

Thesis_Draft1

November 27, 2017

Contents

1	Abstract	2
2	Acknowledgements	2
3	Model	2
3.1	Introduction	2
3.1.1	Marginal Structural Models.	2
3.2	Time Dependent Confounding	3
3.3	Inverse Probability Weighting	4
3.4	Positivity	5
3.5	Static vs. Dynamic Strategies	6
3.6	Related work	6
3.7	Motivation	9
4	Simulating from a static MSM	9
4.1	Data Structure	9
4.2	Simulation Algorithm	10
4.2.1	Algorithm	10
4.2.2	Discussion of how algorithm works	12
4.3	Constructing IPT weights	12
4.4	Simulation Set-up	13
4.5	Results	13
5	Dynamic Case	13
5.1	The problem of simulating from a MSM under a dynamic strategy	13
6	Violations of Positivity	13
6.1	Extended discussion of algorithm linking to positivity	13
7	Application	14
8	Discussion and Conclusion	14
8.1	Limitations	14
9	References	14

1 Abstract

2 Acknowledgements

- Carlo
- Marta
- Olivia
- Work

3 Model

3.1 Introduction

The structure of this thesis is as follows. In section 2 of part 1, the model considered in this thesis and its important aspects are explained. In part 2 simulating from this statistical model is discussed in detail. In part 3 the model under dynamic strategies is considered and comparisons are drawn with the static case. In part 4 we entertain violations of positivity in the data, this section represents the novelty in this thesis. Part 5 conducts a simulation study. Part 6 includes a discussion, conclusions and suggestions for future work.

3.1.1 Marginal Structural Models.

We consider simulating from marginal structural models (MSM), a class of models that can be used to estimate causal effects from observational data under time dependent confounding. Specifically, we consider the problem of simulating longitudinal data from a follow-up study where a treatment variable A has a causal effect on an outcome of interest Y . The following definition from [Pearl \(2009\)](#) defines a causal effect:

The purpose of marginal structural models is causal inference.

Definition 1 (Def 3.2.1 from [Pearl \(2009\)](#), abridged) Given two disjoint sets of variables, X and Y , the causal effect of X on Y , denoted as $P(y \mid do(A = a))$, is a function from X to the space of probability distributions on Y . For each realization a of A , $P(y \mid do(A = a))$ gives the probability that $Y = y$ induced by deleting from the model all equations corresponding to variables in A and substituting a into the remaining equations.

A model that parameterises $P(Y \mid do(A = a))$ is called a marginal structural model (MSM) as it is marginal over any covariates and structural in the sense that it represents an interventional rather than observational model. The latter point is expressed in definition 1.1 through the notation $do(a)$ in $P(y \mid do(x))$, as opposed to the observational case $P(y \mid A = a)$. In the presence of confounding $P(y \mid A = a) \neq P(y \mid do(x))$ with the result that a model which is conditional on A will not represent the true causal effect of A on Y . In the former case the system is changed such that variable A takes on a particular value or history, the result represents the system under the intervention $do(A = a)$ ([pearl 2010](#)). On the other hand the first method represents the situation where the model is unchanged and A is observed to be a . This distinction will be particularly important when considering static and dynamic strategies in subsequent sections.

In their paper [Havercroft and Didelez \(2012\)](#) develop an algorithm that allows simulating data that corresponds to a particular parameterisation of an MSM. The DAG in figure 1 represents the one-shot non-longitudinal case. Factorising the joint distributions of the variables in figure 1 yields

$$P(U, L, W, A, Y) = P(W)P(U)P(W)P(L | U)P(A | L, W)P(Y | U, A)$$

Where, following definition 1.1 we delete $P(A | L, W)$, a probability function corresponding to A , and replace $A = a$ in all remaining functions

$$P(U, L, W, Y | do(A = a)) = \begin{cases} P(U)P(L | U)P(Y | U, A) & \text{if } A = a \\ 0 & \text{if } A \neq a \end{cases}$$

Our goal is to simulate from a particular MSM. This means parameterising $P(Y | do(A = a))$. Applying the law of total probability over W, U and L yields

$$P(Y | do(A = a)) = \sum_{w,u,l} P(W)P(U)P(L | U)P(Y | U, L, A = a) = \sum_{u,l} P(U)P(L | U)P(Y | U, L, A = a)$$

Making use of the fact that $P(L, U) = P(L | U)P(U) = P(U | L)P(L)$ and summing over either W and U or W and L yields

$$P(Y | do(A = a)) = \sum_l P(Y | L, A = a)P(L) = \sum_u P(Y | U, A = a)P(U)$$

If we can find suitable forms for either $P(Y | L, A = a)$ and $P(L)$ or $P(Y | U, A = a)$ and $P(U)$ that correspond to the MSM $P(Y | do(A = a))$, then, given suitable values for A, L, U it will be possible to simulate from the chosen MSM.

Choosing a functional form for

$$P(Y | do(A = a))$$

depends on convenience. We need a functional form that can be easily represented by $P(Y | L, A = a)P(L)$. non-linear functions will be hard to work into the analysis.

$U \sim U[0, 1]$ is a good choice because we can use the CDF of Y because $U[0, 1]$ is always between 0 and 1

General health is patient specific but comes from a clear distribution and has a nice medical interpretation. In contrast L would be more difficult to include. It is better as a function of U than a value in of itself.

3.2 Time Dependent Confounding

In the one-shot case the outcome Y depends on the treatment decision, measured covariates and unmeasured covariates. In the longitudinal case, the outcome depends on the histories of the treatment and covariates. One complication is that current values of the covariate may depend on previous treatments, and previous treatments may, in turn, depend on previous covariates. If treatment is succesful this will inform future treatments and also affect the outcome of interest Y . While in the one-shot case in figure 1 regression adjustment using

A time varying confounder affected by prior treatment

Explain in the context of time dependent confounding what a naive estimator is.

L affects whether a patient will receive treatment, and is associated with future survival. It is also itself affected by previus treatment Young etal 2013

Similarly, if treatment is succesful this will affect both the outcome and subsequent treatments. Analogous to the one-shot case, $P(y | \bar{A} = \bar{a}) \neq P(y | do(\bar{A} = \bar{a}))$. Regression adjustment will always introduce other sources of bias in the time dependent case. As a result regression adjustment does not work. Only the IPTW method adjust for selection bias and confounding.

The conclusion is that methods which use regression adjustment will not be applicable in the longitudinal setting

The time dependent confounding becomes a problem because the confounder then depends on both the previous treatment and the previous treatment depends on the confounder. There is always confounding in this case (Hernan, Robins 2004) caused by selection bias.

L_k may be a confounder for later treatment and must be adjusted for, And L_k may be affected by earlier treatment and therefore should not be adjusted by standard methods Robins(2000)

A covariate L is a confounder if it predicts the event of interest and also predicts subsequent exposure. Explain how this actually happens, as U_0 is a common ancestor of A through L and also Y , also that there is selection bias, and L is sufficient to adjust for confounding see Havercroft algorithm code page bottom.

Important to explain why we cannot account for confounding by making regression adjustments.

The reason why U is important is because with U , we can obtain any $P(Y \mid \text{do}(a))$ using the inverse of U ?? We can do this under any counterfactual survival path.

3.3 Inverse Probability Weighting

Inverse Probability of Treatment Weighting (IPTW) is a method for estimating MSMs in the presence of time dependent confounding. For each observation a weight is calculated which is informally the inverse of the probability that a patient receives their own treatment. This has the effect of creating a pseudo population consisting of w_i copies of each subject i where w_i are the IPT weights. In the pseudo population A is no longer confounded with L and crucially $P(Y = 1 \mid A)$ is the same as the true population values [Robins et al. \(2000\)](#).

IN previous simulation studies unstabilized weights show substantial increase in SEs

Weightin creates a pseudo -population in which the exposure is independent of the **measured** confounders (Hernan, Cole, 2008) The weight is informally proportional to the participants probability of receiving her own exposure history

As these weights have high instability we need to stabilize them. The unstabilized weights can be driven by only a small number of observations. Under time dependent confounding it may still be possible to recover the causal effect of A on Y by the method of Inverse Probability of Treatment (IPT) weighting. How does this work?

- true weights are unknown but can be estimated from the data.
- Robins(2000) - when there are no unmeasured confounders, we can control for confounding using weights
- weighting adjusts for confounding and selection bias due to measured time varying covariates affected by prior exposure.
- A_t is no longer affected by L_t , and crucially the causal effect of \bar{A} on Y remains unchanged

No unmeasured confounding time dependent confounding in the set-up we consider, will always introduce bias. In the time dependent case, we will therefore always have bias. And IPT weighting a means of avoiding this.

treatment increases/improves L and this in turn affects the probability of getting treatment A .

lack of adjustment for L precludes unbiased estimation. This is because it introduces selection bias.

Problem is that the weights will have very high variability, so we need to stabilize them otherwise they will affect the estimates. Explain all the little details about why it becomes more variable

etc. IPTW is a method of correcting for time dependent confounding when fitting marginal structural models.

No matter whether we use stabilized or unstabilized weights, under positivity violations the weights will be undefined.

By including the weights we can then include regress Y on A in a model that does not include L at all. This is important for the MSM above, because if we choose the model conditional on L , we lose the L altogether and just have a model conditional on A .

Conditions under which IPTW work are largely untestable (westreich 2012)

Informally a patients weight through visit k is proportional to the inverse of the probability of having her own exposure history through visit k (Cole and Hernan 2008)

Be more specific about what is contained in the weights. The denominator depends on the measured confounders L the numerator does not, but could contain baseline coefficients in the numerator to help stabilize the weights.

weighted regression and MSM are equivalent.

3.4 Positivity

Estimating the weights described in the previous section relies on four assumptions of consistency, exchangeability, positivity and non mispecification of the model [Cole and Hernán \(2008\)](#). In this thesis we examine the effect of positivity violations on our ability to recover the true model parameters from a given MSM.

Positivity is the condition that in every strata of the covariates there is a non-zero probability of receiving each level of treatment. In the one shot case represented by the DAG in figure 1 this would correspond to the formal statement $P(A | A_0, L) > 0$ while in the longitudinal case, incorporating the histories of the treatment and covariates, the positivity assumption can be stated

$$P_{\tau}(A_{\tau} | \bar{A}_{\tau-1}, \bar{L}_{\tau}) > 0$$

As the definition of the weights is:

Also need to condition on survival up to the present time.

When this is not met, the weights will be undefined, and the resulting estimates will be biased.

from Westreich 2012: positivity requires that there are treated and untreated individuals at every combination of the values of the confounders at every combination of the values of the confounders in the observed data.

Positivity can be violated if, for example, medical protocols demand that a doctor always treat a patient if a risk factor falls below a certain threshold. there are both exposed and unexposed individuals at every level of the confounders (Cole, Hernan, 2008). If positivity is violated because doctors have protocols to treat patients when L falls below a certain level, then positivity is violated. If the structural bias occurs within levels of a time-dependent confounder then restriction or censoring may lead to bias whether one uses weighting or other methods (Cole and Hernan 2008). In fact, weighted estimates are more sensitive to random zeroes (Cole, Hernan, 2008)

When positivity is violated, the weights are undefined because In this case the weights may lead to a biased estimate of the causal effect of A on Y . Examining the extent of this bias in MSMs will be the central purpose of this thesis.

Definition 2: positivity (Appendix 2 Cole and Hernan 2008): Positivity states that there is a non-zero (i.e. positive) probability of receiving every level of exposure X_{ij} for every combination of values of exposure and covariate histories that occur in individuals histories.

In section V we will consider violations of positivity to examine the extent of the bias introduced into the estimates when the assumption of positivity is not met. The inverse probability weights are equal to the stabilized weights, only if positivity holds. If positivity does not hold, the weights are undefined and may result in biased estimates of the causal effects.

As described above, the model only holds as long as key assumptions are met, and one of these is positivity. Introducing violations of positivity can be achieved by censoring observations.

- Point out difficulty in simulation from this model.
- talk about blocking a back door from U to Y?
- Pearl and Robins (1995) probabilistic evaluation of sequential plans from causal models with hidden variables.

3.5 Static vs. Dynamic Strategies

So far we have considered static strategies, in this section we describe the differences between static and dynamic strategies. A static strategy is one where the value where the values that A will take depend on for A is represented as:

Where $a(t) = 1$ if the strategy specifies that a is to take the value 1 at time t . In contrast, a dynamic regime is any well-specified way of adjusting the choice of the next decision (treatment or dose to administer) in the light of previous information constitutes a dynamic decision (or treatment) strategy (Didelez arxiv).

To our knowledge positivity violations in the dynamic case have not been considered in the literature.

Hernan et al (2005) for looking at comparing dynamic regimes using artificial censoring.

3.6 Related work

The aim is to be able to simulate the survival function of a desired MSM under the intervention $do(\bar{a})$

Several studies have developed algorithms for simulating data from marginal structural models in the presence of time dependent confounding. An early example is ? who study estimators of the causal effect of a time dependent treatment on survival outcomes. They compare naive estimators with IPTW and a treatment orthogonalized estimator which is also developed in the paper. This study shares similarities with [Havercroft and Didelez \(2012\)](#)

- stay on treatment after treatment starts
- treatment regime is determined by t^* (starting point of treatment because it is a vector of $\{0, 0, 0, 1, 1, 1\}$)
- they motivate a logistic model for the hazard function, they use a discrete equivalent to the hazard function (link to citation about farington study.)
- the survival time U is directly linked to the survival outcome -> here it is good to provide more intuition.
- need to understand why it is linked.

Work that proposes a different method to solve the same problem. Work that uses the same proposed method to solve a different problem. A method that is similar to your method

that solves a relatively similar problem. A discussion of a set of related problems that covers your problem domain.

These previous studies have lacked a systematic investigation of positivity violations in a simulation setting. It is unknown, for example, how large an effect a violation of positivity has and how it is affected by the sample size, threshold etc etc.

1. marginal structural models literature
2. Simulating from marginal structural models
3. positivity
4. connect limited work on simulating from models to positivity
5. Explain why we choose the Havercroft simulation of the others, specifically why it helps us to incorporate violations of positivity in our analysis.

Bryan et al 2010 have a similar focus as Havercroft 2012 in that they develop an algorithm for simulation from a given MSM and they use this algorithm to compare IPTW methods to naive regression methods.

Several studies have considered simulation from marginal structural models. The finite-sample properties of marginal structural proportional hazards models has been considered by ?.

- What is their focus?
- how do they simulate
- what do they find (in terms of MSE, SE, etc.
- how does it differ from HD (2012)

Young et al (2014) also provide a simulation algorithm for simulating longitudinal data from a known Cox MSM.

- What is their focus?

to compare IPW and standard regression based techniques. This is not the subject of this thesis.

Comparing IPW and standard regression based estimates in the absence of model misspecification. This allows for complete isolation of any type of bias. This approach involves simulating data from a standard parametrization of the likelihood and solving for the underlying Cox MSM

- how do they simulate
- what do they find (in terms of MSE, SE, etc.
- how does it differ from HD (2012)

we describe an approach to Cox MSM data generation that allows for a comparison of the bias of IPW estimates versus that of standard regression-based estimates in the complete absence of model misspecification

- could do this section in comparison to Havercroft and didelez, explain how their algorithm works first and then describe the related work by linking differences in their algorithm to earlier or other algorithms.

ALgorithm:

1. generate survival under no treatment from a weibull
2. generate survival times under the ten non-zero treatment levels.
- 3.

Bryan et al (2010) Cole (2008) for more of a discussion about positivity, with an actual observed data example. While Cole (2008) have looked at positivity in an observational setting, to our knowledge no study has looked at positivity violations within a simulation setting.

Cole and Hernan 2008 have examined the four assumptions underlying IPW using a study on real data from the HAART SWISS study.

1. 2 paper on simulation + Bryan paper - why this simulation method is different from other sim papers
2. related work - Judea pearl, Robins, econometrics, related and broader literature on MSM
3. positivity long discussion in Hernan and cole and the warnings but no simulation study in that paper. They study the positivity but not the effect and there is nothing in the havercroft paper on this either.

-

- Major point is that there are a number of ways of simulating from marginal structural models. But, we need one where we can mess with the positivity. Other methods are not suitable for this.
- G formula simulation
- exposure, confounder feedback loop
- treatment
- outcome variable
- causal effect of the treatment variable on the outcome.
- external intervention
- Explain the do notation, and what precisely is meant in the case
- used to estimate the joint effect of time dependent treatments on survival
- Need to strongly link the time dependent confounding to the MSM, do we choose this class of models because of their relationship with time dep confounding? Yes, marginal structural models are used with TDC
- The causal graph helps (according to Pearl pp. 40) to bridge statistics into causality
- There is a key part to this, which is that we do not observe confounding, this seems to be what motivates the use of the MSM class of models.

- Counterfactuals need to be addressed here to make it clear this is not the purpose of the thesis.
- Robins (2000) have demonstrated that in the presense of time dependent confounders, standard approaches for asjusting for confounding are biased.
- A covariate that is a risk factor for, or predictor of the event of interest (Y) (from Robins 2000). This defines a time dependent covariate
- And also past exposure determines the level of the covariate.
- Works under a set of assumptions (consistency, exchanchability, positivity and no misspecification of the model used to estimate the weights
-

3.7 Motivation

Positivity masking study

Explain connection between the MSM and the survival function

4 Simulating from a static MSM

4.1 Data Structure

We wish to simulate survival data in discrete time $t = 0, \dots, T$ for n subjects. At baseline $t = 0$ all subjects are assumed to be at risk of failure so that $Y_0 = 0$. For each time period $t = 0, \dots, T$ a subject may either be on treatment, $A_t = 1$, or not on treatment, $A_t = 0$. All patients are assumed to be not on treatment before the study begins. Once a patient commences treatment, they remain on treatment in all subsequent periods until failure or the end of follow-up. In each time period L_t is the value of a covariate measured at time t . In the simulated data, L_t behaves in a similar manner to CD4 counts such that a low value of L_t represents a more severe illness and hence a higher probability of both tratemnt and failure in the following period. In addition to L_t , the variable U_t represents subject specific general health at time t . Although we will simulate U_t , in a real world application U_t is an unmeasured confounder which

Each time period is either a check up visit or is between two check up visits. If t is a check-up visit and treatment has not yet commenced, L_t is measured and a decision is made on whether to commence treatment. Between visits, treatment remains unchanged at the value recorded at the previous visit. Similarly, L_t which is only measured when t is a visit, alos remains unchanged.

We represent the history of a random variable with an over bar. For example, the vector representing the treatment history of the variable A is represented by $\bar{A} = [a_0, a_1, \dots, a_m]$ where $m = T$ if the subject survives until the end of follow-up, or $m < T$ otherwise. Prior to baseline both $A = 0$ for all subjects.

- explain what U is and how it relates to the simulation design/algorithm
- Be more specific on Y
- L_t is a measured confounder
- U_t is an unmeasured confounder.

4.2 Simulation Algorithm

4.2.1 Algorithm

Next, we describe the algorithm used to simulate data from our chosen marginal structural model under time dependent confounding. In the following section we discuss in detail how the algorithm works and the salient features for this thesis. The algorithm is taken from [Havercroft and Didelez \(2012\)](#) who generate data on n patients, for k time periods. The outer loop in the following algorithm $i \in 1, \dots, n$, refers to the patients while the inner loop $t \in 1, \dots, T$ refers to the subject specific time periods from baseline to failure or the end of the study. There will be at least one,

and at most T records for each patient.

Algorithm 1: Simulation Algorithm MSM

Result: Marginal Structural Model Under Time Dependent Confounding

```

for  $i$  in  $1, \dots, n$  do
   $U_{0,i} \sim U[0, 1]$ 
   $\epsilon_{0,i} \sim N(\mu, \sigma^2)$ 
   $L_{0,i} \leftarrow F_{\Gamma(k, \theta)}^{-1}(U_{i,0}) + \epsilon_{0,i}$ 
   $A_{-1,i} \leftarrow 0$ 
   $A_{0,i} \leftarrow \text{Bern}(\text{expit}(\theta_0 + \theta_2(L_{0,i} - 500)))$ 
  if  $A_{0,i} = 1$  then
     $T^* \leftarrow 0$ ;
  end
   $\lambda_{0,i} \leftarrow \text{expit}(\gamma_0 + \gamma_2 A_{0,i})$ 
  if  $\lambda_{0,i} \geq U_{0,i}$  then
     $Y_{1,i} \leftarrow 0$ 
  else
     $Y_{1,i} \leftarrow 1$ 
  end
  for  $k$  in  $1, \dots, T$  do
    if  $Y_{t,i} = 0$  then
       $\Delta_{t,i} \sim N(\mu_2, \sigma_2^2)$ 
       $U_{t,i} \leftarrow \min(1, \max(0, U_{t-1,i} + \Delta_{t,i}))$ 
      if  $t \neq 0 \pmod k$  then
         $L_{t,i} \leftarrow L_{t-1,i}$ 
         $A_{t,i} \leftarrow A_{t-1,i}$ 
      else
         $\epsilon_{t,i} \sim N(100(U_{t,i} - 2), \sigma^2)$ 
         $L_{t,i} \leftarrow \max(0, L_{t-1,i} + 150A_{t-k,i}(1 - A_{t-k-1,i}) + \epsilon_{t,i})$ 
        if  $A_{t-1,i} = 0$  then
           $A_{t,i} \sim \text{Bern}(\text{expit}(\theta_0 + \theta_1 t + \theta_2(L_{t,i} - 500)))$ 
        else
           $A_{t,i} \leftarrow 1$ 
        end
        if  $A_{t,i} = 1 \wedge A_{t-k,i} = 0$  then
           $T^* \leftarrow t$ 
        end
      end
       $\lambda_{t,i} \leftarrow \text{expit}(\gamma_0 + \gamma_1[(1 - A_{t,i})t + A_{t,i}T^*] + \gamma_2 A_{t,i} + \gamma_3 A_{t,i}(t - T^*))$ 
      if  $1 - \prod_{\tau=0}^t (1 - \lambda_{\tau,i}) \geq U_{0,i}$  then
         $Y_{t+1,i} = 1$ 
      else
         $Y_{t+1,i} = 0$ 
      end
    end
  end
end

```

Within the inner loop ($t \in 1, \dots, T$) we see that the data is only updated at time $t \neq 0 \pmod k$,

where k refers to evenly spaced check-up visits. If t is not a check-up visit the values of A_t and L_t are the same as in $t - 1$. When t is a visit A_t and L_t are updated.

- if treatment has been commenced then a subject may feel extra benefit if more time has elapsed since treatment began
- L_t affects A_t and also Y_t
- explain starting values for A and Y are all zero (except L maybe)

In order to operationalize the Algorithm 1 we need to choose parameters for (\cdot) . In their paper [Havercroft and Didelez \(2012\)](#) use values that simulate data with a close resemblance to the Swiss HIV Cohort Study. We postpone discussion of the parameters in Algorithm 1 to section 2.4.

4.2.2 Discussion of how algorithm works

The algorithm of [Havercroft and Didelez \(2012\)](#) works by factorizing the joint density of the histories of the four variables in the analysis.

- Important is that the form of the MSM is not specified until the last stage
- role of $U_{0,i}$
- How does positivity enter the analysis?
- Why this model is important in terms of positivity.

4.3 Constructing IPT weights

Inverse Probability of Treatment weights can be used to adjust for measured confounding and selection bias in marginal structural models. Link back to pseudo population idea in previous section. This method relies on four assumptions consistency, exchangeability, positivity and no misspecification of the model used to estimate the weights [Cole and Hernán \(2008\)](#). Unstabilized weights are defined as:

$$w_{t,i} = \frac{1}{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})}$$

Where the denominator is the probability that the subject received the particular treatment history that they were observed to receive up to time t , given their prior observed treatment and covariate histories (Havercroft, Didelez, 2012). The probabilities $p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})$ may vary greatly between subjects when the covariate history is strongly associated with treatment. In terms of the resulting pseudopopulation, very small values of the unstabilized weights for some subjects would result in a small number of observations dominating the weighted analysis. The result is that the IPTW estimator of the coefficients will have a large variance, and will fail to be normally distributed. This variability can be mitigated by using the following stabilized weights

$$sw_{it} = \frac{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i})}{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})}$$

In the case that there is no confounding the denominator probabilities in the stabilized weights reduce to $p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i})$ and $sw_{it} = 1$ so that each subject contributes the same weight. In the case of confounding this will not be the case and the stabilized weight will vary around 1.

In practice, we estimate the weights from the data using a pooled logistic model for the numerator and denominator probabilities. The histories of the treatment and covariates are included

in the probabilities. In practice Specifically, following Havercroft and Didelez (2012), we estimate the model where the visit is only the visits every check up time. Between check ups both the treatment and covariate remain the same. Other ways of doing this include a spline function over the months to create a smooth function between the visits. Another difference might be to use a coxph function instead of logistic function

$$\text{logit } p_{\tau}(A_{\tau,i} | \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i}) = \alpha_0 + \alpha_1 k + \text{alpha}_2 a_{k-1} + \dots + \text{alpha}_k a_0 +$$

We have several options for estimating these weights. We could use a coxph model, or a logistic model.

4.4 Simulation Set-up

We follow the simulation set-up of Havercroft, Didelez (2012) which is based on parameters that closely match the Swiss HIV Cohort Study (HAART).

4.5 Results

- check the distribution of the weights that come out of the model (see Cole 2008). This would allow us to see weight model mispecifications. Not a problem in the simulation case.
- compare the bias, se, MSE, and 95% confidence interval
- compare all of these in the positivity violation and non-positivty violation case.
- explain to some extent monte-carlo standard error.
- we don't confirm the results of the havercroft of Bryan papers, instead refer readers to these papers to see how IPTW outperforms the naive estimators.

5 Dynamic Case

5.1 The problem of simulating from a MSM under a dynamic strategy

6 Violations of Positivity

- creating an artificial population in which positivity is violated in specific ways.

6.1 Extended discussion of algorithm linking to positivity

As described in the introduction, one assumption of the model is that there is a non-zero probability of the event occuring at every startum of the covariate.

- When previous covariates like CD4 count are strongly associated with treatment the probabilities in the denominator of the ustabilized weights may vary greatly. Because we are foric-ing positivity by using a treatment rule when L falls below a threshold and A is then eausal to one, we create a strong association between A and L -> hence the unstabilized weights would vary. (Robins et al 2000 pp. 553)
- present the algorithm again with positivity violations.

7 Application

8 Discussion and Conclusion

8.1 Limitations

9 References

References

Stephen R Cole and Miguel A Hernán. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6):656–664, 2008.

WG Havercroft and V Didelez. Simulating from marginal structural models with time-dependent confounding. *Statistics in medicine*, 31(30):4190–4206, 2012.

Judea Pearl. *Causality*. Cambridge university press, 2009.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology, 2000.