*Statistical Science*
2017, Vol. 32, No. 3, 331–345
DOI: 10.1214/16-STS571


Chapter 8 Unification

# A Paradox from Randomization-Based Causal Inference[1]

**Peng Ding**

*Abstract.* Under the potential outcomes framework, causal effects are defined as comparisons between potential outcomes under treatment and control. To infer causal effects from randomized experiments, Neyman proposed to test the null hypothesis of zero average causal effect (Neyman's null), and Fisher proposed to test the null hypothesis of zero individual causal effect (Fisher's null). Although the subtle difference between Neyman's null and Fisher's null has caused a lot of controversies and confusions for both theoretical and practical statisticians, a careful comparison between the two approaches has been lacking in the literature for more than eighty years. We fill this historical gap by making a theoretical comparison between them and highlighting an intriguing paradox that has not been recognized by previous researchers. Logically, Fisher's null implies Neyman's null. It is therefore surprising that, in actual completely randomized experiments, rejection of Neyman's null does not imply rejection of Fisher's null for many realistic situations, including the case with constant causal effect. Furthermore, we show that this paradox also exists in other commonly-used experiments, such as stratified experiments, matched-pair experiments and factorial experiments. Asymptotic analyses, numerical examples and real data examples all support this surprising phenomenon. Besides its historical and theoretical importance, this paradox also leads to useful practical implications for modern researchers.

$H_{0N}: \tau = 0$

$H_{0F}: \tau_i = 0 \ \forall i$

*Key words and phrases:* Average null hypothesis, Fisher randomization test, potential outcome, randomized experiment, repeated sampling property, sharp null hypothesis.

**Commentary** — Statistics in Medicine

(wileyonlinelibrary.com) DOI: 10.1002/sim.6764 — Published online in Wiley Online Library

$H_{0F}: \tau_i = 0 \ \forall i$

p-value
exact

# Exact confidence intervals for the average causal effect on a binary outcome

**Xinran Li[a] and Peng Ding[b*†]**

$\hat{\tau} \pm 1.96\sqrt{\hat{V}}$

$n \to \infty$

CLT

**Based on the physical randomization of completely randomized experiments, in a recent article in *Statistics in Medicine*, Rigdon and Hudgens propose two approaches to obtaining exact confidence intervals for the average causal effect on a binary outcome. They construct the first confidence interval by combining, with the Bonferroni adjustment, the prediction sets for treatment effects among treatment and control groups, and the second one by inverting a series of randomization tests. With sample size $n$, their second approach requires performing $O(n^4)$ randomization tests. We demonstrate that the physical randomization also justifies other ways to constructing exact confidence intervals that are more computationally efficient. By exploiting recent advances in hypergeometric confidence intervals and the stochastic order information of randomization tests, we propose approaches that either do not need to invoke Monte Carlo or require performing at most $O(n^2)$ randomization tests. We provide technical details and R code in the Supporting Information.** 

精确 $\Pr(\tau \in [L, U]) \geq 1-\alpha$

$\forall n$

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

*J. R. Statist. Soc.* B (2016)
**78**, *Part* 3, *pp.* 655–671

Rosenbaum

$H_0(c)$: $\tau_i = c$ $\forall i$

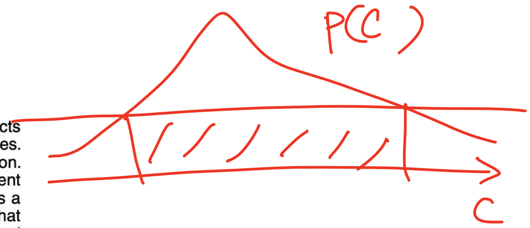
constant

# Randomization inference for treatment effect variation

Peng Ding, Avi Feller and Luke Miratrix
*Harvard University, Cambridge, USA*

[Received May 2014. Revised March 2015]

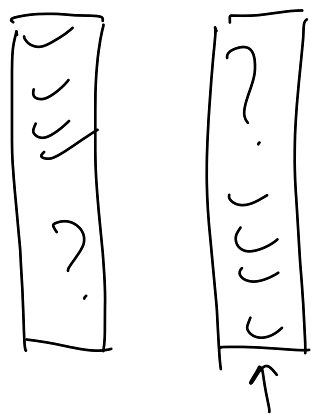
P(C)




**Summary.** Applied researchers are increasingly interested in whether and how treatment effects vary in randomized evaluations, especially variation that is not explained by observed covariates. We propose a model-free approach for testing for the presence of such unexplained variation. To use this randomization-based approach, we must address the fact that the average treatment effect, which is generally the object of interest in randomized experiments, actually acts as a nuisance parameter in this setting. We explore potential solutions and advocate for a method that guarantees valid tests in finite samples despite this nuisance. We also show how this method readily extends to testing for heterogeneity beyond a given model, which can be useful for assessing the sufficiency of a given scientific theory. We finally apply our method to the National Head Start impact study, which is a large-scale randomized evaluation of a Federal preschool programme, finding that there is indeed significant unexplained treatment effect variation.

$H_0$: $\exists c$, s.t. $\tau_i = c$ $\forall i$

$Y_i(1) - Y_i(0) = c$ 常数作用



置换

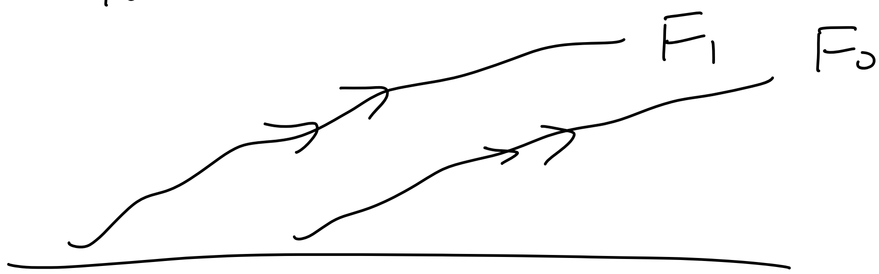
边缘分布不变，联合分布变

无法验证

可以证伪 falsify

如果 $Y_i(1) = Y_i(0) + C$

那么




反之，若 $\hat{F}_1(y) \neq \hat{F}_0(y)$

那么非常数作用.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2017, VOL. 112, NO. 520, 1759–1769, Review
https://doi.org/10.1080/01621459.2017.1295865



# General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference

Xinran Li[a] and Peng Ding[b]

[a]Department of Statistics, Harvard University, Cambridge, MA; [b]Department of Statistics, University of California, Berkeley, CA

**ABSTRACT**

Frequentists' inference often delivers point estimators associated with confidence intervals or sets for parameters of interest. Constructing the confidence intervals or sets requires understanding the sampling distributions of the point estimators, which, in many but not all cases, are related to asymptotic Normal distributions ensured by central limit theorems. Although previous literature has established various forms of central limit theorems for statistical inference in super population models, we still need general and convenient forms of central limit theorems for some randomization-based causal analyses of experimental data, where the parameters of interests are functions of a finite population and randomness comes solely from the treatment assignment. We use central limit theorems for sample surveys and rank statistics to establish general forms of the finite population central limit theorems that are particularly useful for proving asymptotic distributions of randomization tests under the sharp null hypothesis of zero individual causal effects, and for obtaining the asymptotic repeated sampling distributions of the causal effect estimators. The new central limit theorems hold for general experimental designs with multiple treatment levels, multiple treatment factors and vector outcomes, and are immediately applicable for studying the asymptotic properties of many methods in causal inference, including instrumental variable, regression adjustment, rerandomization, cluster-randomized experiments, and so on. Previously, the asymptotic properties of these problems are often based on heuristic arguments, which in fact rely on general forms of finite population central limit theorems that have not been established before. Our new theorems fill this gap by providing more solid theoretical foundation for asymptotic randomization-based causal inference. Supplementary materials for this article are available online.

**ARTICLE HISTORY**
Received June 2016
Revised January 2017

**KEYWORDS**
Conservative confidence set; Fisher randomization test; Potential outcome; Randomization inference; Repeated sampling property; Sharp null hypothesis

SRE = K个独立 CRE

离散 $X \in \{1, 2 \cdots K\}$

随机化: $Z_i \perp\!\!\!\perp (Y_i(1), Y_i(0))$

分层随机化: $Z_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) \mid X_i$

后面在观察性研究中核心假设

无混杂假设

可忽略性假设

why SRE?

① 避免"坏的"随机化

② 提高估计精度

估计 $\tau$

比较 CRE 和 SRE ($X \in \{1 \cdots K\}$)

$$\begin{cases} \text{CRE}: & \dfrac{n_1}{n} \text{ 处理比率 } \hat{\tau} \\ \text{SRE}: & \dfrac{n_{[k]1}}{n_{[k]}} \text{ 处理比率} \\ & = \dfrac{n_1}{n} \text{ 相同 } \hat{\tau}_S \end{cases}$$

$\Rightarrow$ $\hat{\tau} = \hat{\tau}_S$ 数值上一样

比较 $\mathrm{var}_{CRE}(\hat{\tau})$ 和 $\mathrm{var}_{SRE}(\hat{\tau}_S)$

$$\mathrm{var}_{CRE}(\hat{\tau}) = \underbrace{\frac{S^2(1)}{n_1} + \frac{S^2(0)}{n_0} - \frac{S^2(\tau)}{n}}_{\text{依赖总方差}}$$

依赖总方差 = 组内方差 + 组间方差 ANOVA

$$\mathrm{var}_{SRE}(\hat{\tau}_S) = \sum_{k=1}^{K} \pi_{[k]}^2 \underbrace{\mathrm{var}(\hat{\tau}_{[k]})}$$

依赖于 $\left\{ \begin{array}{l} S^2_{[k]}(1) \\ S^2_{[k]}(0) \\ S^2_{[k]}(\tau) \end{array} \right.$ 组内方差

结论: $\mathrm{var}_{CRE}(\hat{\tau}) \geq \mathrm{var}_{SRE}(\hat{\tau}_S)$

$\underbrace{\phantom{\geq}}_{\text{近似}}$

$n_{[k]1}, n_{[k]0} \to \infty$

分层可以提高精度

走向极端可能损害精度:

MPE $\quad n_{[k]1} = 1, \; n_{[k]0} = 1$

matched-pairs experiment

实际问题:

1. 如何选 $K$?

$K$ 越大越好 (理想)

如果 $n_{[k]} \to \infty$

$X_1 \in \{1, 0\}$

$X_2 \in \{1, 0\}$

$\vdots$

$x_p \in \{1, 0\}$

---

$2^p$ 层

取决于 $X \sim Y$ 相关性

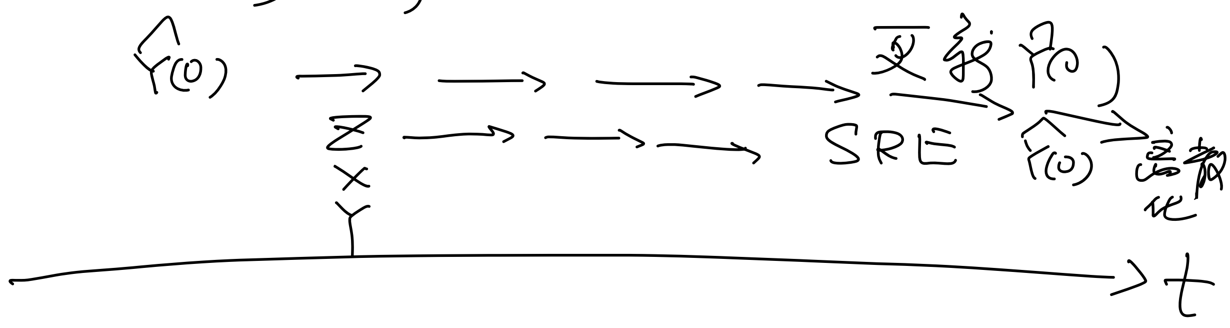
2. 若有一部分 $(X, Y)$ 数据

那么 ① $Y \sim X$ 变量选择

② $Y \sim X \Rightarrow \hat{Y}(0)$

离散化 $\hat{Y}(0)$，用作分层

③ 序贯实验



3. 连续或混合 $X$

Chapter 6 Rerandomization

CRE: $(Z_i, X_i)_{i=1}^{n}$

$$\hat{\tau}_X = \frac{1}{n_1}\sum_{i=1}^{n} Z_i X_i - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i) X_i$$

期望 0

实现值 $\neq 0$

避免 $\hat{\tau}_X$ "大"

$$M = \hat{\tau}_X' \left[\operatorname{cov}(\hat{\tau}_X)\right]^{-1} \hat{\tau}_X$$

$$\approx \chi^2_K, \quad \dim(X) = K$$

协变量调整　CRE

$(X_i, Z_i, Y_i)_{i=1}^n$

目标：用协变量 $X_i$ 提高估计精度

如：$lm(Y_i \sim 1 + Z_i + X_i)$

$coef(Z_i) = \hat{\tau}_F$

Fisher's ANCOVA

协方差分析

David Freedman (2008) 批评 $\hat{\tau}_F$：

0. $(Y_i(1), Y_i(0), X_i)_{i=1}^n$ 固定

$Z_i$'s 随机 $\sim$ CRE

1. $\hat{\tau}$ 无偏

$\hat{\tau}_F$ 有偏

2. $var_a(\hat{\tau}_F)$ 可能比

$var(\hat{\tau})$ 大

3. lm $\Longrightarrow$ $\hat{se}^2_F$ 可能低估 (假定同方差)

$var_a(\hat{\tau}_F)$.

Lin (2013) 博士论文

1. $\hat{\tau}_F$ 有偏，但 $\hat{\tau}_F \xrightarrow{P} \tau$

相合

2. $lm(Y_i \sim 1 + Z_i + X_i + Z_i \cdot X_i)$

交互项

且 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 0$

为了方便

$coef(Z_i) = \hat{\tau}_L$

$\hat{\tau}_L \xrightarrow{P} \tau$ 相合

且 $var_a(\hat{\tau}_L) \le var(\hat{\tau})$

$\le var_a(\hat{\tau}_F)$

直观？ $\left.\begin{array}{l} Y_i(1) \sim X_i \\ Y_i(0) \sim X_i \end{array}\right\}$ 不平行

不直观？ 参数越多，估计越不精确

MLE $f(y;\ \alpha, \beta)$

$\alpha$ 关心

$\beta$ 讨厌参数 nuisance

$\hat{\alpha}, \hat{\beta}$

$$cov_a\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\beta\alpha} & I_{\beta\beta} \end{pmatrix}^{-1}$$

$\beta = 0$ : 只估计 $\alpha$ : $\tilde{\alpha}$

$$var_a(\tilde{\alpha}) = I_{\alpha\alpha}^{-1}$$

EHW 方差估计：保守

3. $\hat{\tau}_F$：lm( $Y_i \sim 1 + Z_i + X_i$ )

EHW：保守

为什么 $\hat{\tau}_L$？

答：四种写法

其中一种解释：$\bar{X} = 0$

$$\hat{\tau}(\beta_1, \beta_0) = \frac{1}{n_1}\sum_{i=1}^{n} Z_i\left(Y_i - \beta_1' X_i\right) - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)\left(Y_i - \beta_0' X_i\right)$$

$$= \left(\hat{\bar{Y}}(1) - \beta_1' \hat{\bar{X}}(1)\right) - \left(\hat{\bar{Y}}(0) - \beta_0' \hat{\bar{X}}(0)\right)$$

无偏

方差 $\mathrm{var}\left(\hat{\tau}(\beta_1, \beta_0)\right)$:

潜在结果 $\left(Y_i(1) - \beta_1' X_i ,\ Y_i(0) - \beta_0' X_i\right)_{i=1}^{n}$

Neyman 方差估计:

$$\hat{V}(\beta_1, \beta_0) = \frac{\hat{S}^2(1, \beta_1)}{n_1} + \frac{\hat{S}^2(0, \beta_0)}{n_0}$$

其中
$$\hat{S}^2(1, \beta_1) = \frac{1}{n_1 - 1} \sum_{i=1}^{n} Z_i \left(Y_i - \gamma_1 - \beta_1' X_i\right)^2$$
$$\hat{S}^2(0, \beta_0) = \frac{1}{n_0 - 1} \sum_{i=1}^{n} (1 - Z_i) \left(Y_i - \gamma_0 - \beta_0' X_i\right)^2$$

优化估计精度:

$$\left\{\begin{array}{l} \min\limits_{\gamma_1, \beta_1} \sum\limits_{i=1}^{n} Z_i \left(Y_i - \gamma_1 - \beta_1' X_i\right)^2 \\ \min\limits_{\gamma_0, \beta_0} \sum\limits_{i=1}^{n} (1 - Z_i) \left(Y_i - \gamma_0 - \beta_0' X_i\right)^2 \end{array}\right\} \text{OLS}$$

$$\Rightarrow (\hat{\gamma}_1, \hat{\beta}_1),\ (\hat{\gamma}_0, \hat{\beta}_0)$$

来自 $Y_i \sim X_i$ OLS

处理　　对照

$\Rightarrow$ 估计量 $\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$

$$= \left(\hat{\bar{Y}}(1) - \hat{\beta}_1' \hat{\bar{X}}(1)\right) - \left(\hat{\bar{Y}}(0) - \hat{\beta}_0' \hat{\bar{X}}(0)\right)$$

$$= \hat{\gamma}_1 - \hat{\gamma}_0$$

因为 OLS 保证: $\begin{cases} \hat{\bar{Y}}(1) = \hat{\gamma}_1 + \hat{\beta}_1' \hat{\bar{X}}(1) \\ \hat{\bar{Y}}(0) = \hat{\gamma}_0 + \hat{\beta}_0' \hat{\bar{X}}(0) \end{cases}$

$$\Rightarrow \hat{\tau}_L = \hat{\tau}(\hat{\beta}_1, \hat{\beta}_0)$$

数学漏洞: $\hat{\beta}_1, \hat{\beta}_0$ 随机了

$$\hat{\tau}(\hat{\beta}_1, \hat{\beta}_0) - \hat{\tau}(\tilde{\beta}_1, \tilde{\beta}_0) = N,$$

随机　　固定（极限）

实验的解释：因果推断
= 缺失数据

结论：$\hat{\tau}_L$ = OLS 填补缺失的潜在结果

总结：CRE

$\hat{\tau}_N$: lm($Y_i \sim Z_i$)

$\hat{\tau}_F$: lm($Y_i \sim Z_i + X_i$)

$\hat{\tau}_L$: lm($Y_i \sim Z_i + X_i + Z_i \cdot X_i$)

EHW s.e. 保守的

$\hat{\tau}_L \xrightarrow{P} \tau$

不要求线性模型时　方差估计

SRE = K 个独立 CRE

⇒ 每层作协方差分析