Threshold-Rule Policy Learning via Strata-Means

Roberto Vacante

September 22, 2025

Let $i = 1, \ldots, N$ index individuals, $T_i \in \{0, 1\}$ the experimental assignment, $S_i \in \mathcal{S}$ the randomization stratum (female $\times$ location), and Y_i the outcome. Let X_i denote pre-treatment features used to build a *score* $\hat{\tau}(X_i)$ that orders units by predicted treatment benefit; only the ranking is required, so any strictly monotone transform of $\hat{\tau}$ leaves the rules below unchanged. In our baseline we set $X_i := S_i$ and form stratum-level mean contrasts: for each $s \in \mathcal{S}$,

$$\hat{\mu}_1(s) = \frac{\sum_i Y_i \, \mathbb{1}\{T_i = 1, S_i = s\}}{\sum_i \mathbb{1}\{T_i = 1, S_i = s\}}, \qquad \hat{\mu}_0(s) = \frac{\sum_i Y_i \, \mathbb{1}\{T_i = 0, S_i = s\}}{\sum_i \mathbb{1}\{T_i = 0, S_i = s\}},$$

and define $\hat{\tau}(X_i) := \hat{\tau}(S_i) = \hat{\mu}_1(S_i) - \hat{\mu}_0(S_i)$, which is piecewise constant within strata.

A *threshold rule* treats those whose score clears a cutoff:

$$d_\theta(x) = \mathbb{1}\{\hat{\tau}(x) \geq \theta\}, \qquad \theta \in \mathbb{R}.$$

To summarize the coverage/precision trade-off, we trace a *percentile path* by setting $\theta = q_p(\{\hat{\tau}_i\}_{i=1}^N)$, the empirical pth percentile with $p \in \{5, 10, \ldots, 95\}$. The induced treated share ("coverage") at θ is

$$\text{Coverage}(\theta) = \frac{1}{N} \sum_{i=1}^{N} d_\theta(X_i).$$

Because $X_i := S_i$, the score takes only $|\mathcal{S}|$ values—one per stratum—so moving θ switches whole strata on/off and produces discrete jumps in coverage.

Two diagnostics guide interpretation along the path. First, the departure from the experimental assignment is captured by the joint cells

$$N_{ab}(\theta) = \sum_{i=1}^{N} \mathbb{1}\{d_\theta(X_i) = a, \ T_i = b\} \quad (a, b \in \{0, 1\}),$$

and the *mismatch rate* $\text{mismatch}(\theta) = [N_{10}(\theta) + N_{01}(\theta)]/N$. Substantively, mismatch quantifies reallocations relative to the RCT (administrative load and distributional shifts); statistically, large mismatch or strata with p_s near 0 or 1 inflate sampling variability under inverse-probability weighting.[1] Second, we report *group composition* as shares $\frac{1}{N}\sum_i \mathbb{1}\{d_\theta(X_i) = 1, S_i = s\}$ across $s \in \mathcal{S}$.

The *policy value* is the finite-sample mean outcome were the rule implemented on these N units:

$$W(d_\theta) = \frac{1}{N}\sum_{i=1}^{N}\Big(d_\theta(X_i)\,Y_i(1) + [1 - d_\theta(X_i)]\,Y_i(0)\Big).$$

In a stratified RCT, $p_i := \Pr(T_i = 1 \mid S_i) = p_{S_i} \in (0,1)$ are known by design.[2] A Horvitz–Thompson (HT) estimator uses inverse match probabilities (Horvitz & Thompson, 1952):

$$\widehat{W}(d_\theta) = \frac{1}{N}\sum_{i=1}^{N} Y_i\left(\frac{T_i\,d_\theta(X_i)}{p_i} + \frac{(1-T_i)\,[1 - d_\theta(X_i)]}{1-p_i}\right), \qquad \Delta(\theta) = \widehat{W}(d_\theta) - \bar{Y}_{\mathrm{RCT}},$$

where $\bar{Y}_{\mathrm{RCT}} = \frac{1}{N}\sum_i Y_i$ is the experimental mean. We report $\Delta(\theta)$ along the percentile path and quantify uncertainty with a stratified bootstrap that resamples within strata (pointwise percentile bands; the lower band provides a "lower confidence bound").

As for identification, we assume SUTVA/consistency, blocked randomization with known propensities $p_s \in (0,1)$ and $T_i \perp (Y_i(0), Y_i(1)) \mid S_i$, strictly pre-treatment X_i, and positivity within used strata $(0 < p_s < 1)$. Under these conditions, *for any fixed, pre-specified rule* d_θ (or when valuation is carried out on held-out data), HT is design-unbiased for $W(d_\theta)$, so $\mathbb{E}[\Delta(\theta)] = W(d_\theta) - \bar{Y}_{\mathrm{RCT}}$. In our implementation the rule is learned and valued on the same sample, so the path $\theta \mapsto \Delta(\theta)$ is descriptive (potentially optimistic).

One might collapse the path to a single cutoff (the percentile maximizing the estimated value) to avoid cherry-picking. With discrete stratum-level scores, however, the curve is often (nearly) monotone, so the maximizer sits at a boundary (very low or very high p), effectively implying "treat almost all" or "treat almost none." Such boundary optima disregard budget/capacity, raise mismatch (increasing HT variance), and are fragile to small grid/sample changes. Absent pre-specified program inputs, we therefore refrain from selecting a single threshold and instead report the full percentile path with bootstrap bands and diagnostics.

[1] The Horvitz–Thompson valuation below uses observations whose realized assignment matches the rule, weighted by $1/p_s$ or $1/(1-p_s)$.

[2] By contrast, Kitagawa & Tetenov (2018) study policy learning when propensities are unknown and must be estimated. In our blocked RCT with known p_s, HT/IPW reduces to simple design-based normalizations, making the HT form natural here.

References

Horvitz, D. G., & Thompson, D. J. 1952. A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, **47**(260), 663–685.

Kitagawa, Toru, & Tetenov, Aleksey. 2018. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, **86**(2), 591–616.