

伽马分布结局下基于目标最大似然估计的因果推断研究

周新茹, 侯文*

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2026年6月7日; 录用日期: 2026年7月1日; 发布日期: 2026年7月8日

摘要

传统的因果推断方法(如G计算和逆概率加权)在估计平均处理效应时, 往往依赖于结果模型或倾向得分模型的正确设定, 当模型被错误指定时可能产生有偏估计, 存在稳健性不足的问题。目标最大似然估计通过引入波动参数对初始结果模型进行靶向更新, 提高了模型的双重稳健性和估计效率。本文基于伽马分布假设, 通过构造结果回归模型和倾向得分模型的得分函数, 利用目标最大似然估计方法得到了平均处理效应的估计结果。蒙特卡洛模拟结果表明, 在所有模型设定场景下目标最大似然估计均表现稳健, 当任一模型正确时保持无偏估计, 估计精度随样本量增加而提高。对比于单一的G计算方法, 基于目标最大似然估计的估计结果在模型错误设定时具有更小的偏差和更优的均方误差。将所提方法应用于ACTG175艾滋病数据集, 估计结果显示联合治疗可显著提升CD4细胞计数, 验证了方法的稳健性和实用性。

关键词

因果推断, 目标最大似然估计, 平均处理效应, 双重稳健性

Research on Causal Inference Based on Targeted Maximum Likelihood Estimation for Gamma-Distributed Outcomes

Xinru Zhou, Wen Hou*

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: June 7, 2026; accepted: July 1, 2026; published: July 8, 2026

Abstract

Traditional causal inference methods, such as G-computation and inverse probability weighting, *通讯作者。

often rely on the correct specification of the outcome model or the propensity score model when estimating average treatment effects. Incorrect model specification may lead to biased estimates and a lack of robustness. Targeted maximum likelihood estimation (TMLE) improves double robustness and estimation efficiency by introducing a fluctuation parameter to update the initial outcome model in a targeted manner. Assuming a gamma distribution, this paper constructs the score functions for both the outcome regression model and the propensity score model, and obtains the estimated average treatment effect using TMLE. Monte Carlo simulation results show that TMLE remains robust across all model specification scenarios, maintaining unbiased estimates when either model is correctly specified, and estimation accuracy improves as the sample size increases. Compared with the standalone G-computation method, TMLE yields smaller bias and superior mean squared error under model misspecification. Applying the proposed method to the ACTG175 AIDS dataset, the estimation results indicate that combination therapy significantly increases CD4 cell counts, validating the robustness and practicality of the method.

Keywords

Causal Inference, Targeted Maximum Likelihood Estimation, Average Treatment Effect, Double Robustness

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在观察性研究的因果推断中, 存在多种常用估计方法, 其中目标最大似然估计(Targeted Maximum Likelihood Estimation, TMLE)作为一种半参数估计方法, 近年来已在生物医学、流行病学、公共卫生等领域得到广泛应用[1]。与此同时, G 计算(G-computation, G-comp)、逆概率加权(Inverse Probability of Treatment Weighting, IPTW)以及增强型逆概率加权(Augmented Inverse Probability of Treatment Weighting, AIPTW)等也是该领域的经典方法[2] [3], 这些方法共同构成了观察性研究因果推断的常用技术体系, 但各自的性能与适用场景存在差异。

具体而言, 传统常用方法均存在明显局限性: G-comp 高度依赖结果模型的正确设定, 一旦模型设定出现偏差, 将直接导致估计结果有偏[4], IPTW 的有效性依赖于倾向得分模型的合理性, 且极端权重的存在易造成估计量方差膨胀, 影响估计精度[5], AIPTW 虽同样具备双重稳健性, 但在数据稀疏或违背正则性假设的场景下, 易产生超出参数空间的估计值(如概率预测小于 0 或大于 1), 在有限样本中表现不够稳定[6]。这些局限性为观察性研究因果推断的方法选择带来了挑战, 也推动了更优估计方法的探索与应用。

相较于上述传统方法, TMLE 展现出显著优势, 其兼具双重稳健性和代入估计量的核心特点: 一方面, TMLE 无需两个模型同时正确设定, 只需倾向得分模型或结果模型之一设定合理, 即可获得一致估计[7], 另一方面, 作为代入估计量, 其估计值始终处于合理参数空间内, 有效规避了 AIPTW 可能出现的异常值问题[7]。此外, TMLE 还可与超级学习器(Super Learner)等机器学习算法结合, 进一步降低模型错误设定的风险[8], 在处理高维数据、复杂数据结构方面的优势已被多项实证研究证实——Luque-Fernandez 等将其用于癌症治疗效果评估[1], Pirracchio 等将其应用于危重病人预后预测[2], Decker 等则借助其评估青少年肥胖干预措施的效果[3]。然而, 现有研究对 TMLE 在伽马分布的不同模型错误设定场景下的表现尚缺乏系统性评估, 基于此, 本文通过蒙特卡洛模拟研究, 比较 TMLE、AIPTW、G-comp 和

Naive 估计四种方法在伽马分布场景下的估计性能, 重点考察在模型全部正确、倾向得分模型错误、结果模型错误三种设定下各方法的偏差与均方根误差, 以期为因果推断实践中方法的选择提供参考依据。

2. TMLE 介绍

目标最大似然估计是一种半参数估计方法, 其核心思想是在初始估计的基础上, 通过构造波动模型对初始结果模型进行靶向修正, 使得估计量满足有效影响函数方程, 从而兼具双重稳健性与渐近有效性。

2.1. 符号与假设

设 Y 为服从伽马分布的结局变量, $A = \{0, 1\}$ 为二元处理变量, 其中 $A=1$ 表示接受处理, $A=0$ 表示接受对照, W 为协变量向量。定义潜在结果 $Y(1)$ 和 $Y(0)$ 分别表示个体接受处理或对照时的潜在结局。平均处理效应(Average Treatment Effect, ATE)定义为 $\varphi = E[Y(1) - Y(0)]$ 。为了识别 ATE, 需要满足以下三个标准假设。

- (i) 可忽略性 $Y(0), Y(1) \perp A | W$, 即给定协变量 W 的条件下, 处理分配与潜在结果相互独立。
- (ii) 正值性 $0 < P(A=1|W) < 1$, 即每个个体都有非零的概率接受处理或对照。
- (iii) 一致性 $Y = AY(1) + (1-A)Y(0)$, 即观察到的结局等于其潜在结果。

2.2. 方法介绍

在上述假设下, TMLE 的估计过程遵循“初始估计→靶向修正→统计推断”的三步范式。首先, 根据结局变量为伽马分布特征, 拟合结果模型, 同时采用 logit 模型拟合倾向得分模型, 其次, 构造有利协变量并拟合波动模型进行靶向修正, 最后, 基于有效影响函数计算平均处理效应的点估计、标准误及置信区间。具体步骤如下。

步骤 1 初始模型的拟合

初始估计的目标是对结果模型(Q 模型)和倾向得分模型(g 模型)进行初始拟合。这两个模型虽不是研究者最终关心的目标参数, 但却是估计目标参数的必要条件。其中结果模型(Q 模型), 定义为给定处理变量和协变量时结果的条件期望为

$$Q(A, W) = E(Y | A, W)。 \quad (2.1)$$

对于结局变量服从伽马分布, 采用伽马广义线性模型来进行初始拟合。我们把初始预测记为 $Q^0(A, W)$ 。

倾向得分模型(g 模型)则用于估计给定协变量时, 个体接受处理的条件概率, 即

$$g(W) = P(A=1|W)。 \quad (2.2)$$

通常采用 logit 模型进行拟合, 估计值记为 $\hat{g}(W)$ 。需要说明的是, 初始模型并不局限于参数回归方法, 也可结合机器学习算法进行拟合, 以降低模型误设风险。

步骤 2 靶向修正

靶向修正是 TMLE 的核心步骤, 其目的是修正初始估计的偏差, 使其满足有效影响函数方程, 从而达到半参数效率界。即使初始模型存在偏倚, 靶向修正也能将其调整为渐近有效的估计量。

首先构造有利协变量。有利协变量是靶向修正的关键构造, 其作用是在更新步骤中引入倾向得分的信息, 以纠正初始结果回归的偏差。定义为

$$H_1(A, W) = \frac{A}{g(W)}, \quad H_0(A, W) = \frac{1-A}{1-g(W)}。 \quad (2.3)$$

其次, 以初始估计 $Q^0(A, W)$ 为中心, 沿着有利协变量的方向构造波动模型。当结局变量为伽马分布, 则波动模型为

$$Q_\varepsilon(A, W) = Q^0(A, W) + \exp(\varepsilon \cdot H(A, W)), \quad (2.4)$$

其中 ε 为波动参数, 控制修正的幅度, 模型中不包含截距项, $Q^0(A, W)$ 作为偏移量固定不变。

接着通过最大化似然函数来估计波动参数 $\hat{\varepsilon}$ 。对于伽马分布结局, 这一最大化过程可通过迭代重加权最小二乘来实现。波动参数 $\hat{\varepsilon}$ 反映了初始预测与数据之间的偏差程度, 若初始模型已正确, $\hat{\varepsilon}$ 应接近 0。

最后利用估计得到的 $\hat{\varepsilon}$, 对初始预测进行靶向更新, 为

$$Q^*(1, W) = Q^0(1, W) + \exp(\hat{\varepsilon} \cdot H_1(W)), \quad Q^*(0, W) = Q^0(0, W) + \exp(\hat{\varepsilon} \cdot H_0(W)). \quad (2.5)$$

这一更新步骤是 TMLE 的靶向所在, 使得更新后的预测能够更好地拟合数据, 同时保持对目标参数的低偏倚和低方差。

步骤 3 统计推断

完成靶向修正后, 需要对目标参数进行统计推断, 包括点估计、标准误和置信区间的计算

在点估计层面, 平均处理效应的 TMLE 估计量为

$$\hat{\phi}_{TMLE} = \frac{1}{n} \sum_{i=1}^n [Q^*(1, W_i) - Q^*(0, W_i)]. \quad (2.6)$$

即对所有个体分别预测其在处理组和对照组下的潜在结局, 取个体处理效应的平均值。这一估计量是相合的, 且具有渐近正态性。

在此基础上, TMLE 的标准误基于有效影响函数(Efficient Influence Function, EIF)的样本方差计算——这是其实现渐近有效性的核心保障。标准误的计算公式为

$$SE(\hat{\phi}_{TMLE}) = \sqrt{\frac{1}{n} \sum_{i=1}^n [IF(Q_i)]^2}, \quad (2.7)$$

其中, $IF(Q_i)$ 是第 i 个样本的有效影响函数估计值。对于 ATE, 其影响函数的具体形式为

$$IF(Q_i) = \frac{A_i - \hat{g}(W_i)}{\hat{g}(W_i)(1 - \hat{g}(W_i))} (Y_i - Q^*(A_i, W_i)) + Q^*(1, W_i) - Q^*(0, W_i) - \hat{\phi}_{TMLE}.$$

这一公式将每个样本的贡献分解为“偏差修正项”和“中心化项”, 保证了影响函数的均值为 0, 其方差能准确反映估计的不确定性。

基于上述点估计和标准误, 可构造目标参数的 $1 - \alpha$ 置信区间为

$$\hat{\phi}_{TMLE} \pm z_{\alpha/2} \times SE(\hat{\phi}_{TMLE}), \quad (2.8)$$

其中 $z_{\alpha/2}$ 是标准正态分布的上位数。

3. 模拟研究

3.1. 数据生成机制

本研究参考了 Luque-Fernandez 等(2018)针对二分类结局的 TMLE 教程中的数据生成框架, 并将其扩展至结局变量为伽马分布。原始框架包含四个协变量, 沿用该设定并进行了适当修改。

具体而言, 研究生成四个混杂变量 W_1, W_2, W_3, W_4 , 其中 $W_1 \sim B(1, 0.5)$, $W_2 \sim B(1, 0.65)$, W_3 在集合 $\{0, 1, 2, 3, 4\}$ 上服从离散均匀分布, W_4 在集合 $\{0, 1, 2, 3, 4, 5\}$ 上服从离散均匀分布。对于处理变量 A ($A=1$

时表示接受处理, $A=0$ 时表示未接受处理)由 logit 模型生成, 具体模型见式(3.1)。

$$P(A=1|W) = \frac{\exp(0.96 + 0.15W_1 + 0.12W_2 + 0.08W_3 + 0.10W_4 + 0.35W_2W_4)}{1 + \exp(0.96 + 0.15W_1 + 0.12W_2 + 0.08W_3 + 0.10W_4 + 0.35W_2W_4)}. \quad (3.1)$$

在给定处理变量 A 和协变量 W 的条件下, 结果变量 Y 服从伽马分布, 其条件均值 $\mu(A, W)$ 决定分布参数为

$$Y|A, W \sim \text{Gamma}(\nu, \theta), \quad \text{其中 } \nu = 10, \quad \theta = \frac{\mu(A, W)}{\nu},$$

这里 ν 是形状参数, θ 是尺度参数。对应的条件概率密度函数为

$$f_{Y|A, W}(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu(A, W)} \right)^\nu y^{\nu-1} \exp\left(-\frac{\nu y}{\mu(A, W)}\right), y > 0.$$

该参数化方式属于指数散布族, 满足方差与均值的平方成正比即 $\text{Var}(Y|A, W) = \frac{\mu(A, W)^2}{\nu}$ 。条件均值 $\mu(A, W)$ 通过对数链接函数定义为

$$\log(\mu(A, W)) = 0.6A - 0.10W_1 + 0.15W_2 + 0.06W_3 + 0.06W_4 + 0.35W_2W_4. \quad (3.2)$$

该真实平均处理效应通过大样本(5×10^6)模拟计算得到, 为 $\text{ATE} \approx 2.6$ 。

为评估各方法的估计性能并验证 TMLE 的双重稳健性, 主要考虑三种模型设定, 分别是结果模型和倾向得分模型均设定正确、仅倾向得分模型存在设定错误(遗漏交互项 W_2W_4), 以及仅结果模型存在设定错误(遗漏交互项 W_2W_4)。同时, 引入三种对比方法与 TMLE 共同开展模拟分析, 分别是 G-comp、AIPTW 和 Naive 估计。对于 G-comp, 其估计量为

$$\hat{\phi}_{G\text{-comp}} = \frac{1}{n} \sum_i^n [\hat{Q}(1, W_i) - \hat{Q}(0, W_i)],$$

其中 $\hat{Q}(a, W)$ 结果模型的预测值。该方法仅依赖于结果模型的正确设定, 若结果模型错误则估计会产生偏倚。将其纳入比较是为了验证 TMLE 在结果模型错误时的稳健性优势。

对于 AIPTW, 其估计量为

$$\hat{\phi}_{AIPTW} = \frac{1}{n} \sum_{i=1}^n \left[\frac{A_i Y_i}{\hat{g}(W_i)} - \frac{(1-A_i) Y_i}{1-\hat{g}(W_i)} - \frac{A_i - \hat{g}(W_i)}{\hat{g}(W_i)(1-\hat{g}(W_i))} \hat{Q}(A_i, W_i) \right].$$

该方法结合了倾向得分模型和结果模型, 同样具有双重稳健性。将其纳入比较是为了与 TMLE 进行对比, 考察两者在有限样本下的表现差异。

对于 Naive, 其估计量为

$$\hat{\phi}_{Naive} = \frac{1}{n_1} \sum_{i: A_i=1} Y_i - \frac{1}{n_0} \sum_{i: A_i=0} Y_i.$$

即直接比较处理组与对照组的均值差。该方法不控制任何混杂因素, 将其纳入比较作为基准, 用以说明不调整混杂会导致的偏倚程度。

为了衡量各方法的估计性能, 采用了三个评价指标, 估计均值(Mean), 其计算公式为 $\frac{1}{R} \sum_r \hat{\phi}_r$, 反映估计值的集中趋势。绝对偏差(AbsBias), 计算公式为 $|\hat{\phi} - \phi|$, 衡量估计值与真实值的偏离程度, 数值越小

表示估计越准确。均方根误差(RMSE), 计算公式为 $\sqrt{\frac{1}{R} \sum_r^R (\hat{\varphi}_r - \varphi)^2}$ 。综合衡量估计的偏差和方差, 数值越小表示估计性能越好。其中, φ 为真实平均处理效应, $\hat{\varphi}_r$ 为第 r 次模拟的估计值, $R=1000$ 为模拟次数。模拟过程中, 样本量分别取 $n=50, 100, 200, 500, 1000$ 。

3.2. 模拟结果

基于上述数据生成机制, 各方法的估计具体结果见表 1~3。

当结果模型和倾向得分模型均设定正确时, 各方法的估计结果见表 1。

Table 1. Both the outcome model and the propensity score model yielded correct results

表 1. 结果模型和倾向得分模型都正确情形

N	Method	Mean	AbsBias	RMSE
50	TMLE	2.5894	0.4099	0.5182
	Gcomp	2.5931	0.4100	0.5185
	AIPTW	2.5925	0.4510	0.5742
	Naive	4.5577	1.9939	2.2821
100	TMLE	2.5790	0.2835	0.3567
	Gcomp	2.5795	0.2830	0.3565
	AIPTW	2.5574	0.3320	0.4165
	Naive	4.5287	1.9293	2.1056
200	TMLE	2.6111	0.1978	0.2502
	Gcomp	2.6117	0.1972	0.2500
	AIPTW	2.6198	0.2354	0.2953
	Naive	4.5758	1.9736	2.0540
500	TMLE	2.5972	0.1339	0.1677
	Gcomp	2.5973	0.1339	0.1676
	AIPTW	2.5985	0.1533	0.1937
	Naive	4.5472	1.9448	1.9842
1000	TMLE	2.6036	0.0940	0.1170
	Gcomp	2.6037	0.0938	0.1167
	AIPTW	2.6040	0.1096	0.1368
	Naive	4.5654	1.9630	1.9831

当倾向得分模型被错误设定时, 各方法的估计结果见表 2。

Table 2. Error scenarios of the propensity score model

表 2. 倾向得分模型错误情形

N	Method	Mean	AbsBias	RMSE
50	TMLE	2.5848	0.4222	0.5329
	Gcomp	2.5876	0.4206	0.5321

续表

50	AIPTW	2.5925	0.4644	0.5883
	Naive	4.4914	1.9345	2.2269
	TMLE	2.6011	0.2806	0.3551
100	Gcomp	2.6011	0.2801	0.3545
	AIPTW	2.6139	0.3272	0.4183
	Naive	4.5402	1.9463	2.115
200	TMLE	2.6068	0.2048	0.2576
	Gcomp	2.6067	0.2044	0.2576
	AIPTW	2.6013	0.2356	0.2939
500	Naive	4.5467	1.9474	2.0378
	TMLE	2.5986	0.1316	0.1620
	Gcomp	2.5989	0.1318	0.1622
1000	AIPTW	2.5945	0.1552	0.1916
	Naive	4.5314	1.9321	1.9663
	TMLE	2.5988	0.0926	0.1167
	Gcomp	2.5990	0.0926	0.1166
	AIPTW	2.6007	0.1094	0.1374
	Naive	4.5529	1.9536	1.9717

当结果模型被错误设定时, 各方法的估计结果见表 3。

Table 3. Error scenarios in the result model

表 3. 结果模型错误情形

N	Method	Mean	AbsBias	RMSE
50	TMLE	2.6562	0.4636	0.5868
	Gcomp	2.7293	0.4988	0.6347
	AIPTW	2.7160	0.5119	0.6488
	Naive	4.5801	2.0147	2.3176
100	TMLE	2.6393	0.3052	0.3896
	Gcomp	2.7339	0.3444	0.4417
	AIPTW	2.6635	0.3617	0.4530
	Naive	4.5763	1.9785	2.1588
200	TMLE	2.6110	0.2145	0.2687
	Gcomp	2.7005	0.2455	0.3077
	AIPTW	2.6371	0.2478	0.3149
	Naive	4.5633	1.9626	2.0500
500	TMLE	2.6071	0.1351	0.1695
	Gcomp	2.7014	0.1691	0.2123

续表

	AIPTW	2.6133	0.1575	0.1989
	Naive	4.5444	1.9437	1.9818
1000	TMLE	2.6026	0.0953	0.1183
	Gcomp	2.6981	0.1327	0.1637
	AIPTW	2.6089	0.1166	0.1444
	Naive	4.5454	1.9447	1.9649

模拟结果表明, 当模型正确或倾向得分模型错误时, TMLE、Gcomp 与 AIPTW 均表现良好。当结果模型错误时, Gcomp 产生明显的正向系统偏倚, 而 TMLE 保持稳健, 体现了其双重稳健性。AIPTW 在结果模型错误时表现居中。Naive 估计在所有情景下均有严重偏倚, 不可使用。综上, 在伽马分布场景下, TMLE 是最为可靠的选择。

4. 实例分析

为进一步验证 TMLE 在真实数据中的适用性, 本研究选取了真实数据集开展实例分析, 该数据集来源于 R 语言中 speff2trial 包中的 ACTG175 临床试验, 该试验比较了核苷类单药治疗与联合治疗在 CD4 细胞计数为 200~500/mm³ 的 HIV 感染成人患者中的疗效, 共纳入 2139 名 HIV 感染者的临床资料, 由于 ACTG175 为随机对照试验, 治疗分配随机, 理论上处理组与对照组的协变量分布已基本平衡。此时采用 TMLE 等协变量调整方法, 其主要目的并非控制混杂, 而是提高估计精度, 即通过调整与结局相关的基线协变量, 减小估计量的方差。本研究将原始的四分类治疗方案变量(arms)重新编码为二分类处理变量 A, 其中联合治疗组(AZT + ddI 或 AZT + ddC)编码为 1, 单药治疗组(AZT 或 ddI)编码为 0。两组样本量达到较好平衡, 联合治疗组共计 1046 人, 单药治疗组共计 1093 人。具体变量定义与赋值见表 4。

Table 4. Variables in the ACTG175 AIDS clinical trial study

表 4. ACTG175 艾滋病临床试验研究的变量

变量名	定义	意义
cd420	随访时的 CD4 计数	结局变量 Y (单位: 个/mm ³)
arms	原始治疗方案分组	四分类变量(0, 1, 2, 3)
A	处理变量(联合治疗 vs 单药治疗)	1 = 联合治疗(AZT + ddI 或 AZT + ddC), 0 = 单药治疗(AZT 或 ddI)
age	年龄	协变量(患者年龄)
wtkg	体重	协变量(患者体重, 单位: kg)
karnof	Karnofsky 评分	协变量(身体状况评分 0~100)
preanti	曾接受抗病毒治疗情况	协变量(0 = 存在, 1 = 不存在)
oprior	曾接受机会性感染预防情况	协变量(0 = 存在, 1 = 不存在)
z30	基线 CD4 ≤ 30	协变量(0 = 是, 1 = 不是)
symptom	有症状情况	协变量(0 = 存在, 1 = 不存在)
offtrt	中断治疗情况	协变量(0 = 存在, 1 = 不存在)
gender	性别	协变量(0 = 女性, 1 = 男性)
race	种族	协变量(0 = 白人, 1 = 非白人)
cd40	基线 CD4 计数	协变量(治疗前的 CD4 水平)

为验证伽马回归模型的适用性, 我们对 cd420 变量进行了分布检验。图 1 显示 cd420 的直方图呈典型的右偏分布特征, 取值范围为正值, 红色实线(拟合的伽马密度曲线)与直方图高度吻合, 表明数据分布与理论伽马分布非常接近。同时描述性统计显示, cd420 的偏度为 0.730, 峰度为 4.066。采用极大似然法拟合伽马分布, 得到形状参数 $\alpha = 6.390$, 尺度参数 $\beta = 0.0172$ 。拟合优度检验中, Kolmogorov-Smirnov 检验的 p 值为 0.752 (不显著), 表明伽马分布是合适的分布假设。因此, cd420 近似服从伽马分布, 采用伽马广义线性模型是合理的。

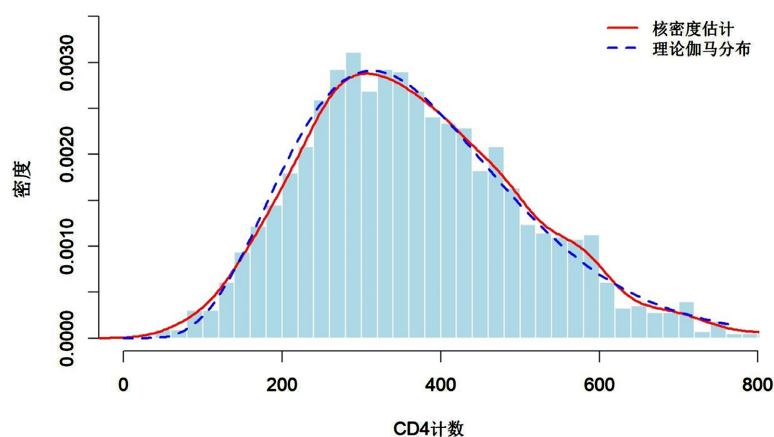


Figure 1. Gamma distribution test of cd420

图 1. cd420 的伽马分布检验

基于上述验证, 进一步采用 TMLE 方法估计联合治疗的效应, 在进行 TMLE 靶向修正之前, 首先对结果模型和倾向得分模型进行初始拟合。

由于结局变量 $Y = \text{cd420}$ 近似服从伽马分布, 结果模型按公式(2.1)采用伽马回归进行拟合, 即

$$\log(Q(A, W)) = \eta_0 + \eta_1 A + \eta_2 \text{age} + \eta_3 \text{wtkg} + \eta_4 \text{karnof} + \eta_5 \text{preanti} + \eta_6 \text{oprior} \\ + \eta_7 \text{z30} + \eta_8 \text{symptom} + \eta_9 \text{offtrt} + \eta_{10} \text{gender} + \eta_{11} \text{race} + \eta_{12} \text{cd40}.$$

估计具体结果见表 5。

Table 5. Parameter estimation of the outcome model

表 5. 结果模型参数估计

变量	η	SE	z 值	p 值
截距	5.1355	0.1276	40.24	<0.001
A	0.0934	0.0138	6.77	<0.001
age	-0.0003	0.0008	-0.32	0.749
wtkg	0.0000	0.0005	-0.07	0.942
karnof	0.0025	0.0012	2.07	0.038
preanti	0.0000	0.0000	-0.18	0.86
oprior	-0.2351	0.0475	-4.95	<0.001
z30	-0.1132	0.0186	-6.08	<0.001
symptom	-0.0542	0.0185	-2.93	0.003
offtrt	-0.098	0.0146	-6.72	<0.001

续表

gender	-0.0233	0.0196	-1.19	0.235
race	-0.0427	0.0161	-2.66	0.008
cd40	0.0018	0.0001	29.95	<0.001

从表 5 可以看出, 处理变量 A 的系数 0.0934 ($p < 0.001$), 表明联合治疗相比单药治疗对 CD4 计数有显著的正向影响。基线 CD4 计数 ($\alpha = 0.0018, p < 0.001$)、Karnofsky 评分 ($\alpha = 0.0025, p = 0.038$) 等协变量也显著影响结局。年龄、体重、性别等协变量未达到统计显著水平。

倾向得分模型按公式(2.2)采用 logit 回归进行拟合, 估计个体接受联合治疗的概率, 即

$$g(W) = \frac{\exp(\lambda_0 + \lambda_1 \text{age} + \lambda_2 \text{wtkg} + \lambda_3 \text{karnof} + \lambda_4 \text{preanti} + \lambda_5 \text{oprior} + \lambda_6 \text{z30} + \lambda_7 \text{symptom} + \lambda_8 \text{offtrt} + \lambda_9 \text{gender} + \lambda_{10} \text{race} + \lambda_{11} \text{cd40})}{1 + \exp(\lambda_0 + \lambda_1 \text{age} + \lambda_2 \text{wtkg} + \lambda_3 \text{karnof} + \lambda_4 \text{preanti} + \lambda_5 \text{oprior} + \lambda_6 \text{z30} + \lambda_7 \text{symptom} + \lambda_8 \text{offtrt} + \lambda_9 \text{gender} + \lambda_{10} \text{race} + \lambda_{11} \text{cd40})}$$

估计结果见表 6。

Table 6. Parameter estimation of the propensity score model

表 6. 倾向得分模型参数估计

变量	λ	SE	z 值	p 值
截距	-0.9545	0.8066	-1.18	0.237
age	0.0034	0.0051	0.67	0.506
wtkg	-0.0049	0.0034	-1.42	0.157
karnof	0.0117	0.0075	1.55	0.121
preanti	-0.0001	0.0001	-0.82	0.412
oprior	-0.0371	0.3001	-0.12	0.902
z30	0.1021	0.1175	0.87	0.385
symptom	0.0593	0.1167	0.51	0.611
offtrt	-0.0112	0.0920	-0.12	0.903
gender	0.0402	0.1241	0.32	0.746
race	-0.1260	0.1014	-1.24	0.214
cd40	0.0001	0.0004	0.17	0.867

表 6 可以看出, 由于 ACTG175 为随机对照试验, 治疗分配与基线协变量无关, 模型结果显示所有协变量均未达到统计显著水平 ($p > 0.05$), 这符合随机化数据的预期特征, 同时也验证了随机化的成功。

在初始模型的基础上进行靶向修正。首先由公式(2.3)构造有利协变量, 由于结局变量服从伽马分布, 构造波动模型如公式(2.4)。通过迭代重加权最小二乘法估计波动参数, 得到

$$\hat{\varepsilon}_1 = 0.00007, \quad \hat{\varepsilon}_0 = 0.00003。$$

两个波动参数均非常接近 0, 表明初始结果模型已具有较好的拟合效果, 靶向修正仅对预测值进行了微小调整。随后对初始预测进行靶向更新即公式(2.5)。

完成靶向修正后, 对目标参数进行统计推断。根据点估计公式(2.6), 计算得到平均处理效应的 TMLE 估计值为 34.82 个/mm³ (CD4 计数)。根据标准误公式(2.7), 基于有效影响函数计算标准误, 得到 $SE = 5.1242$, 根据置信区间公式(2.8), 构造 95% 置信区间为 [24.77, 44.86], $p < 0.001$, 对应的倍数变化为

1.098 (95% CI:[1.068,1.129]), 相对提升百分比为 9.8% (95% CI:[6.8%,12.9%])。作为对比, 仅基于初始结果模型 G-comp 的估计值为 34.80 个/mm³ (CD4 计数), 与 TMLE 估计值高度一致, 这表明靶向修正仅对初始估计进行了微小调整。

5. 结论

本文通过蒙特卡洛模拟和真实数据分析, 比较了 TMLE、G-comp、AIPTW 及 Naive 四种方法在估计平均处理效应时的表现。结果表明: TMLE 在所有设定下均表现最优。当结果模型错误设定时, G-Comp 偏差显著增大, AIPTW 鲁棒性不足, 唯有 TMLE 的估计性能与模型正确设定时基本一致, 偏差和 RMSE 均保持在较低水平。当仅倾向得分模型错误设定时, TMLE 同样保持稳定。随着样本量增大, TMLE 的估计精度持续提升, 验证了其大样本一致性。在真实数据应用中, 联合治疗使 CD4 细胞计数平均增加约 35 个/mm³, 结果稳健可靠。综上, TMLE 兼具双鲁棒性与有限样本下的稳定性能, 在模型设定可能存在不确定性的实际应用中表现最为可靠。建议将 TMLE 作为因果效应估计的首选方法。

但本研究存在以下局限性。模拟场景相对单一, 仅考虑低维连续结局变量, 未涉及高维协变量或生存数据等复杂场景; 初始模型依赖参数假设, 虽 TMLE 具有双重稳健性但严重错误设定仍可能影响有限样本性能; 伽马分布散度参数假定为固定值, 未探讨其估计不确定性对推断的影响。针对上述局限性, 未来可将 Super Learner 等集成学习方法引入 TMLE 框架以降低模型误设风险, 评估 TMLE 在高维协变量、时依混杂等复杂数据结构下的表现, 并探讨散度参数估计不确定性对标准误和置信区间覆盖率的影响, 采用二阶 TMLE 或自助法进行改进。

参考文献

- [1] Luque-Fernandez, M.A., Schomaker, M., Rachet, B. and Schnitzer, M.E. (2018) Targeted Maximum Likelihood Estimation for a Binary Treatment: A Tutorial. *Statistics in Medicine*, **37**, 2530-2546. <https://doi.org/10.1002/sim.7628>
- [2] Pirracchio, R., Petersen, M.L. and van der Laan, M. (2014) Improving Propensity Score Estimators' Robustness to Model Misspecification Using Super Learner. *American Journal of Epidemiology*, **181**, 108-119. <https://doi.org/10.1093/aje/kwu253>
- [3] Decker, A.L., Hubbard, A., Crespi, C.M., Seto, E.Y.W. and Wang, M.C. (2014) Semiparametric Estimation of the Impacts of Longitudinal Interventions on Adolescent Obesity Using Targeted Maximum-Likelihood: Accessible Estimation with the Ltmle Package. *Journal of Causal Inference*, **2**, 95-108. <https://doi.org/10.1515/jci-2013-0025>
- [4] Robins, J.M. (1986) A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect. *Mathematical Modelling*, **7**, 1393-1512. [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6)
- [5] Rosenbaum, P.R. and Rubin, D.B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**, 41-55. <https://doi.org/10.1093/biomet/70.1.41>
- [6] Snowden, J.M., Rose, S. and Mortimer, K.M. (2011) Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique. *American Journal of Epidemiology*, **173**, 731-738. <https://doi.org/10.1093/aje/kwq472>
- [7] Bang, H. and Robins, J.M. (2005) Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, **61**, 962-973. <https://doi.org/10.1111/j.1541-0420.2005.00377.x>
- [8] van der Laan, M.J. and Rose, S. (2011) Targeted Learning: Causal Inference for Observational and Experimental Data. Springer.