5 10
-10 −5 0 5 10
S2
A
Y

-10 −5 0 5 10
-10 −5 0 5 10
S3
A
Y

-10 −5 0 5 10
-10 −5 0 5 10
S5
A
Y

-10 −5 0 5 10
-10 −5 0 5 10
S6
A
Y
Model of interest

• The portion of the stabilized weight which is of primary interest is the joint distribution found in the denominator of the stabilized weight:

\[ p(A_i = a_i | C_i = 0, X_i = x_i) p(C_i = 0 | X_i = x_i) \]

• We consider the situation where confounders are measured with error,

\[ p(A_i = a_i | C_i = 0, X_i^* = x_i^*, Z = z_i) p(C_i = 0 | X_i^* = x_i^*, Z = z_i) \]

where \( Z \) denotes observed and correctly measured confounders.
The joint distribution of the model of interest is

\[ p(A, C, X^*, X|Z = z_i; \theta) \]

where

- \( A \) denotes the treatment,
- \( C \) is binary and indicates censoring (\( C = 0 \), observed),
- \( X^* \) is the measured confounders,
- \( X \) is the unmeasured confounders, and
- \( Z \) are other perfectly measured covariates.
- \( \theta \) is the vector of parameters.
Log-likelihood

The associated complete data log-likelihood for the model of interest is

\[ \ell_c = \sum_{i=1}^{n} \log p(A_i|c = 0, x_i, z_i; \theta^A) + \log p(C_i = 0|x_i, z_i; \theta^C) \]

\[ + \log p(X_i^*|x_i, z_i; \theta^M) + \log p(X_i|z_i; \theta^X) \]

where

- \( \theta = \{ \theta^A, \theta^C, \theta^M, \theta^X \} \),
- \( \theta^A \) parameterizes the treatment model,
- \( \theta^C \) parameterizes the censoring mechanism,
- \( \theta^M \) parameterizes the measurement error model,
- \( \theta^X \) parameterizes the unobserved confounder(s) model.
- Assuming all models are uniquely parameterized.
Observed likelihood

Since $X$ is unobserved, we should use the observed likelihood,

$$L(\theta) = \prod_{i=1}^{n} \int_{\mathcal{X}} p(A_i, C_i, X_i^*, X_i | Z = z_i; \theta) d\nu_{X_i}$$

$$= \prod_{i=1}^{n} \int_{\mathcal{X}} L_{ci}(\theta) d\nu_{X_i}$$

Using the observed likelihood can be more complicated than using the complete-data likelihood, so we would like to use the complete-data likelihood instead.
The approach

We use the EM algorithm, a general iterative algorithm for maximum-likelihood estimation in incomplete-data situations.

We can view measurement error as a type of missing data problem.

- The $X$ are unobservable - missing, latent.
- We observe $X^*$ and we have or assume a functional relationship between $X^*$ and $X$.
- We assume that all observations are mismeasured under the same measurement error model.
Basic idea of the EM algorithm

Objective
Iterative procedure to obtain maximum likelihood parameters when maximum likelihood estimation would be straightforward, but there is the additional complexity of incomplete information.

Principle
The EM algorithm is less an algorithm and more a two-step general principle.
1. E-step: Take the conditional expectation of the complete likelihood, $\ell_c(\theta|\cdot)$, given the observed data.
2. M-step: Maximize the conditional expectation with respect to the parameter.


**EM Algorithm: The procedure**

- Unobservable complete-data log likelihood is replaced by the conditional expectation of the complete-data log likelihood given the observed data and current parameter estimates.
- For the \((t + 1)\)th iteration the **E-step** is:

\[
Q(\theta^{(t+1)}|\theta^{(t)}) = \sum_{i=1}^{n} \mathbb{E} \left[ \ell_c(\theta|\cdot)|\theta^{(t)}, \text{observed} \right]
\]

- For the **M-step**, choose \(\theta^{(t+1)}\) such that \(\theta^{(t+1)} \in \Theta\) and maximizes \(Q(\theta|\theta^{(t)})\),
  - i.e. \(\theta^{(t+1)} = \arg\max_{\theta \in \Theta} Q(\theta|\theta^{(t)})\)
The Generalized EM (GEM) algorithm

We can generalize the EM algorithm by modifying the M-step:

Choose $\theta \in \Theta$ such that $Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)})$

- We choose the updated parameter estimate to increase the $Q$-function rather than maximize it over the entire parameter space, $\Theta$.
- This is sufficient to ensure that $\mathcal{L}(\theta^{(t+1)}) \geq \mathcal{L}(\theta^{(t)})$.
- We are not decreasing the likelihood after each GEM iteration.
The Q-function for our likelihood

Assuming unique parameterization, the EM algorithm can be expressed as

\[
Q(\theta^{(t+1)}|\theta^{(t)}) = Q(\theta^{A(t+1)}|\theta^{A(t)}) + Q(\theta^{C(t+1)}|\theta^{C(t)}) \\
+ Q(\theta^{M(t+1)}|\theta^{M(t)}) + Q(\theta^{X(t+1)}|\theta^{X(t)})
\]
Observations from preliminary trials

It was observed that:

- The EM algorithm would often find a ridge or plateau.
- The expectation-conditional maximization (ECM) (Meng and Rubin 1993) had similar problems.
  - ECM uses a constraint function to define a sequence of conditional maximization steps for each iteration of the EM algorithm.
- Evidence strongly suggested that these problems were linked to the estimation of $\theta^X$. 
Response to the observations

- Keep the essential idea of the ECM, constrained optimization
- Break the problem into smaller and simpler steps
- Permit the use of Monte Carlo integration at each step.
- Combine these features in order to retain the desired properties of a GEM algorithm.
The basic idea of the PMCEM

The Partitioned Monte Carlo Expectation Maximization (PMCEM) algorithm is the result of breaking the problem into two simpler components.

We

1. Estimate \( \{\theta^C(t), \theta^M(t), \theta^X(t)\} \) first, then
2. Estimate \( \theta^A(s) \) while holding \( \{\theta^C(t), \theta^M(t), \theta^X(t)\} \) fixed.
The stages

- The first stage, estimation of \( \{\theta^C, \theta^M, \theta^X\} \), is the EM algorithm to the joint distribution

\[
p(C, X^*, X|Z = z_i; \theta)
\]

- The second stage, estimation of \( \theta^A \), is the EM algorithm on the conditional distribution

\[
p(A|C = 0, X^* = x^*, X = x, Z = z_i, \theta^C(t), \theta^M(t), \theta^X(t); \theta^A)
\]
The $Q$-functions for the two stages are

1. $Q(\theta^1(t) | \theta^1(t-1)) = Q(\theta^C(t), \theta^M(t), \theta^X(t) | \theta^C(t-1), \theta^M(t-1), \theta^X(t-1))$
2. $Q(\theta^2(s) | \theta^2(s-1)) = Q(\theta^A(s) | \theta^A(s-1))$

Such that $Q(\theta^{(k)} | \theta^{(k)}) = Q(\theta^1(t) | \theta^1(t)) + Q(\theta^2(s) | \theta^2(s))$

where $k = s + t$ and

$$Q(\theta^A(s) | \theta^A(s-1)) = \sum_{i=1}^{n} \mathbb{E} \left[ \log p(A_i | c = 0, x_i, z_i; \theta^A) | a_i, c, x_i^*, z_i, \theta^1(t); \theta^A(s-1) \right]$$
Immediate tasks

With this construction, there are two basic questions that need addressing,

1. Does it search over the entire parameter space?
2. Is this still a GEM, and
3. Do we still have stationary points?
A likelihood, $L(\theta|x)$, where $\theta \in \Theta$ and $X$ are vector valued, is said to be uniquely parameterized if it can be written as a product of conditional distributions,

$$L(\theta|x) = \prod_{j=1}^{p} p(X_{p-j+1}|x_1, \ldots, x_{p-j}; \theta^{p-j+1}),$$

such that $\theta^i \perp \theta^j$ for all $i \neq j$, $i = 1, \ldots, p$, and $j = 1, \ldots, p$ where $\theta^i \subset \Theta$ and $\theta^j \subset \Theta$. 
Proposition about unique parameterization

A likelihood may be uniquely parameterized if and only if

$$\Theta = \bigoplus_{j \in \{1, \ldots, p\}} \Theta^j,$$

for $p \geq 1$, where $\bigoplus$ denotes the direct sum of the direct summands $\Theta^j \subseteq \Theta$.

It is this definition of and proposition concerning unique parameterization which permits the aforementioned partitioning of the $Q$-function.
Formal definition of the PMCEM

A $K$-stage EM algorithm is called a Partitioned Monte Carlo EM algorithm (PMCEM) if the E-step can be written as

$$Q(\theta^{(t)}|\theta^{(t-1)}) = \sum_{j=1}^{K} Q(\theta^{(t_j)}|\theta^{(t_j-1)})$$

which finds $Q(\theta^{(t)}|\theta^{(t-1)})$ as a function of $\theta$ as in $Q(\theta|\theta^{(t)}) = E\{\ell(\theta|Y)|Y_o; \theta^{(t)}\}$, uses Monte Carlo integration, and maximizes $Q(\theta^{(t)}|\theta^{(t-1)})$ subject to

$$G = \{g_k(\theta) = \theta/\theta^k|k = 1, \ldots, K, K \leq p\}$$

using $K$ MCEM stages where $G$ defines the set of constraint functions over the parameter space.
Space-filling preliminaries

As shown by Meng and Rubin (1993), we use the equivalence between using feasible directions defined by the constraint functions and the closure of the resultant tangent cones with the span of the gradient of the constraint functions via the polar and bipolar theorems.
Constraint function gradient column space

- Use $G$
- Assume unique parameterization

then,

- The gradient of $g_k(\theta)$, $\nabla g_k(\theta)$, is full rank at $\theta^r \in \Theta$ for all $r$
- $C_k(\theta) = \{ \eta \nabla g_k(\theta) | \eta \in R^{d_k} \}$ is the column space of $\nabla g_k(\theta)$
  - where $\eta \in R^{d_k}$,
  - $R^{d_k} \subseteq R^{d_\Theta} \equiv \Theta$ and
  - $d_\Theta = dim(\Theta)$ and $d_k = dim(\theta^k)$
Space-filling

Now,

\[ C(\theta) = \bigcap_{k} C_k(\theta) = \emptyset, \]

thus

\[ C(\theta)^c = \bigcup_{k} \{ \eta \nabla g_k(\theta) | \eta \in R^{d_k} \}^\perp = R^{d_\theta}. \]

At each location in the parameter space, we can search in all directions for the next maximizing location!
The PMCEM is a GEM algorithm

Given that $G$ satisfies the space-filling criterion, we can show that the PMCEM algorithm is a GEM algorithm.

If \( Q(\theta^{(t)}|\theta^{(t-1)}) = \sum_{k=1}^{K} Q(\theta^{k(t_k)}|\theta^{k(t_k-1)}) \) is uniquely parameterized and the MCEM algorithm can be applied to the $k$th component, subject to $G = \{ g_k(\theta) = \theta/\theta^k | k = 1, \ldots, K, K \leq p \}$, such that there exists a $t_k$ for which $Q(\theta^{k(t_k+1)}|\theta^{k(t_k)}) \geq Q(\theta^{k(t_k)}|\theta^{k(t_k)})$ for all subsequent steps, $k = 1, \ldots K$, then the PMCEM algorithm is a GEM algorithm.
Limiting points are stationary points

If all $K$ maximizations of

$$Q(\theta^{(t)}|\theta^{(t-1)}) = \sum_{j=1}^{K} Q(\theta^{(t_j)}|\theta^{(t_j-1)})$$

are unique, then all the limit points of the $K$ MCEM sequences

$$\{\theta^{k(t_k)}, t_k \geq 0\}$$

are stationary points of $L_0(\theta|x_0)$ if $G$ is space filling at all $\theta^{k(t_k)}$. 

Implementation for each component

Use Monte-Carlo (MC) integration to approximate the expectation

\[
\tilde{Q}(\theta^p | \theta^{p(t)}) = \sum_{i=1}^{n} \frac{1}{m_{g_i}(t)} \sum_{l=1}^{m_{g_i}(t)} \ell_c(\theta^p)
\]

where \(m_{g_i}(t)\) is the Monte Carlo (MC) sample size as a function of the \(t\)th step for the \(i\)th subject
Making a long story short

- Use Gibbs Adaptive Rejection Sampling (GARS) algorithm (Wild 1993)
- Implicit in this choice is a restriction to log-concave functions (Gilks 1992 deals with the exponential family)
M-step

- Maximization of $\tilde{Q}(\theta^p|\theta^p(t))$ is equivalent to component-wise maximization.
- In many situations, this can be done using standard software.
Simulation set-up: Simulation structure

- 100 Simulations.
- Sample size, n=500 for each simulation.
- Monte Carlo sample size, 2500.
- Burn-in for MC integration, 1000.
- Dissimilarity criterion for each stage of the PMCEM:
  1. \(|\theta^1(t) - \theta^1(t-1)| \leq 0.0025\)
  2. \(|\theta^2(s) - \theta^2(s-1)| \leq 0.0025\)
Data generation DAG

Data was generated according to the following DAG.

- Dashed lines indicate that the censoring mechanism was not included as a covariate in the data generating models, but that it does affect what is observable.
Simulation set-up: Confounders

- $f(X_1, X_2) \sim \text{MVN}$.
- $\mu = (0, 0)$.
- $\sigma_{ii} = 1$ for $i = 1, 2$.
- $\sigma_{1,2} = 0.2$. 
Simulation set-up: Measurement error model

- Chose $X_1$ to be unobservable, but has observable surrogate $X_1^*$
- Unbiased classic measurement error model:
  - $X_1^* = X_1 + \epsilon$
  - $\epsilon \sim N(0, \tau)$
  - $\tau \in \{0.1, 0.3, 0.5, 0.7\}$
- We are assuming $\tau$ to be known
Simulation set-up: Model parameterization

- Censoring mechanism: logit[Pr(C = 1|x)] = \( \theta^C_0 + \theta^C_1 x_1 + \theta^C_2 x_2 \)
- Treatment model: logit[Pr(A = 1|x)] = \( \theta^A_0 + \theta^A_1 x_1 + \theta^A_2 x_2 \)

<table>
<thead>
<tr>
<th>Parameterization</th>
<th>( \theta^C_0 )</th>
<th>( \theta^C_1 )</th>
<th>( \theta^C_2 )</th>
<th>( \theta^A_0 )</th>
<th>( \theta^A_1 )</th>
<th>( \theta^A_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (P1)</td>
<td>0.025</td>
<td>0.378</td>
<td>-1.881</td>
<td>0.4</td>
<td>2.630</td>
<td>2.307</td>
</tr>
<tr>
<td>2 (P2)</td>
<td>0.025</td>
<td>-0.184</td>
<td>-2.821</td>
<td>0.4</td>
<td>2.446</td>
<td>0.864</td>
</tr>
<tr>
<td>3 (P3)</td>
<td>0.025</td>
<td>1.037</td>
<td>0.893</td>
<td>0.4</td>
<td>1.860</td>
<td>-0.683</td>
</tr>
</tbody>
</table>
Goal

Unbiased estimation of \( \{\theta^C, \theta^A\} \)
# Results for step 1: $\theta^1$

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\theta_0^C$</th>
<th>$\theta_1^C$</th>
<th>$\theta_2^C$</th>
<th>$\theta_0^C$</th>
<th>$\theta_{PMCEM}^C$</th>
<th>$\theta_2^C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.1</td>
<td>-1.1 (0.4)</td>
<td>0.5 (0.2)</td>
<td>-0.8 (0.3)</td>
<td>-1.1 (0.4)</td>
<td>0.5 (0.2)</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-1.0 (0.4)</td>
<td>-0.1 (0.2)</td>
<td>-0.6 (0.3)</td>
<td>-1.1 (0.4)</td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-1.1 (0.4)</td>
<td>0.5 (0.2)</td>
<td>-0.8 (0.3)</td>
<td>-1.1 (0.4)</td>
<td>0.5 (0.2)</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>-0.6 (0.4)</td>
<td>-1.0 (0.2)</td>
<td>&lt;0.1 (0.3)</td>
<td>-1.0 (0.4)</td>
<td>0.5 (0.3)</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>0.1</td>
<td>-0.3 (0.3)</td>
<td>-0.2 (0.2)</td>
<td>-0.5 (0.3)</td>
<td>-0.3 (0.3)</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.8 (0.4)</td>
<td>&lt;0.1 (0.2)</td>
<td>-0.6 (0.4)</td>
<td>-0.8 (0.4)</td>
<td>-0.2 (0.2)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-1.3 (0.4)</td>
<td>0.4 (0.2)</td>
<td>-1.3 (0.4)</td>
<td>-1.3 (0.4)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>-0.2 (0.4)</td>
<td>0.6 (0.1)</td>
<td>-0.4 (0.3)</td>
<td>-0.4 (0.4)</td>
<td>-0.1 (0.2)</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>0.1</td>
<td>-0.5 (0.4)</td>
<td>0.1 (0.3)</td>
<td>0.1 (0.2)</td>
<td>-0.5 (0.4)</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.1 (0.4)</td>
<td>-0.6 (0.2)</td>
<td>0.2 (0.3)</td>
<td>-0.6 (0.4)</td>
<td>0.4 (0.3)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3 (0.3)</td>
<td>-2.3 (0.3)</td>
<td>0.6 (0.2)</td>
<td>-0.9 (0.4)</td>
<td>0.1 (0.4)</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.8 (0.4)</td>
<td>-3.7 (0.2)</td>
<td>0.6 (0.3)</td>
<td>-1.1 (0.4)</td>
<td>0.3 (0.3)</td>
</tr>
</tbody>
</table>
Results for step 2: $\theta^2$

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\theta^A_0$</th>
<th>$\theta^A_1$</th>
<th>$\theta^A_2$</th>
<th>$\theta^A_0$</th>
<th>$\theta^A_1$</th>
<th>$\theta^A_2$</th>
<th>$\theta^A_{PMCEM}$</th>
<th>$\theta^A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>$&lt;0.1(0.1)$</td>
<td>-0.6 (0.2)</td>
<td>-0.1 (0.2)</td>
<td>-0.1 (0.1)</td>
<td>$&lt;0.1(0.2)$</td>
<td>0.2 (0.2)</td>
<td>0.2 (0.2)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2 (0.1)</td>
<td>-3.9 (0.2)</td>
<td>-0.9 (0.2)</td>
<td>-0.1 (0.2)</td>
<td>0.6 (0.3)</td>
<td>0.8 (0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.6 (0.2)</td>
<td>-9.0 (0.2)</td>
<td>-2.8 (0.2)</td>
<td>$&lt;0.1(0.2)$</td>
<td>0.7 (0.4)</td>
<td>0.5 (0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>1.0 (0.1)</td>
<td>-13.0 (0.1)</td>
<td>-4.1 (0.2)</td>
<td>-0.1 (0.2)</td>
<td>2.0 (0.6)</td>
<td>1.2 (0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.2 (0.1)</td>
<td>-0.3 (0.2)</td>
<td>$&lt;0.1(0.2)$</td>
<td>0.2 (0.1)</td>
<td>0.2 (0.2)</td>
<td>$&lt;0.1(0.2)$</td>
<td>0.2 (0.2)</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.4 (0.1)</td>
<td>-3.6 (0.2)</td>
<td>-0.3 (0.1)</td>
<td>0.2 (0.1)</td>
<td>0.2 (0.3)</td>
<td>-0.2 (0.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.7 (0.1)</td>
<td>-8.0 (0.2)</td>
<td>-0.7 (0.1)</td>
<td>-0.1 (0.2)</td>
<td>0.7 (0.4)</td>
<td>0.1 (0.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.9 (0.1)</td>
<td>-11.9 (0.1)</td>
<td>-0.7 (0.1)</td>
<td>$&lt;0.1(0.2)$</td>
<td>1.0 (0.5)</td>
<td>0.2 (0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>-0.1(0.1)</td>
<td>$&lt;0.1(0.1)$</td>
<td>$&lt;0.1(0.1)$</td>
<td>-0.1 (0.1)</td>
<td>0.3 (0.2)</td>
<td>$&lt;0.1(0.1)$</td>
<td>0.3 (0.2)</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.2 (0.1)</td>
<td>-2.0 (0.1)</td>
<td>0.4 (0.1)</td>
<td>$&lt;0.1(0.1)$</td>
<td>0.7 (0.2)</td>
<td>-0.3 (0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.5 (0.1)</td>
<td>-5.5 (0.1)</td>
<td>1.3 (0.1)</td>
<td>0.1 (0.1)</td>
<td>$&lt;0.1(0.2)$</td>
<td>-2.0 (0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.6 (0.1)</td>
<td>-8.2 (0.1)</td>
<td>2.1 (0.1)</td>
<td>-0.1 (0.1)</td>
<td>0.6 (0.3)</td>
<td>-0.2 (0.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- General conclusions
  - We formalize the idea of unique parameterization.
  - The PMCEM extends the EM algorithm by using key ideas of the ECM and MCEM algorithms.
  - The PMCEM produces unbiased estimates of model parameters for a class of models which has constrained covariates for one of the conditional models.
    - For example, the denominator of the stabilized weight when using inverse probability of treatment weights.
  - The PMCEM is a GEM and has its properties
  - The PMCEM permits searches over the entire parameter space.

- Simulation conclusions
  - The bias of the parameter of primary interest, $\theta^A$ was greatly reduced with little inflation of the standard error
  - The bias associated with the parameter of secondary interest, $\theta^C$, but not deemed a nuisance parameter, was also reduced.
  - Given these results, it is reasonable to posit that it may be possible to constrain $\theta$ to obtain a desirable efficiency for the targeted estimates.
References