Solutions to Exercises of Chapter 3

September 25, 2025
Ting-Chih Hung
nuit0815@gmail.com

Problem 3.1 (Exactness of p_{FRT})

We are asked to show that the p-value

$$p_{\text{FRT}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}(T(\mathbf{z}^m, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))$$

is finite-sample exact under the sharp null hypothesis; i.e., for any $u \in [0, 1]$,

$$P(p_{\text{FRT}} \leq u) \leq u.$$

In a completely randomized experiment (CRE), $\mathbf{Z}$ is uniformly distributed over the set $\{\mathbf{z}^1, \dots, \mathbf{z}^M\}$, where $M = \binom{n}{n_1}$ is the total number of possible treatment assignments. That is,

$$P(\mathbf{Z} = \mathbf{z}^m) = \frac{1}{M}, \quad m = 1, \dots, M.$$

Under the sharp null hypothesis, $Y_i(1) = Y_i(0)$ for all $i = 1, \dots, n$, so the observed outcome $\mathbf{Y}$ is invariant to the treatment assignment $\mathbf{Z}$. Therefore, the test statistic $T(\mathbf{Z}, \mathbf{Y})$ is also uniformly distributed over the set $\{T(\mathbf{z}^1, \mathbf{Y}), \dots, T(\mathbf{z}^M, \mathbf{Y})\}$, and the p-value p_{FRT} is uniformly distributed over the set $\{1/M, 2/M, \dots, 1\}$. This implies that for any $k \in \{1, \dots, M\}$,

$$P\left(p_{\text{FRT}} \leq \frac{k}{M}\right) = \frac{k}{M}.$$

For any $u \in [0, 1]$, let $k = \lfloor Mu \rfloor$ be the largest integer less than or equal to Mu. Then,

$$P(p_{\text{FRT}} \leq u) = P\left(p_{\text{FRT}} \leq \frac{k}{M}\right) = \frac{k}{M} \leq u,$$

which completes the proof.

Problem 3.2 (Monte Carlo error of $\hat{p}_{\text{FRT}}$)

In practice, we usually approximate p_{FRT} by

$$\hat{p}_{\text{FRT}} = \frac{1}{R} \sum_{r=1}^{R} \mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y})),$$

where $\mathbf{z}^1, \dots, \mathbf{z}^R$ are R independent draws from the uniform distribution over the set $\{\mathbf{z}^1, \dots, \mathbf{z}^M\}$.

We are asked to show that given the observed data $(\mathbf{Z}, \mathbf{Y})$,

$$\text{E}_{\text{mc}}(\hat{p}_{\text{FRT}}) = p_{\text{FRT}},$$

and

$$\mathrm{Var}_{\mathrm{mc}}\left(\widehat{p}_{\mathrm{FRT}}\right) \leq \frac{1}{4R},$$

where E_{mc} and $\mathrm{Var}_{\mathrm{mc}}$ denote the expectation and variance with respect to the Monte Carlo draws $\mathbf{z}^1, \dots, \mathbf{z}^R$, given the observed data $(\mathbf{Z}, \mathbf{Y})$.

For any $r \in \{1, \dots, R\}$, $\mathbf{z}^r$ is drawn from the uniform distribution over the set $\{\mathbf{z}^1, \dots, \mathbf{z}^M\}$, so we have

$$\begin{aligned}
\mathrm{E}_{\mathrm{mc}}\left(\mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right) &= P_{\mathrm{mc}}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y})) \\
&= \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}(T(\mathbf{z}^m, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y})) \\
&= p_{\mathrm{FRT}}.
\end{aligned}$$

Applying the linearity of expectation, we have

$$\begin{aligned}
&\mathrm{E}_{\mathrm{mc}}\left(\widehat{p}_{\mathrm{FRT}}\right) \\
&= \mathrm{E}_{\mathrm{mc}}\left(\frac{1}{R} \sum_{r=1}^{R} \mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right) \\
&= \frac{1}{R} \sum_{r=1}^{R} \mathrm{E}_{\mathrm{mc}}\left(\mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right) \\
&= p_{\mathrm{FRT}},
\end{aligned}$$

which proves the first part.

For the second part, notice that for any r,

$$\begin{aligned}
&\mathrm{Var}_{\mathrm{mc}}\left(\mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right) \\
&= \mathrm{E}_{\mathrm{mc}}\left(\mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))^2\right) - \left(\mathrm{E}_{\mathrm{mc}}\left(\mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right)\right)^2 \\
&= p_{\mathrm{FRT}} - p_{\mathrm{FRT}}^2 \\
&\leq \frac{1}{4}.
\end{aligned}$$

Since $\mathbf{z}^1, \dots, \mathbf{z}^R$ are independent,

$$\begin{aligned}
&\mathrm{Var}_{\mathrm{mc}}\left(\widehat{p}_{\mathrm{FRT}}\right) \\
&= \mathrm{Var}_{\mathrm{mc}}\left(\frac{1}{R} \sum_{r=1}^{R} \mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right) \\
&= \frac{1}{R^2} \sum_{r=1}^{R} \mathrm{Var}_{\mathrm{mc}}\left(\mathbf{1}(T(\mathbf{z}^r, \mathbf{Y}) \geq T(\mathbf{Z}, \mathbf{Y}))\right) \\
&\leq \frac{1}{R^2} \sum_{r=1}^{R} \frac{1}{4} \\
&= \frac{1}{4R},
\end{aligned}$$

which proves the second part.

Problem 3.3 (A finite-sample valid Monte Carlo approximation of p_{FRT})

Problem 3.4 (Fisher's exact test)

Problem 3.5 (More details for lady tasting tea)

Problem 3.6 (Covariate-adjusted FRT)

Problem 3.8 (An algebraic detail)

We are asked to show that

$$(n-1)s^2 = \sum_{Z_i=1} \left(Y_i - \hat{\bar{Y}}(1)\right)^2 + \sum_{Z_i=0} \left(Y_i - \hat{\bar{Y}}(0)\right)^2 + \frac{n_1 n_0}{n}\hat{\tau}^2. \tag{3.7}$$

First, notice that

$$\bar{Y} \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\left(\sum_{Z_i=1} Y_i + \sum_{Z_i=0} Y_i\right) = \frac{1}{n}\left(n_1\hat{\bar{Y}}(1) + n_0\hat{\bar{Y}}(0)\right),$$

which implies

$$\bar{Y} = \hat{\bar{Y}}(1) - \frac{n_0}{n}\underbrace{\left(\hat{\bar{Y}}(1) - \hat{\bar{Y}}(0)\right)}_{\hat{\tau}} = \hat{\bar{Y}}(0) + \frac{n_1}{n}\underbrace{\left(\hat{\bar{Y}}(1) - \hat{\bar{Y}}(0)\right)}_{\hat{\tau}}. \tag{1}$$

Therefore, we have

$$\begin{aligned}
&(n-1)s^2 \\
\equiv\ & \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 && \text{(Definition)} \\
=\ & \sum_{Z_i=1}\left(Y_i - \bar{Y}\right)^2 + \sum_{Z_i=0}\left(Y_i - \bar{Y}\right)^2 && \text{(Partition by } Z_i\text{)} \\
=\ & \sum_{Z_i=1}\left(Y_i - \hat{\bar{Y}}(1) + \hat{\bar{Y}}(1) - \bar{Y}\right)^2 + \sum_{Z_i=0}\left(Y_i - \hat{\bar{Y}}(0) + \hat{\bar{Y}}(0) - \bar{Y}\right)^2 && \text{(Add and subtract)} \\
=\ & \sum_{Z_i=1}\left(Y_i - \hat{\bar{Y}}(1) + \frac{n_0}{n}\hat{\tau}\right)^2 + \sum_{Z_i=0}\left(Y_i - \hat{\bar{Y}}(0) - \frac{n_1}{n}\hat{\tau}\right)^2 && \text{(Use Equation (1))} \\
=\ & \sum_{Z_i=1}\left(Y_i - \hat{\bar{Y}}(1)\right)^2 + \frac{n_0^2 n_1}{n^2}\hat{\tau}^2 + \sum_{Z_i=0}\left(Y_i - \hat{\bar{Y}}(0)\right)^2 + \frac{n_1^2 n_0}{n^2}\hat{\tau}^2 && \Big(\sum_{Z_i=z}(Y_i - \hat{\bar{Y}}(z)) = 0\Big) \\
=\ & \sum_{Z_i=1}\left(Y_i - \hat{\bar{Y}}(1)\right)^2 + \sum_{Z_i=0}\left(Y_i - \hat{\bar{Y}}(0)\right)^2 + \frac{n_1 n_0}{n}\hat{\tau}^2, && \text{(Combine terms)}
\end{aligned}$$

which completes the proof.

Acronyms

CRE completely randomized experiment. 1