Active Decision Experimentation

I-han Lai[1] Puheng Li[2] Xiaojie Mao[3] Renyuan Xu[1] (alphabetical order)

[1]Stanford MS&E [2]Stanford Statistics [3]Tsinghua SEM

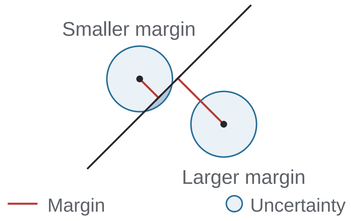
Motivation

Experiments guide decisions, but experimentation can be costly.

With a limited budget, which data improve future decisions?

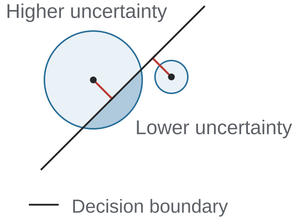
- **Goal:** Learn which action works best for each individual or context.
- **Challenge:** More accurate estimates do not necessarily lead to better decisions.
- **High-level ideas:** Adaptively choose **whom to enroll** and **which action to test**, prioritizing information that helps avoid costly mistakes.

Same uncertainty



A larger margin reduces the chance of a mistake.

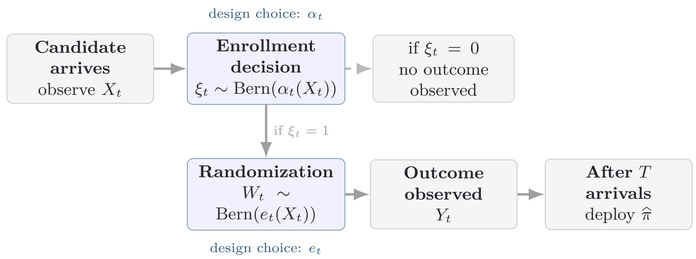
Same margin



Greater uncertainty increases the chance of a mistake.

Problem setup

For each of T arrivals, covariates X_t are observed at no cost. Enrolled individuals are randomized to treatment or control.



$\alpha_t(x)$ is enrollment probability and $e_t(x)$ is treatment probability if enrolled. ξ_t indicates enrollment. $W_t = 1$ denotes treatment and $W_t = 0$ control.

$$\mathbb{E}\left[\sum_{t=1}^{T} \xi_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} \alpha_t(X_t)\right] \le B.$$

B is the outcome budget. $Y(1)$ and $Y(0)$ are potential outcomes under treatment and control. Y_t is the observed outcome. The treatment effect is

$$\tau(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x].$$

The optimal policy $\pi^\star$ assigns treatment when $\tau(x) > 0$ and control otherwise. The learned policy $\hat{\pi}$ substitutes an estimate of τ.
The design objective is to minimize **deployment regret**, the expected outcome loss relative to $\pi^\star$ on new individuals:

$$R(\hat{\pi}) = \mathbb{E}_X[|\tau(X)| \times \Pr\{\hat{\pi}(X) \neq \pi^\star(X)\}].$$

Regret and allocation

Kernel smoothing defines a local effect $\theta_h(x)$ at bandwidth h. Its asymptotically normal estimator yields the Gaussian regret surrogate

$$\ell_h(x, v) = |\theta_h(x)| \times \Phi\left(-\frac{|\theta_h(x)|}{\sqrt{v}}\right).$$

v denotes estimator variance and Φ the standard normal CDF. **The average surrogate loss converges to deployment regret as the budget grows.**

Decision relevance

$\Gamma_t(u)$ measures the sensitivity of surrogate regret to noise at u.

$$\Gamma_t(u) \;\propto\; \mathbb{E}_X[q_h(X, u)^2\, \partial_v \ell_h(X, v_t(X))].$$

$q_h(x, u)$ is the normalized kernel weight at u for estimating $\theta_h(x)$. v_t is cumulative predictable variance scaled by T^{-2}.

Adaptive design

Phase 1 **Burn-in**

Estimate effects and conditional outcome standard deviations σ_1, σ_0 under uniform enrollment and equal randomization. The burn-in fraction may vanish asymptotically.

Phase 2 **Active enrollment and assignment**

Allocation combines estimated decision relevance $\hat{\Gamma}_t$ with **conditional noise variance of the weighted contrast, $\hat{H}_t$:**

$$\hat{H}_t(u) = \frac{\hat{\sigma}^2_{1,t}(u)}{e_t(u)} + \frac{\hat{\sigma}^2_{0,t}(u)}{1 - e_t(u)}.$$

Enrollment

$$\alpha_t(u) \propto \sqrt{\hat{\Gamma}_t(u)\, \hat{H}_t(u)}.$$

Treatment assignment

$$e_t(u) = \frac{\hat{\sigma}_{1,t}(u)}{\hat{\sigma}_{1,t}(u) + \hat{\sigma}_{0,t}(u)}.$$

Enrollment is normalized to the budget. Both rules satisfy exploration bounds. Nuisance estimates are refitted after each stage.

Phase 3 **Deployment**

Deploy the sign rule based on the final local effect estimate.

Optimal allocation near the decision boundary

As the budget grows, regret concentrates near the treatment boundary. In a small neighborhood $\mathcal{N}$, the optimal allocation is

$$a^\star(x) = a_{\max}, \qquad e^\star(x) = \frac{\sigma_1(x)}{\sigma_1(x) + \sigma_0(x)}, \quad x \in \mathcal{N}.$$

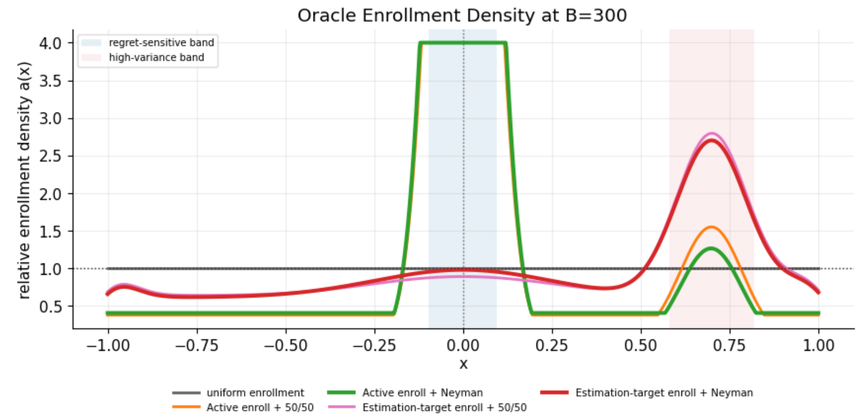
$a = (T/B)\alpha$ denotes relative enrollment with upper bound $a_{\max}$.

The adaptive design converges to this optimal rule near the decision boundary under the stability conditions of our analysis.

Its regret is **asymptotically optimal among stable, bounded designs**.

Simulation

The simulation separates regions contributing most to **decision regret** and **estimation error**.

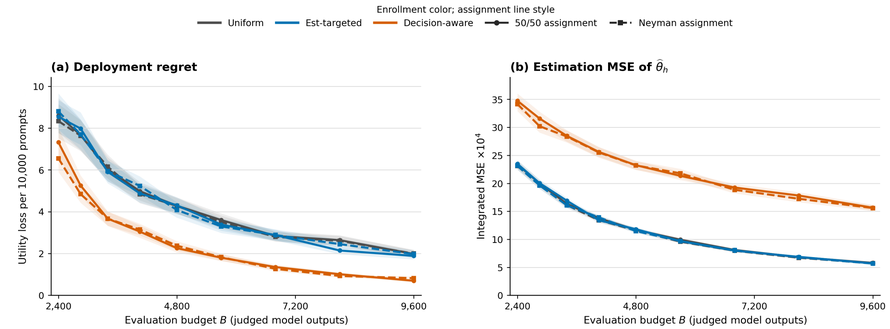


- **Decision-aware enrollment** concentrates near the boundary.
- **Estimation-targeted enrollment** concentrates in the high-noise region.

Uniform enrollment corresponds to relative density $a(x) = 1$.

Language model routing

Each prompt is routed to **GPT-4 or Mixtral-8x7B** to balance response quality and inference cost. Using **RouterBench** (Hu et al., 2024), we learn this routing rule by choosing which prompts to judge and which model to evaluate under a limited judgment budget.



Decision-aware enrollment lowers regret despite higher estimation error.

At B = 9,600, regret is approximately one third and mean squared error is 2.7 times that of uniform and estimation-targeted designs.

Hu et al. (2024). *RouterBench: A Benchmark for Multi-LLM Routing System.* arXiv:2403.12031.