International Journal of Development, Vol.3, No.(2) (2014): 11-20
ISSN: 2314-5536 **e-ISSN: 2314-5544 (Online)**
www. ijd.byethost13.com **e-mail: fas_ijd@yahoo.com**

Estimating optimal decisions using stochastic dynamic models

Deyadeen Ali Alshibani
Libyan Academy
Department of Mathematics and Statistics

ABSTRACT

Stochastic Dynamic Models are functions of decision and covariate history which are used to advice on decisions to be taken. Murphy (2003) and Robins (2004) have proposed models and developed semi-parametric methods for making inferences about the optimal dynamic treatment regime in a multi-interval study that provide clear advantages over traditional parametric approaches.

In this paper the author investigates the estimation of optimal dynamic treatment decisions based on a full parametric approaches: Inverse Probability Treatment Weighted and Regret-Regression of Henderson *et al*. (2010). A numerical example on determination of optimal decisions is presented in detail.

Key words: Optimal decisions, stochastic dynamic models.

INTRODUCTION

The treatment decision strategies play a critical role in the method of several diseases such as AIDS or cancers. For such diseases, physicians face the difficult problem of deciding when and which drugs to administer to a patient. Furthermore, for such diseases, the strength and interaction of the treatment with the immune system are so complex that the design of the optimal treatment strategy may require some complicated analysis. A comparison of the regret-regression and inverse probability of treatment weighting is presented. We have more efficiency by choosing a fully parameterized model for the mean. The regret-regression method for estimating the causal effect of the actions on the final response allows diagnostic model assessment and model comparisons. Probability of treatment weighting is presented. We have more efficiency by choosing a fully parameterized model for the mean. The regret-regression method for estimating the causal effect of the actions on the final response allows diagnostic model assessment and model comparisons.

Notation

At the start, we need some notation. Our development extends to the general case. K denotes the number of intervals, j is a specific interval so that $j = 1, 2, 3, ..., K$, M_j represents a status variable available at the start of the j^{th} interval, M_j may be scalar or multivariate. T_j is the treatment at interval j given subsequent to observing M_j, Y is the outcome observed at the end of the K^{th} interval, and large values of Y are preferred. So, the occurrence is $(M_1, T_1, M_2, T_2, \cdots, M_K, T_K, Y)$, $\bar{M}_j$ denotes a status variable at time j and its history, *e.g.* $\bar{M}_j$ =$(M_1, M_2, ..., M_j)$. Also $\bar{T}_j = (T_1, T_2, ..., T_j)$, specific values are denoted with the lower case, e.g. $m_1, t_1, m_2, \cdots, t_K, m_K$. Also $\bar{t_j} = (t_1, t_2, ..., t_j)$ and d_j a rule or regime and $\underline{d}_{j+1}^{opt}$ is interpreted to mean that optimal rules are followed from time $j + 1$ onward.

Assumption for causal inference

Three assumptions sufficient to identify the average causal effect are

Consistency:

The potential outcome under any particular treatment or action regime corresponds to the actual outcome if that regime is followed. This states that the results of a subject's treatment allocation are not affected by other subjects' treatment allocation, Formally, the assumption implies that $Y_i(T=a)=Y_i(T=b)$ whenever $a_i=b_i$ (Rosenbaum and Rubin, 1983).

No unmeasured cofounders:

The assumption says that any regime t of T, received in any interval is conditional on history, but is independent of any future potential outcome. This means $Y(t)\perp T|M=m$ for each possible value t of T and m of M. This assumption is sometimes referred as the conditional exchangeability (sequential randomization). The sequential version of no unmeasured assumption at time j.

$$T_j\perp M_{j+1}(\bar{t}_j),\cdots,M_K(\bar{t}_{K-1}),Y(\bar{t}_K)|\bar{M}_j,\bar{M}_{j-1}$$

and at the last time point K

$$T_K\perp Y(\bar{t}_K)|\bar{M}_K,\bar{M}_{K-1}.$$

Positivity:

The treatment is not deterministically allocated within any level m of the covariates M. That is, not all source population subjects with a given value m of M are assigned to be treated or untreated *(Hernan and Robins, 2006)*. If $P(M=m)\neq 0$ (the population marginal probability that M takes the value m) then $P(T=t|M=m)>0$. Generally those hold in a randomized experiment. Without additional assumptions, the optimal regime might be estimated from among the set of feasible regimes (Robins 2004).

Regret-Regression:

In samples of modest size there is no realistic alternative to parametric modeling of at least some components of the terms needed to determine an optimal regime. In turn this brings the risk that the chosen model is not suitable for the data. Fundamental statistical practice of model building, checking and comparison has had little attention so far in this literature.

A direct class can be based on modeling of $E[Y|\bar{M}_j,\bar{T}_{j1}]$ or $E[Y|\bar{T}_j]$. The problem is then how to tease out the causal effects of actions, which may require some form of dynamic programming as well as additional modeling of M_{j+1} given $(\bar{M}_j,\bar{T}_j)$.

The computational burden of such an approach scales dramatically with K and soon becomes infeasible. Structural nested mean also fall within the direct class and have an advantage in interpretability. Computational issues remain formidable however. The indirect approach by contrast does not attempt to model the response Y. Instead, causal effects expressed as differences between counterfactuals (outcomes that might have occurred) are parameterized. Examples of these are the regrets of Murphy (2003) and the blips of Robins (2004). Interpretation of estimates is then easier but now model adequacy is less straightforward, since there is no model for the observed response. Computational problems are reduced but not removed.

Estimating optimal decisions using stochastic dynamic models

Regret-Regression proposes a modeling and estimation strategy which incorporates the regret functions of Murphy (2003) into a regression model for observed responses. It present and apply a method which is straightforward to implement, item provides direct estimates of causal parameters, allows diagnostic model assessment and model comparisons. With the regrets defined as in Murphy (2003, equation 12) showed that

$$E(Y|\bar{M}_K,\bar{T}_K) = \beta_0(M_1)+\sum_{j=2}^{K}\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j)-\sum_{j=1}^{K}\mu_j(T_j|\bar{M}_j,\bar{T}_{j-1}),$$

where

$$phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j) = E\{Y(\underline{d}_j^{opt})|\bar{M}_{j-1},\bar{T}_{j-1},M_j\}-E\{Y(\underline{d}_j^{opt})|\bar{M}_{j-1},\bar{T}_{j-1}\}$$

which compares the expected response under the optimal rule *after* M_j is revealed with the corresponding expected value *before* M_j is revealed. Thus the achieved response $Y$is affected by the initial condition (through β_0), the chosen actions (T_j) (through the regrets $\mu$) and the chance development over time of the states (M_j) (through the ϕ terms). Turning to estimation, both Murphy iterative minimization and Robins G-estimation methods require knowledge of the action probability distribution used in data generation, $P(t_j|\bar{M}_j,\bar{T}_{j-1})$ In a randomized trial this would of course be known, but more generally it needs to be estimated. Regret-Regression proposal is that instead of avoiding the $\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j)$ terms. It explicitly parameterizes them, as $\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j;\beta)$ say, and then simultaneously estimate β and ψ by regressing the observed responses on their associated expectations. It is not free to parameterize $\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j)$ arbitrarily however. As

$$\begin{aligned}\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j) &= E\{Y(\underline{d}_j^{opt})|\bar{M}_{j-1},\bar{T}_{j-1},M_j\} - E\{Y(\underline{d}_j^{opt})|\bar{M}_{j-1},\bar{T}_{j-1}\}\\ &= E\{Y(\underline{d}_j^{opt})|\bar{M}_{j-1},\bar{T}_{j-1},M_j\}\\ &- E_{M_j|\bar{M}_{j-1},\bar{T}_{j-1}}[E\{Y(\underline{d}_j^{opt})|\bar{M}_{j-1},\bar{T}_{j-1},M_j\}],\end{aligned}$$

It is shown that by construction $E_{M_j|\bar{M}_{j-1},\bar{T}_{j-1}}\{\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j)\}=0.$ Any parameterization needs to respect this condition: the expected value over M_j of each $\phi_j(.)$ term, given the past, needs to be zero. The Regret-Regression proposal is straightforward: model $\phi_j(\bar{M}_{j-1},\bar{T}_{j-1},M_j)$ as a linear combination of residuals between M_j (or functions thereof) and the corresponding conditional expectation given$(\bar{M}_{j-1},\bar{T}_{j-1})$. Hence It defines $Z_j = M_j - E(M_j|\bar{M}_{j-1},\bar{T}_{j-1})$ and note that the expectation is identified from observational data for $(\bar{M}_{j-1},\bar{T}_{j-1})$ values of interest. Then assume

$$E(Y|\bar{M}_K,\bar{T}_K) = \beta_0(M_1)+\sum_{j=2}^{K}\beta_j^T(\bar{M}_{j-1},\bar{T}_{j-1})Z_j-\sum_{j=1}^{K}\mu_j(T_j|\bar{M}_j,\bar{T}_{j-1}).$$

Here $\beta_j(\bar{M}_{j-1},\bar{T}_{j-1})$is a coefficient vector measuring the effect of M_j $after\ allowing$ for $\bar{M}_{j-1},\bar{T}_{j-1}$ and assuming optimal actions are chosen from time j onward.

Formally there can be a different coefficient for each possible history but in practice we may choose to simplify. Note (that the zero mean requirement) given history is immediate since the $\phi_j(.)$ terms are replaced by linear combinations of residuals.

The Inverse Probability of Treatment Weighted (IPTW):

As alternative approach to estimate causal parameters is to use the Inverse Probability of Treatment Weighted (IPTW) method. The IPTW formula based on $\bar{M}$ for the counterfactual mean $E(\bar{Y}_t)$ is the average of Y among subjects with $\bar{T} = \bar{t}$ in a stabilized or un stabilized pseudo-population constructed by weighting each subject by their subject-specific stabilized IPTW

$$SW = \prod_{j=1}^{K} \frac{f(T_j|\bar{T}_{j-1})}{f(T_j|\bar{T}_{j-1}, \bar{M}_j}).$$

or their un stabilized IPTW. When the three conditions hold

$$W = \prod_{j=1}^{K} \frac{1}{f(T_j|\bar{T}_{j-1}, \bar{M}_j})$$

either IPTW creates a pseudo-population in which the mean of Y_t is identical to that in the actual population but the randomization probabilities at each time j depend at most on past treatment history. The only difference is that in the un stabilized pseudo-population

$$P_{ps} = (T_K = 1|\bar{T}_{K-1}, \bar{M}_K) = 0.5$$

The only difference between stabilized and un stabilized IPTW is that in the un stabilized pseudo-population $P(T = t) = 0.5$ while in the stabilized pseudo-population $P(T = t)$ is as in the actual population. Thus $E[Y_t]$ in the actual population is $E_{ps}(Y|T = t)$ where the subscript ps is to remind us that we are taking the average of Y among subjects with $T = t$ in either pseudo-population. In summary, when the three conditions hold, the average causal effect $E(Y_{t=t_1}) - E(Y_{t=t_0})$ in the population is the difference $E_{ps}(Y|T = t_1) - E_{ps}(Y|T = t_0)$ in the pseudo-population.

Regret-Regression and IPTW:

In this section a comparison of the regret-regression and inverse probability of treatment weighting is presented. We have more efficiency by choosing a fully parameterized model for the mean. The regret-regression method for estimating the causal effect of the actions on the final response allows diagnostic model assessment and model comparisons. It requires modeling of $E(M_j|\bar{M}_{j-1}, \bar{T}_{j-1})$, to remove the effect of the time-dependent confounders on the counterfactuals, which is not used for the Murphy (2003) iterative estimation or the Robins (2004) G-estimation methods. On the other hand regret-regression has not required $E[T_j|\bar{M}_j, \bar{T}_{j-1}]$, which is needed for the others.

Simple example:

The following is an extremely simple example described in the previous section. All subjects are assumed to start in the same state M_1, which can thus be ignored. We consider a sequentially randomized trial in which $N =1000$. Patients are randomly assigned at time $K = 1$ to treatment $T_1 = 1$ with probability 0.5 and to placebo $T_1 = 0$ otherwise. Patients continue on treatment or placebo until their next visit to clinic at time $K = 2$, where they are again randomly assigned to take treatment with their probabilities. Table 1 provides the number of subjects and the average value of the outcome $E[Y|T_1, M_2, T_2]$. Y_t denote the counterfactual or potential outcome for a subject under treatment level $T = t$. We have two counterfactual variables $Y_{t=1}$ and $Y_{t=0}$. For example, if a subject's outcome would be 8 under treatment and would be 3 under non treatment, then we can write $Y_{t=1} = 8$, $Y_{t=0} = 3$ and $Y_{t=1} - Y_{t=0} = 5$. For the actual study, if this subject was treated, then his observed Y will be 8. Furthermore an observed outcome Y is the counterfactual outcome Y_t for a subject who is treated with level t.

Table 1. The data

Row	T_1	M_2	T_2	N	$E[Y\|T_1, M_2, T_2]$
1	0	0	0	50	3
2	0	0	1	50	8
3	0	1	0	120	2
4	0	1	1	280	7
5	1	0	0	280	6
6	1	0	1	70	5
7	1	1	0	135	4
8	1	1	1	15	1

Optimal actions can be taken by working from the final time point and choosing the actions when regrets are equal to zero. To choose optimal actions at the final time point, regrets are directly calculated. At other time point $j = 1, \cdots, k-1$, we might choose optimal actions by calculation of regrets through the expectation of optimal final rewards given the history of previous states and actions. Regrets for making the wrong decision at the second time-point can be read directly from the figure, as 5,5,1 and 3 for (T_1, M_2) equal to (0,0), (0,1), (1,0) and (1,1) respectively. Choice $T_1 = 0$ is optimal for the first time-point and the regret for choosing $T_1 = 1$ can be worked out to be 1.8.

A direct approach, which involves two stages. The first stage, regression, involves modeling the observable data. The second stage, dynamic programming (DP) or backward induction, uses the models to determine optimal actions, working iteratively from the last time stage. For the simple example there are eight different $T_1M_2T_2$ sequences and hence eight parameters in a saturated model for the response Y. Using the standard main effects and interaction formulation these are

Const	T_1	M_2	T_1M_2	T_2	T_1T_2	M_2T_2	$T_1M_2T_2$
3	3	-1	-1	5	-6	0	-2

From this we can calculate the mean response at each of the eight $T_1M_2T_2$ sequences and hence the regrets due to choices T_2 for each(T_1, M_2). Dealing with the first decision time is trickier: for each of the two values of T_1 we need to calculate

$$\sum_{M_2} E[Y|T_1, M_2, T_2^{opt}]P(M_2|T_1)$$

from which the optimal choice and regret can be found. In total, for general K we need $2^{2K-1} + 2^{2K-3} + \ldots + 2$ expectations. Moreover, we need the multivariate distribution of possible states M as well as our model for Y.

Regret-Regression Method:

In order to complete our treatment of the simple example, we will anticipate a little and apply the regret-regression method to be described in the next section. Let $I(t_1) = I(T_1 = t_1), I(t_1 m_2) = I(T_1 = t_1, M_2 = m_2)$
and $I(t_1 m_2 t_2) = I(T_1 = t_1, M_2 = m_2, T_2 = t_2)$. Further, let $Z_2(t_1)$ be the residual between M_2 and its expected value given$T_1 = t_1$. So $Z_2(0) = M_2 - p_0$ and $Z_2(1) = M_2 - p_1$ where $p_0 = 0.8$ and $p_1 = 0.3$, both of which would need to be estimated in practice. Instead of the eight-parameter main effects and interaction model summarized in the previous section, we obtain exactly the same saturated fit using a linear model with the eight covariates given below, along with their associated parameter values.

Const	$I(1)$	$Z_2(0)I(0)$	$Z_2(1)I(1)$	$I(000)$	$I(010)$	$I(101)$	$I(111)$
7.2	-1.8	-1	-2	-5	-5	-1	-3

This time we can read off directly that the mean response if the optimal regime is always followed is 7.2. Choosing the wrong action at the first decision time will cost 1.8 in mean. The wrong actions at the second time lead to regrets of 5,5,1 and 3 depending on the earlier sequence(T_1, M_2). The other two terms measure the effects of M_2 after allowing for the effect of T_1. For each value of T_1 in this example having $M_1 = 1$ is associated with a decrease in mean: by $1 \times (1 - 0.8)$ if $T_1 = 0$ and by $2 \times (1 - 0.3)$ if $T_1 = 1$, in both cases assuming the optimal T_2 is chosen. In fitting this model we have chosen the covariates knowing which action is optimal at each time. Reaching this point is trivial: we first take a working optimal for each decision and define the covariates accordingly. For example we might have included $I(001)$ instead of $I(000)$. At the first fit the signs of the final four coefficients determine which actions are indeed optimal: a positive value appears if the working version is wrong. We thus obtain the true optimal decisions at the second time and then re-fit the model. This time the sign of the second coefficient - the regret at time 1 - determines the optimal. Thus only two model fits are required. Each requires a linear model and negligible effort .

IPTW Method

We now describe how to use IPTW for estimating the counterfactual means $E(Y_{\bar{t}})$ under the four regimes $\bar{t} = \{t_1, t_2\}$. The first step is to create a stabilized pseudo population by weighting the subjects in each row in Table 2 by the stabilized weights using the IPTW that

$$SW = \frac{P(T_1 = t_1)P(T_2 = t_2|T_1 = t_1)}{P(T_1 = t_1|M_1 = m_1)P(T_2 = t_2|M_2 = m_2, T_1 = t_1)}$$

Because all subjects are assumed to start in the same state M_1, the factor $P(T_1 = t_1)$ cancels, because in our study the potential confounder M_1 is absent. So the formula will be as follows

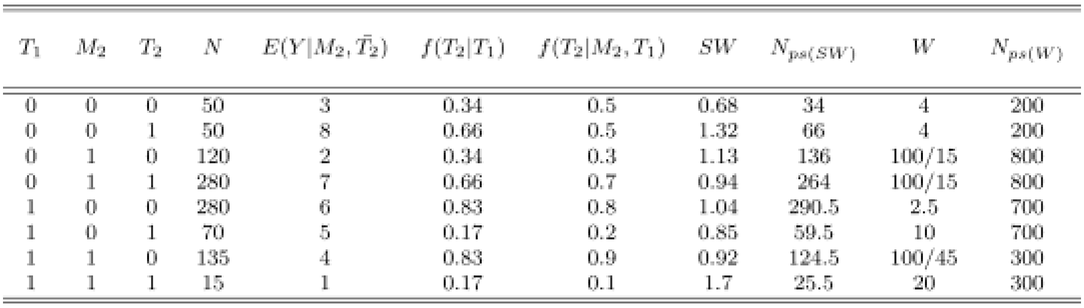
$$SW = \frac{P(T_2 = t_2 | T_1 = t_1)}{P(T_2 = t_2 | M_2 = m_2, T_1 = t_1)}$$

or by un stabilized weights, the formula is

$$W = \frac{1}{P(T_1 = t_1 | M_1 = m_1) P(T_2 = t_2 | M_2 = m_2, T_1 = t_1)}.$$

Then using one of these formulas we can obtain the pseudo population, then estimate directly the means under each of the regimes.

Table 2. Data to use IPTW for estimating E(Y_t^-)



T_1	M_2	T_2	N	$E(Y \mid M_2, \bar{T}_2)$	$f(T_2 \mid T_1)$	$f(T_2 \mid M_2, T_1)$	SW	$N_{ps(SW)}$	W	$N_{ps(W)}$
0	0	0	50	3	0.34	0.5	0.68	34	4	200
0	0	1	50	8	0.66	0.5	1.32	66	4	200
0	1	0	120	2	0.34	0.3	1.13	136	100/15	800
0	1	1	280	7	0.66	0.7	0.94	264	100/15	800
1	0	0	280	6	0.83	0.8	1.04	290.5	2.5	700
1	0	1	70	5	0.17	0.2	0.85	59.5	10	700
1	1	0	135	4	0.83	0.9	0.92	124.5	100/45	300
1	1	1	15	1	0.17	0.1	1.7	25.5	20	300

For example, for the first row: $f(T_2|T_1) = P(T_2 = 0|T_1 = 0) = 170/500 = 0.34$ and $f(T_2|T_1, M_2) = P(T_2 = 0|T_1 = 0, M_2 = 0) = 50/100 = 0.5$. Each of the 50 subjects in the first row therefore receives the weight $SW_{000} = 0.34/0.5 = 0.68$. Hence, the row has $0.68 \times 50 = 34$ subjects in the stabilized pseudo-population. This is in column $N_{ps(SW)}$ in Table 2. The other terms in the column are calculated similarly. Also each of the subjects in the first row can be received the weight$W_{000} = \frac{1}{(0.5)(0.5)} = 4$. Hence, the row has $4 \times 50 = 200$ subjects in the un stabilized pseudo-population. The IPTW weights cancel the arrow between M_2 and T_1 in the pseudo-population as shown in Table 2. The absence of the arrow can be easily confirmed by checking whether $T_2 \perp M_2 | T1$, where $\perp$ represents independence in the pseudo-population. For example, this conditional independence holds in the stabilized pseudo-population of our example because $P_{ps}(T_2 = 1|T_1 = 0, M_2 = 0) = \frac{66}{100} = P_{ps}(T_2 = 1|T_1 = 0, M_2 = 1) = \frac{264}{400} = 0.66,$ and $P_{ps}(T_2 = 1|T_1 = 1, M_2 = 0) = \frac{59.5}{350} = P_{ps}(T_2 = 1|T_1 = 1, M_2 = 1) = \frac{25.5}{150} = 0.17.$ Using the IPTW stabilized or un stabilized pseudo-population formula, the four means under each of the regimes are

$$E_{ps(SW)}(Y_{\bar{t}=\{0,0\}}) = \frac{3 \times 34 + 2 \times 136}{170} = 2.2,$$

$$E_{ps(SW)}(Y_{\bar{t}=\{0,1\}}) = \frac{8 \times 66 + 7 \times 264}{330} = 7.2,$$

$$E_{ps(SW)}(Y_{\bar{t}=\{1,0\}}) = \frac{6 \times 290.5 + 5 \times 124.5}{415} = 5.4,$$

$$E_{ps(SW)}(Y_{\bar{t}=\{1,1\}}) = \frac{4 \times 59.5 + 1 \times 25.5}{85} = 3.8, \quad \text{or}$$

$$E_{ps(W)}(Y_{\bar{t}=\{0,0\}}) = \frac{3 \times 200 + 2 \times 800}{1000} = 2.2,$$

$$E_{ps(W)}(Y_{\bar{t}=\{0,1\}}) = \frac{8 \times 200 + 7 \times 800}{1000} = 7.2,$$

$$E_{ps(W)}(Y_{\bar{t}=\{1,0\}}) = \frac{6 \times 700 + 5 \times 300}{1000} = 5.4,$$

$$E_{ps(W)}(Y_{\bar{t}=\{1,1\}}) = \frac{4 \times 700 + 1 \times 300}{1000} = 3.8.$$

The following table shows the average values of the outcome $E_{ps}[Y|T_1, M_2, T_2]$ of the four static regimes, using the both of the stabilized and the un stabilized pseudo-population

$Table\ 3:\ Stabilized\ and\ unstabilized\ pseudo{-}population\ for\ E_{ps}[Y|T_1, M_2, T_2].$

T_1	T_2	$N_{ps(SW)}$	$N_{ps(W)}$	$E_{ps}[Y\|T_1, T_2]$
0	0	170	1000	2.2
0	1	330	1000	7.2
1	0	415	1000	5.4
1	1	85	1000	3.8

As expected, the values of $E_{ps}(Y|T_1, T_2)$ obtained by IPTW, in the pseudo population, are equal to those obtained by the G-formula. In this example, we do not need to use models to estimate the inverse probability weights because we can be easily calculated by hand from the data. Also, we do not need models for the counterfactual means $E_(Y_{\bar{t}})$ because these means can also be calculated by hand .

Let us consider the marginal structural mean model

$$E_(Y_{\bar{t}}) = \gamma_0 + \gamma_1 t_1 + \gamma_2 t_2 + \gamma_3 t_1 t_2$$

The model is referred to as a marginal structural mean model (MSM) because it models the marginal mean of the counterfactuals Y_t and models for counterfactuals are often referred to as structural models, Herna'n at al (2006). If we simply fit the model to calculate the parameters $\gamma_0, \gamma_1, \gamma_2$ and γ_3. We obtain

$$E(Y|T_1, T_2) = 2.2 + 2.5T_1 + 5T_2 - 7T_1T_2,$$

This gives biased estimates of $\gamma' s$ because of confounding by M_2. Now we can use the pseudo-population data because

$$\begin{aligned} E[Y_{\bar{t}=\{0,0\}}] &= \gamma_0, \\ E[Y_{\bar{t}=\{1,0\}}] &= \gamma_0 + \gamma_1, \\ E[Y_{\bar{t}=\{0,1\}}] &= \gamma_0 + \gamma_2, \\ E[Y_{\bar{t}=\{1,0\}}] &= \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3. \end{aligned}$$

Using the estimates for $E_{ps}(Y|T_1, T_2)$ in Table 3, $\gamma_0 = 2.2$, $\gamma_1 = 5.4 - 2.2 = 3.2$, $\gamma_2 = 7.2 - 2.2 = 5$ and $\gamma_3 = 3.8 - 2.2 - 5 - 3.2 = -6.6$. This estimation procedure is equivalent to fitting a linear model with each subject weighted by SW as follows

$$E_{ps}(Y|T_1, T_2) = 2.2 + 3.2T_1 + 5T_2 - 6.6T_1T_2$$

Simulation Results

As shown, the regret-regression and inverse probability of treatment weighting methods have the ability to estimate an optimal dynamic treatment regime by choosing a fully parameterized model for the mean. The regret-regression estimates the direct effect of treatments on the potential final responses. Basically it requires modeling of $E[M_j|\bar{M}_{j-1}, \bar{T}_{j-1}]$, to remove the effect of the confounders (the intermediate states $M_2, \cdots, M_k$) on those outcomes. Instead of that the inverse probability of treatment weighting needs modeling of $E[T_j|\bar{M}_j, \bar{T}_{j-1}]$, which is needed as well for both Murphy iterative minimization (IMOR) and Robins G-estimation.

The following are simulations results on estimation of the optimal final response. For each simulation we use 100 datasets of samples size 100 and 1000 .

Table 4 : Comparing Regret−regression and IPTW for estimating optimal regimes

Sample size	SD	Reget-regression		IPTW	
		Mean optimal response	SE	Mean optimal response	SE
100	0	7.2	0.06	7.2	0.06
	0.25	7.2	0.11	7.2	0.11
	0.5	7.19	0.13	7.19	0.13
	1	7.18	0.22	7.18	0.22
1000	0	7.2	0.02	7.2	0.02
	0.25	7.2	0.02	7.2	0.02
	0.5	7.2	0.04	7.2	0.04
	1	7.2	0.06	7.2	0.06

Discussion

As we seen inverse probability of treatment weighting and regret-regression give exactly the same results. An important advantage of the regret-regression is to remove the effect of the confounder's covariates and use a unique model. By using Regret-Regression Method, We are able to find optimal dynamic regimes using the regret idea by choosing the optimal actions of T_j which are depend on the value of M_j, where $j = 2, \ldots, k$. Hence the optimal dynamic roles using all the two methods are the same, but the regret-regression method avoids problems which arise when using samples of modest size or when there is need to estimate high-dimensional parameters.

REFERENCES

Hernan, M.A. and Robins, J.M. (2006). Estimating causal effects from epidemiological data. J. Epidemiol. and Community Health, 60: 578-586.

Murphy, S. (2003). Optimal dynamic treatment regimes (with discussion). J. Royal Statistical Soc., ser. B, 65:331-366.

Henderson, R.; Ansell, P. and Alshibani, D. (2010). Regret-regression for optimal dynamic regimes. Biometrics, 66(4): 1192-1201.

Robins, J.M. (2004). Optimal structured nested models for optimal sequential decision. Proc. 2nd Seattle Symposium on Biostatics, ed D.Y. Lin and P.J. Heagerty, New York: Springer, 189-326.

تقدير لقرارات المثلى للنماذج كاملة المعالم

ضياء الدين الشيبانى
قسم الرياضيات بالاكاديمية الليبية فرع مصراته

المستخلص

النماذج الديناميكية العشوائية هي دوال للناتج النهائي في بدائل ومتغيرات عشوائية ومتغيرات لبيانات من الماضي خلال فترات زمنية متعددة، تستعمل لتحديد القرارات المثلى. وفي هذا المجال حقق كل من مورفي (2003) وروبن (2004)، نتائج متقدمة باستخدام نماذج نصف معلمية بدلا من نماذج الانحدار التقليدية التي يتعذر استخدامها بسبب مشاكل العشوائية. في هذه الورقة، سنبحث في كيفية تقدير القرارات المثلى للنماذج كاملة المعالم مثل طريقة الاحتمال الموزون المعكوس وطريقة دالة انحدار الندم باستخدام دراسة مقارنة وتوضح ذلك من خلال افتراض حالة عددية.