Risk-Controlled Post-Processing of Decision Policies

Sunay Joshi[1,*] Tao Wang[1,*] Hamed Hassani[1] Edgar Dobriban[1]

[1]University of Pennsylvania [*]Equal contribution

arXiv

Why Post-Process a Policy?

A *baseline policy* $\pi_0 : \mathcal{X} \to \mathcal{A}$ is already deployed.

We seek a wrapper that **minimally modifies it** to satisfy a specified *risk constraint*.

Define the *violation risk*

$$R(\pi) := \mathbb{P}\{\ell(\pi(X), Y) \geq c\},$$

where c is the loss threshold and ε is the tolerance.

Agreement-Maximizing Risk Control

Constrained optimization framework:

$$\max_{\pi} \ \mathbb{P}\{\pi(X) = \pi_0(X)\} \quad \text{s.t.} \quad R(\pi) \leq \varepsilon.$$

Objective: preserve baseline behavior.

Constraint: chance-constraint on downstream loss.

Conditional Risk Reduction

Define *conditional violation risk* and the *best fallback action:*

$$g(a, x) := \mathbb{P}\{\ell(a, Y) \geq c \mid X = x\}, \qquad \pi_*(x) \in \arg\min_{a \in \mathcal{A}} g(a, x).$$

The *benefit* of switching at context x is

$$\Delta(x) := g(\pi_0(x), x) - g(\pi_*(x), x).$$

keep baseline $\pi_0(x)$ —$\Delta(x)$→ use fallback $\pi_*(x)$

Large $\Delta(x)$ means switching, as it decreases conditional violation risk.

Population-Optimal Post-Processor

Let $G_0 = R(\pi_0)$ and $G_* = R(\pi_*)$.

- If $G_0 \leq \varepsilon$, keep π_0 everywhere.
- If $G_* > \varepsilon$, infeasible.
- Otherwise, switch on the *largest* values of Δ.

Under atomless P_X, there exists threshold τ and subset $E \subseteq \{\Delta = \tau\}$ such that

$$G_* + \mathbb{E}[\Delta(X)\mathbf{1}\{\Delta(X) < \tau\}] + \tau P_X(E) = \varepsilon,$$

and an optimal deterministic policy is

$$\pi^\star(x) = \begin{cases} \pi_0(x), & x \in \{\Delta < \tau\} \cup E, \\ \pi_*(x), & x \in \{\Delta \geq \tau\} \setminus E. \end{cases}$$

Optimality principle: change the baseline only where switching reduces risk most.

Learn Where Switching Helps

On training data, *estimate* conditional risks, then form *plug-in estimates*

$$\hat{\pi}_*(x) \in \arg\min_{a \in \mathcal{A}} \hat{g}(a, x), \qquad \hat{\Delta}(x) = \hat{g}(\pi_0(x), x) - \hat{g}(\hat{\pi}_*(x), x).$$

For *threshold* τ, deploy

$$\hat{\pi}_\tau(x) = \begin{cases} \pi_0(x), & \hat{\Delta}(x) < \tau, \\ \hat{\pi}_*(x), & \hat{\Delta}(x) \geq \tau. \end{cases}$$

keep π_0 — $\hat{\tau}$ — use $\hat{\pi}_*$ → $\hat{\Delta}$

Decide How Much Switching is Safe

On *calibration data* $Z_i = (X_i, Y_i)$, define *bumped empirical risk*

$$\hat{R}_n^+(\tau) := \frac{1 + \sum_{i=1}^n \mathbf{1}\{\ell(\hat{\pi}_\tau(X_i), Y_i) \geq c\}}{n+1}.$$

Choose the *least-conservative* feasible threshold:

$$\hat{\tau} = \max\left\{\tau \in \{0, \top, \hat{\Delta}_1, \ldots, \hat{\Delta}_n\} : \hat{R}_n^+(\tau) \leq \varepsilon\right\}.$$

Why Conformalization Can Be Non-Monotone

For baseline / fallback violation indicators I_i^0, I_i^*,

$$L_i(\tau) = I_i^* + \underbrace{(I_i^0 - I_i^*)}_{W_i \in \{-1,0,1\}} \mathbf{1}\{\hat{\Delta}_i < \tau\}.$$

Since W_i can have either sign, $L_i(\tau)$ **need not be monotone.**

If K is *leave-one-out rank-instability,* then

$$\frac{1}{N}\sum_{i=1}^N |L_i(\hat{\tau}_{-i}) - L_i(\hat{\tau}_{1:N})| \leq \frac{2K}{N}.$$

Positive drift implies $\mathbb{E}[K] = O(\log N)$.

Risk and Agreement Guarantees

Exact-safe fallback. If $\ell(\pi_{\text{safe}}(x), y) < c$ for every (x, y), exchangeability gives

$$\mathbb{E}[L(Z_{n+1}; \hat{\tau})] \leq \varepsilon.$$

Learned fallback. Conditionally on the fit, under i.i.d. sampling, and positive drift,

$$\mathbb{E}[L(Z_{n+1}; \hat{\tau})] \leq \varepsilon + C\frac{\log n}{n}.$$

Near-optimal agreement (exact-safe regime). Under local regularity, w.p. $1 - \delta$,

$$(J^* - \mathbb{P}\{\hat{\pi}_{\hat{\tau}}(X) = \pi_0(X)\})_+ = O\left(\sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon_\Delta\right).$$

Synthetic: Tracks the Oracle

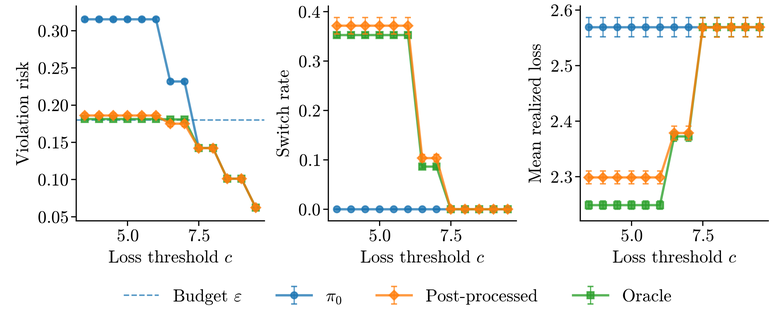


Figure 1. Dashed: $\varepsilon = 0.18$. Fitted post-processor tracks oracle in risk, switch rate, realized loss.

Medical Diagnosis and LLM Routing

Task	Method	Risk	Switch	Cost / loss
COVID, $c = 4$	Post-processed	.10	.09	**.81**
	Random mixing	.10	.22	1.07
LLM, $\varepsilon = .30$	Post-processed	.30	.48	**92.7T**
	Random mixing	.30	.63	114.8T

Table 1. Test means over 20 seeds. COVID reports mean loss; LLM reports selected-model TFLOPs.

COVID-19: at same risk, algo changes clinical policy $\mathbf{50}\%$ **less often.**

LLM routing: at same error rate, algo uses $\mathbf{19}\%$ **less compute.**

Conclusion

- **Learn where switching helps; calibrate how much switching is safe.**
- The oracle solution is a threshold on conditional risk reduction.
- Targeted switching *preserves more baseline behavior* than random mixing.

References

Angelopoulos et al. (2024), *Conformal Risk Control*.
Kiyani et al. (2025), *Decision Theoretic Foundations for Conformal Prediction*.