

# 新文速递

使用说明：

1. 新文速递markdown模板(md文件)可在如下三种软件上编辑：

- 1) Rstudio(2023.03以上)：通过ctex宏包(支持TeX)可由"render"生成pdf（保留可编辑tex文件）；也可通过
- 2) 在VScode中可进行可视化编写，并通过"Export to Pdf"由md转成pdf(带无用的格式)；也通过"Edit in VS
- 3) 在typora中编辑(可视化或源代码方法)

2. 对于熟悉TeX不熟悉markdown的作者，也允许他们直接按所提供的tex模板进行编辑。

3. For English author(s), they are required to provide: (1) background/motivation; (2) main results

YAML 代码块删除，可提供三种预览：

- HTML
- PDF

新文速递 | 具有二元结果的随机化临床试验中经协变量调整的无条件治疗效果的稳健方差估计

Robust Variance Estimation for Covariate-Adjusted Unconditional Treatment Effect in Randomized Clinical Trials with Binary Outcomes

引用: Ting Ye, Marlena Bannick, Yanyao Yi & Jun Shao (2023), Robust variance estimation for covariate-adjusted unconditional treatment effect in randomized clinical trials with binary outcomes, Statistical Theory and Related Fields, DOI: [10.1080/24754269.2023.2205802](https://doi.org/10.1080/24754269.2023.2205802)

作者:

1. Ting Ye, Department of Biostatistics, University of Washington, USA
2. Marlena Bannick, Global Statistical Sciences, Eli Lilly and Company, USA
3. Yanyao Yi, School of Statistics, East China Normal University, China
4. Jun Shao, Department of Statistics, University of Wisconsin, USA

#### 摘要:

为了提高具有二元结果的随机化临床试验中无条件治疗效应估计的精度和检验假设的功效, 研究人员和监管机构建议使用 g-计算作为协变量调整的可靠方法。然而, g-计算的实际应用由于缺乏可用于不同场合下感兴趣的无条件处理效果的稳健且可显式表示的方差估计公式而受到阻碍。为了填补这一空白, 我们为 g-计算估计量提供了显式且稳健的方差估计量, 并通过模拟证明了其在实际使用中的可行性。

#### 背景:

在随机化临床试验中, 调整基线协变量被认为是一种提高估计精度和检测治疗效果的方法。2021 年 5 月, 美国 FDA 发布了在随机化临床试验分析中使用协变量的指南草案, 并推荐 g-计算作为“统计上具有二元结果的无条件治疗效果协变量调整的可靠方法”。然而, (1) 目前没有可用于不同场合感兴趣无条件治疗效果推断的 g-计算估计量的稳健且显式表示的方差估计公式; (2) 已有文献 (如 Ge 等, 2011) 中关于风险差异和两个治疗组的公式, 它是基于模型的, 不适合作为模型稳健性推理范式; (3) Ge 等 (2011) 的公式没有考虑到 logistic 模型的中由于协变量和非线性而导致的变异性来源, 这可能导致置信区间的覆盖概率偏低的情况出现。

考虑有  $n$  个受试者参与的  $k$  臂试验 ( $n$  个受试者被随机分配到  $k$  个实验组)。令  $A_i$  是  $k$  维处理示性向量, 其取值  $a_t$  表示第  $i$  个病人接受处理  $t$  ( $t = 1, \dots, k$ ), 其第  $t$  个元素为 1, 其余元素为 0.  $Y_i(t), i = 1, 2, \dots, n, t = 1, 2, \dots, k$  是在处理  $t$  下的二元潜在结果,  $X_i$  是用于调整的基线协变量向量。令  $\theta_t = E(Y^{(t)}), \theta = (\theta_1, \dots, \theta_k)^\top$  (其中省略了个体下标  $i$ )。

论文考虑的模型为  $E(Y | A, X) = \text{expit}(\beta_A^\top A + \beta_X^\top X)$  (称为 logistic 工作模型)。  $\hat{\mu}_t(X_i) = \text{expit}(\hat{\beta}_A^\top a_t + \hat{\beta}_X^\top X_i)$  为响应变量  $Y$  在处理  $t$  下的预测概率, 其中  $\hat{\beta}_A$  和  $\hat{\beta}_X$  分别为  $\beta_A$  和  $\beta_X$  的极大似然估计。则  $\theta$  的 g-计算估计为  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)^\top$ , 其中  $\hat{\theta}_t = n^{-1} \sum_{i=1}^n \hat{\mu}_t(X_i)$ , 目标参数函数  $f(\theta)$  的 g-计算估计为  $\hat{f}(\theta)$ , 其中  $f(\theta)$  的常用形式有

$$f(\theta) = \begin{cases} \theta_t - \theta_s, & (\text{风险差}) \\ \log \frac{\theta_t}{\theta_s} & (\text{风险比}) \\ \log \frac{\theta_t/(1-\theta_t)}{\theta_s/(1-\theta_s)} & (\text{优势比}) \end{cases}$$

论文的目的是要得到  $\hat{f}(\theta)$  的方差的估计.

主要结果:

### 1. 理论结果

$\hat{f}(\theta)$  的一致稳健的方差估计为

$$n^{-1}\{\nabla f(\hat{\theta})\}^T \hat{V} \{\nabla f(\hat{\theta})\},$$

其中

$$\hat{V} = \begin{pmatrix} \hat{v}_{11} & \hat{v}_{12} & \cdots & \hat{v}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{v}_{1k} & \hat{v}_{2k} & \cdots & \hat{v}_{kk} \end{pmatrix},$$

$$\begin{aligned} \hat{v}_{tt} &= \pi_t^{-1} S_{rt}^2 + 2Q_{ytt} - S_{\mu t}^2, \quad t = 1, \dots, k, \\ \hat{v}_{ts} &= Q_{yts} + Q_{yst} - Q_{\mu ts}, \quad 1 \leq t < s \leq k, \end{aligned}$$

记号  $S_{rt}^2$ ,  $Q_{ytt}$ ,  $S_{\mu t}^2$ ,  $Q_{yts}$ ,  $Q_{yst}$  和  $Q_{\mu ts}$  为一些样本方差 (具体见原文).

### 2. 模拟结果

考虑三种情形:

- 情形 I:  $P(Y = 1 \mid A, X) = \text{expit} \{-2 + 5I(A = a_2) + X\}$ , 相应的  $(\theta_1, \theta_2) = (0.2830, 0.8057)$ ;
- 情形 II:  $P(Y = 1 \mid A = a_1, X) = \text{expit}(-2 + X)$ ,  $P(Y = 1 \mid A = a_2, X) = \text{expit}(3 + 1.5X - 0.01X^2)$ , 相应的  $(\theta_1, \theta_2) = (0.2830, 0.7297)$ ;
- 情形 III:  $P(Y = 1 \mid A, X) = \text{expit}(-2 + 2I(A = a_2) + 4I(A = a_3) + X)$ , 相应的  $(\theta_1, \theta_2, \theta_3) = (0.2827, 0.5004, 0.7172)$ .

表 1: 在情况 I-II 和简单随机化下,  $\hat{\theta}_2 - \hat{\theta}_1$  的模拟平均值和标准差 (SD), 平均标准误差 (SE), 和  $\theta_2 - \theta_1$  的渐近置信区间的 95% 覆盖概率 (CP).

Case	$\theta_2 - \theta_1$	$n$	$\hat{\theta}_2 - \hat{\theta}_1$		Robust SE in (1)		SE in Geet al. (2011)	
			Mean	SD	SE	CP (%)	SE	CP (%)
I	0.5227	200	0.5228	0.0464	0.0464	94.44	0.0415	91.27
		500	0.5227	0.0295	0.0294	94.70	0.0264	91.94
II	0.4467	200	0.4469	0.0457	0.0458	94.56	0.0404	91.08
		500	0.4463	0.0289	0.0290	94.90	0.0257	91.77

表 2: 在情况 III 和简单随机化下, g-计算估计量的均值和标准差 (SD), 平均标准误差 (SE), 和基于稳健 SE(1) 的渐近置信区间的 95% 覆盖概率 (CP).

Parameter	Truth	$n = 200$				$n = 500$			
		Mean	SD	SE	CP	Mean	SD	SE	CP
$\theta_2 - \theta_1$	0.2177	0.2176	0.0578	0.0573	94.34	0.2170	0.0366	0.0363	94.82
$\log(\theta_2/\theta_1)$	0.5711	0.5798	0.1701	0.1664	94.50	0.5726	0.1053	0.1042	94.59
$\log \frac{\theta_2/(1-\theta_2)}{\theta_1/(1-\theta_1)}$	0.9328	0.9440	0.2620	0.2586	94.63	0.9341	0.1637	0.1624	94.79
$\theta_3 - \theta_1$	0.4346	0.4348	0.0581	0.0568	94.15	0.4347	0.0360	0.0360	94.84
$\log(\theta_3/\theta_1)$	0.9311	0.9432	0.1653	0.1611	94.43	0.9353	0.1018	0.1009	94.92
$\log \frac{\theta_3/(1-\theta_3)}{\theta_1/(1-\theta_1)}$	1.8621	1.8852	0.2920	0.2851	94.57	1.8712	0.1791	0.1788	95.01