LEARNING UNDER MECHANISM-INDUCED COVARIATE SHIFT: FINITE-SCALE COVERAGE AND SOURCE ALLOCATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Simulators and physical acquisition systems can generate abundant data yet still leave deployment-relevant neighborhoods poorly observed, even when those neighborhoods are reachable. We study this as mechanism-induced covariate shift: the conditional law $P(Y \mid Z)$ is shared, but a generator's acquisition law and forward map determine where training examples appear in task space. The key quantity is the finite-scale local-mass field—the probability that one generated example falls within a neighborhood of each deployment point. This quantity connects the generator directly to finite-budget learning. For Nadaraya–Watson (NW) regression, it yields a finite-sample risk certificate with constants uniform over the generator family, and on a controlled one-dimensional common-support family the optimized certificate matches minimax risk up to family-uniform constants. Keeping the full spatial field matters: generators can have identical normalized-Jacobian distributions and identical distributions of source-density values at deployment draws while having different finite-scale coverage. The same field also turns upstream data collection into a pre-training decision problem: task matching removes generator-dependent risk differences at fixed labels, exact rejection sampling exposes the corresponding proposal cost, and source mixing gives a convex allocation problem that can exploit spatial complementarity. On a controlled 23-design family, the coverage score tracks exact finite-n NW risk with median target-wise Spearman correlation 0.97–1.00 across six settings. Estimated from unlabeled forward samples, finite-scale coverage can therefore guide source selection and allocation before training, without learner-side feedback.

1 INTRODUCTION

Structured data generators—including simulators and physical acquisition systems—can provide training data at scale when deployment data are scarce or costly. Yet their acquisition laws and forward maps determine where examples appear in task space. A region may be physically reachable but receive so little probability that a finite training set contains no nearby example, leaving a learner data-rich overall but data-poor where deployment places weight. Reachability is not coverage. We call this mechanism-induced covariate shift: under a fixed deployment law and shared $P(Y \mid Z)$, the training covariate distribution is induced by the generator's acquisition law and forward map.

Existing work starts from different points in this process. Covariate-shift theory analyzes learning once the source and deployment laws are given (Shimodaira, 2000; Pathak et al., 2022; Ma et al., 2023); Pathak et al. (2022) identify reciprocal neighborhood probabilities as learning-relevant while treating the source law as fixed. Data-mixture and data-selection methods instead choose sources using pilot training, proxy losses, validation, or online performance (Hashimoto, 2021; Xie et al., 2023a; Chen et al., 2025; Tan et al., 2026), and therefore rely on learner-side feedback. Adjacent approaches tune generators using target or task signals (Veeravasarapu et al., 2017; Ruiz et al., 2019; Mishra et al., 2022), compare source and deployment distributions with task-metric-aware objectives (Vallender, 1974; Gretton et al., 2012), or select covering subsets from an unlabeled candidate pool (Sener & Savarese, 2018). What remains unclear is how information available from the generator before training can inform both finite-budget learning and upstream data decisions. We ask: *What generator-side information links finite-budget learning under mechanism-induced covariate shift to acquisition and source-allocation decisions before training?*

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

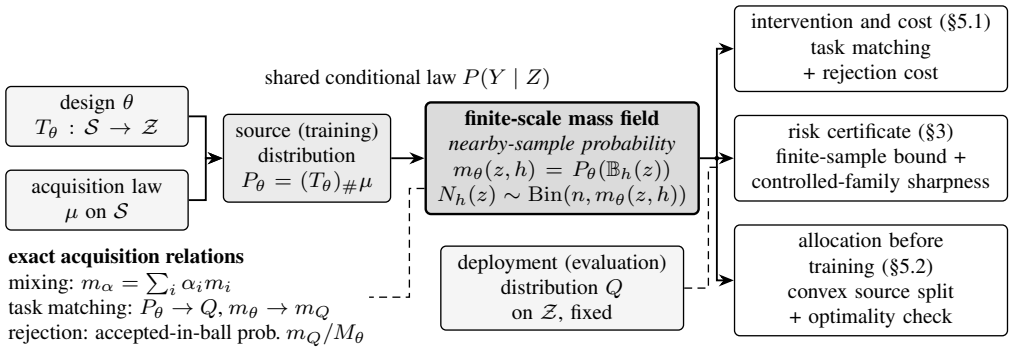

Figure 1: From mechanism to decisions. A design θ and an acquisition law μ induce the source law $P_\theta = (T_\theta)_{\#}\mu$ on task space $\mathcal{Z}$ and the mass field $m_\theta(z,h)$, the probability that one training example falls within h of z. With a fixed deployment law Q and shared $P(Y \mid Z)$, the field links local sample availability to risk certification (section 3), task matching and rejection cost (section 5.1), and source allocation (section 5.2). The acquisition relations use mixture weights α_i, source fields m_i, deployment field m_Q, and rejection envelope M_θ.

Three obstacles make this connection nontrivial. First, reachability is too coarse: support says whether a region can be sampled at all, not whether n examples are likely to place data nearby. Second, value-level summaries can preserve which Jacobian or density values occur while losing where they occur; undersampled deployment neighborhoods can create a finite-budget risk obstruction. Third, learner-side feedback arrives only after training, whereas upstream decisions must anticipate how changing the generator changes local data availability. The relevant generator-side information must therefore retain finite-budget spatial coverage, connect to risk, and behave predictably under feasible operations such as source mixing, task matching, and proposal thinning.

Our answer, in nonparametric regression, is the finite-scale local-mass field of the generator-induced source law. A generator samples $q \sim \mu$ and maps it through T_θ, inducing $P_\theta = (T_\theta)_{\#}\mu$; at scale h, $m_\theta(z,h) := P_\theta(\mathbb{B}_h(z))$ is the probability that one training example falls within distance h of a deployment point z. If $N_h(z)$ counts, among n training examples, those in $\mathbb{B}_h(z)$, then $N_h(z) \sim \mathrm{Binomial}(n, m_\theta(z,h))$. Thus the field has a direct finite-budget meaning while retaining where coverage is weak. It also obeys explicit relations under source mixing, task matching, and rejection sampling and can be estimated from unlabeled forward samples before model fitting. Figure 1 summarizes the chain from generator and acquisition law to source law, local sample availability, risk certification, acquisition cost, and source allocation. We develop the finite-sample theory for NW regression, evaluate the resulting acquisition interventions and allocation rules on controlled planar families, and examine whether the resulting coverage score remains informative for additional regression learners on a seven-joint arm family. We make three main contributions:

- **Finite-scale coverage links generation to learning.** We derive an NW finite-sample risk certificate whose design dependence enters only through the local-mass field, show that its optimized form is sharp up to family-uniform constants on a controlled one-dimensional common-support family, and establish a learner-independent no-observation lower bound (theorem 1 and proposition 10). This turns generator-side coverage available before training into the design-dependent input of a finite-budget risk certificate.
- **Value-level summaries do not determine spatial coverage.** We construct a design-normalized family with identical complete distributions of normalized Jacobian and source-density values at deployment draws but different finite-scale coverage (theorem 2). Thus two generators can look identical under these value summaries while differing in the local coverage relevant to finite-sample learning.
- **The same field supports acquisition and source allocation.** Task matching removes generator-dependent risk differences at fixed labels while exact rejection sampling exposes proposal cost (proposition 1); the mass fields define a convex pre-training allocation problem, and an exact two-source identity shows that single-source profile values do not determine mixture value (proposition 2 and corollary 7). Thus the same pre-training field can be used not only to diagnose coverage but also to decide how to acquire and mix data.

2 FORWARD-MAP LEARNING UNDER A FIXED DEPLOYMENT LAW

Generator and statistical model. Within each design family we fix a source space $\mathcal{S}$ (for a robot, its configuration space) and an acquisition law μ on it. A design θ specifies a measurable forward map $T_\theta : \mathcal{S} \to \mathcal{Z}$ into a common separable metric task space $(\mathcal{Z}, d_{\mathcal{Z}})$. If the source variable $q \sim \mu$ and $Z = T_\theta(q)$, the training covariates follow the pushforward $P_\theta := (T_\theta)_{\#}\mu$, the task-space source distribution induced by the design and acquisition law. Performance is evaluated under a deployment distribution Q on $\mathcal{Z}$, fixed and shared across designs. Labels follow the shared conditional model $Y = f(Z) + \varepsilon$ with f in the Hölder class $\mathcal{F} = \{f : |f| \le B,\ |f(z) - f(z')| \le \Lambda\, d_{\mathcal{Z}}(z, z')^\beta\}$, $0 < \beta \le 1$; the conditional noise law is shared across designs, with $\mathbb{E}[\varepsilon \mid Z] = 0$ and conditional variance at most σ^2. From n independent pairs (Z_i, Y_i) with $Z_i \sim P_\theta$ the learner returns an estimate $\hat{f}$, evaluated by squared loss in $L^2(Q)$. Designs therefore differ statistically only through P_θ; we refer to this statistical model as common-conditional covariate shift.

Relationality principle. With the deployment law, conditional model, function class, estimator class, and sample size fixed, a design affects the statistical experiment only through its induced source law P_θ. Two designs inducing the same P_θ define the same statistical experiment, and replacing P_θ by Q makes all designs identical at a fixed label budget (section B.1). Difficulty is therefore relational: it depends on how the induced source law covers the fixed deployment law. The forward maps and acquisition laws determine the realizable source laws over which acquisition and allocation operate.

Finite-scale occupancy. To describe that relation at the scale seen by a finite sample, for $z \in \mathcal{Z}$ and $h > 0$ let

$$m_\theta(z, h) := P_\theta\big(\mathbb{B}_h(z)\big), \qquad \mathbb{B}_h(z) := \{u \in \mathcal{Z} : d_{\mathcal{Z}}(u, z) \le h\}, \tag{1}$$

the probability that one training covariate falls within distance h of z. If $N_h(z)$ counts the n training covariates in $\mathbb{B}_h(z)$, then exactly $N_h(z) \sim \mathrm{Binomial}(n, m_\theta(z, h))$, so $m_\theta(z, h)$ has an immediate finite-sample meaning: it sets the empty-neighborhood probability $(1 - m_\theta(z, h))^n$ and, for the local estimator of section 3, how many noisy labels are averaged. We call m_θ the mass field; read at a fixed h, it retains where source probability lies relative to task neighborhoods rather than which density values occur. Its deployment-weighted scalar summary is the profile

$$\Psi_\theta(h) := \int \frac{dQ(z)}{m_\theta(z, h)}, \qquad 1/0 := +\infty, \tag{2}$$

which is large when the deployment law places weight on regions with small neighborhood mass. This scalar is the source-only specialization of the ball-probability functional of Pathak et al. (2022). $\Psi_\theta(h)$ summarizes coverage shortage; the full field m_θ locates it and underlies the target-specific comparisons of section 3 and the source complementarity of section 5.2.

Support and acquisition. Finite-scale shortage must be kept apart from structural unreachability: a support deficit means that a design cannot reach part of the deployment region at all, whereas coverage difficulty means that the region is reachable but receives little source mass. Support inclusion guarantees only that every deployment neighborhood has positive source probability; at budget $n \ge 2$, if $m_\theta(z, h) \le 1/n$, that neighborhood is empty with probability at least $1/4$. We therefore make all main cross-design comparisons on common support, where the reachability difference is removed and only finite-scale coverage varies. The same field admits closed-form relations under the three acquisition operations used later: mixing sources with weights α averages their fields; task matching replaces P_θ by Q; and exact rejection sampling toward Q (for $Q \ll P_\theta$ with $w = dQ/dP_\theta$ bounded by an envelope $M_\theta < \infty$) accepts each proposal with probability w/M_θ:

$$m_\alpha(z, h) = \textstyle\sum_i \alpha_i m_i(z, h), \quad m_\theta \mapsto m_Q := Q(\mathbb{B}_h(\,\cdot\,)), \quad \Pr\{\text{accepted in } \mathbb{B}_h(z)\} = m_Q(z, h)/M_\theta.$$

The last relation is the probability that a proposal is accepted and lies in $\mathbb{B}_h(z)$ (section 5.1).

3 FINITE-SCALE COVERAGE CONTROLS RISK

We now ask whether local occupancy controls prediction risk with design-independent constants and whether this control can be sharp.

For radius h, the NW estimator (Nadaraya, 1964; Watson, 1964; Győrfi et al., 2002), $\hat{f}_{n,h}(z)$, averages the labels of the $N_h(z)$ training examples in $\mathbb{B}_h(z)$ and returns 0 when the ball is empty; write $R^{\mathrm{NW}}_{\theta,f}(n,h) := \mathbb{E}\|\hat{f}_{n,h} - f\|^2_{L^2(Q)}$ for its exact deployment risk. Optimizing its certificate over a design-independent admissible scale range gives

$$\mathcal{C}_\theta(n) := \inf_{0<h\le\bar{h}} \Big\{\Lambda^2 h^{2\beta} + \sigma^2 \int g_n(m_\theta(z,h))\, dQ(z) + B^2 \int (1 - m_\theta(z,h))^n\, dQ(z)\Big\}, \quad (3)$$

where $g_n(m) := \mathbb{E}[1/N\,;\, N>0]$ for $N \sim \mathrm{Binomial}(n,m)$ is the expected reciprocal count on a nonempty ball and $\bar{h}$ is a design-independent feasible upper scale shared by the family (section B.2). For the sharpness result, consider a controlled one-dimensional family on $[0,1]$: every source distribution has a density between $1/C$ and $1/c$, as when a uniformly sampled source is pushed through a monotone C^1 bijection T_θ with $0 < c \le |T'_\theta| \le C$; the noise is Gaussian with $\sigma > 0$; the Hölder parameters satisfy $\Lambda > 0$ and $B > 0$; and the deployment density lies in $[q_{\min}, q_{\max}]$ with $q_{\min} > 0$. Bracketed labels state the scope of each formal result and are collected in section A.

Theorem 1 (Finite-scale certificate and controlled-family sharpness)**.** (a) *[uniform constants; fixed design] For every design θ, every $n \ge 1$, and every $h > 0$, if conditionally on $Z_{1:n}$ the noises $\varepsilon_1, \dots, \varepsilon_n$ are independent and satisfy $\mathbb{E}[\varepsilon_i \mid Z_{1:n}] = 0$ and $\mathbb{E}[\varepsilon_i^2 \mid Z_{1:n}] \le \sigma^2$, then*

$$\sup_{f\in\mathcal{F}} \mathbb{E}\|\hat{f}_{n,h} - f\|^2_{L^2(Q)} \;\le\; \Lambda^2 h^{2\beta} + \sigma^2 \int g_n(m_\theta(z,h))\, dQ(z) + B^2 \int (1 - m_\theta(z,h))^n\, dQ(z). \quad (4)$$

The numerical constants are 1 *and do not depend on θ; the design enters only through the mass field m_θ.* (b) *[controlled 1D; $n \ge n_0$] On the controlled one-dimensional family there exist a threshold n_0 and positive constants κ_1, κ_2, all depending only on the controlled-family parameters and not on the design, such that for every design in the family and every $n \ge n_0$,*

$$\frac{\kappa_1}{\kappa_2}\, \mathcal{C}_\theta(n) \;\le\; \inf_{\hat{f}} \sup_{f\in\mathcal{F}} \mathbb{E}\|\hat{f} - f\|^2_{L^2(Q)} \;\le\; \mathcal{C}_\theta(n). \quad (5)$$

Interpretation. The certificate separates Hölder bias at scale h, noise averaged over the random local count, and empty-neighborhood loss. Only the latter two depend on the design, through $m_\theta(z,h)$. Complete proofs are in section B.2.

The theorem is field-level; the profile enters only as a scalar consequence. Under homoscedastic noise, $F_{n,\theta}(h) = \sigma^2 \int g_n(m_\theta(z,h))\, dQ(z)$ is the exact target-independent NW variance term and satisfies $F_{n,\theta}(h) \le 2\sigma^2 \Psi_\theta(h)/n$. For $\sigma > 0$ and fixed h with uniformly positive mass, $nF_{n,\theta}(h)/\sigma^2 \to \Psi_\theta(h)$ (corollary 3). Thus $\Psi_\theta(h)$ summarizes the fixed-scale noise floor.

On a controlled one-dimensional common-support family, the finite-scale NW certificate is minimax-sharp up to constants uniform over the design family. The empty-neighborhood obstruction also applies to arbitrary estimators. With the conditional noise kernel fixed across $f \in \mathcal{F}$, proposition 10 gives the learner-independent minimax lower bound $(1 - m_\theta(z_0,h))^n\, Q(\mathbb{B}_{h/2}(z_0))\, \min\{B, \Lambda(h/2)^\beta\}^2$ for any $z_0 \in \mathcal{Z}$ and $h > 0$; a positive-Q support deficit makes the obstruction independent of sample size (corollary 8). This isolates the no-observation mechanism.

Evidence for the certificate. We study the 23-design common-support planar three-revolute-joint (3R) family with joint-uniform sampling, $\sum_j L_j = 1$, radial task coordinate $z \in [0,1]$, and deployment law $dQ(z) = 2z\, dz$. The median target-wise Spearman correlation between exact NW risk and $\Psi_\theta(h)$ over 30 smooth targets lies within 0.97–1.00 in all six (n,h) cells (fig. 2(a)). A localized-target comparison shows why spatial resolution matters. Across nine scored centers, one design has lower excess risk (risk above the label-noise floor) at six and the other at three (fig. 2(b)). Thus no target-independent scalar ordering of the two designs reproduces all outcomes. The profile supports aggregate screening in this family, while the full mass field retains the spatial information absent from that summary (section C.1).

4 Rigid Local Summaries, Non-Rigid Coverage

Could richer global summaries recover the spatial information lost by scalarization? Not necessarily: even the full distributions of normalized Jacobian values and of density values can be identical across designs while finite-scale coverage differs.

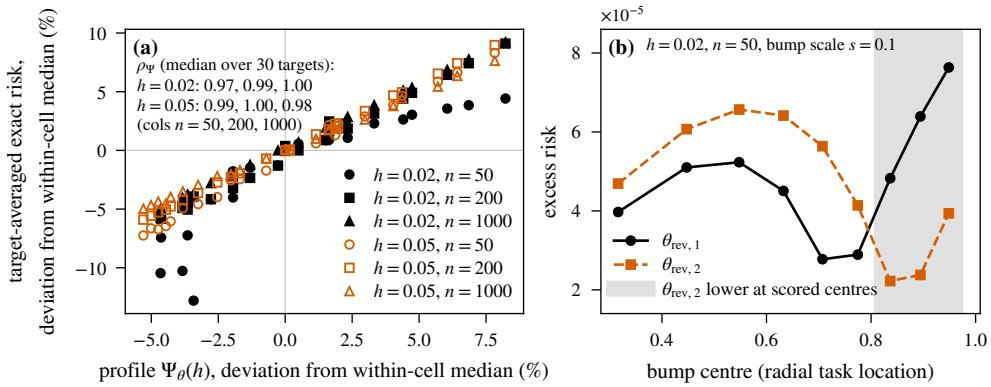

Figure 2: Profile, exact risk, and target-specific reversal. (a) Target-averaged exact finite-n NW risk versus $\Psi_\theta(h)$ for 23 common-support three-link designs, 30 smooth targets, and six (n, h) cells; both axes show within-cell median deviations, and ρ_Ψ is the median target-wise Spearman correlation per cell. (b) Excess risk $R^{\mathrm{NW}}_{\theta,f}(n, h) - F_{n,\theta}(h)$ for the pre-specified pair $\theta_{\mathrm{rev},1} = (0.45, 0.42, 0.13)$ and $\theta_{\mathrm{rev},2} = (0.48, 0.45, 0.07)$ across nine radial bump centers ($h = 0.02$, $n = 50$, bump half-width $s = 0.1$); the first design is lower at the first six centers and the second at the last three (shaded).

Consider a planar two-revolute-joint (2R) arm with link lengths $L_1 = 1$ and $L_2 = a \in (0, 1)$, elbow angle uniform on $[0, 2\pi)$, and radial task quantity $r = |1 + ae^{iq}|$. Since r depends on q only through $\cos q$, folding onto $[0, \pi]$ preserves uniformity. Because the reachable radial interval depends on a, designs are compared only in the normalized coordinate $x = (r - (1 - a))/(2a) \in [0, 1]$ with $Q_x = \mathrm{Unif}[0, 1]$. Let P_a be the law of x, p_a its density, $m_a(x, h) := P_a([x - h, x + h] \cap [0, 1])$ its mass field, and $J_x = |dx/dq|$ the normalized Jacobian.

Theorem 2 (Rigidity of local summaries without finite-scale rigidity)**.** *[design-normalized 2R] For every $a \in (0, 1)$:* (a) *J_x has the same distribution as $\frac{1}{2}\sin U$ with $U \sim \mathrm{Unif}[0, \pi]$, independently of a; hence every power moment of J_x of order $s > -1$ (the orders for which it is finite) is design-invariant. Moreover, the full distribution of the density value $p_a(X)$, $X \sim \mathrm{Unif}[0, 1]$, is design-invariant, and so is every well-defined power- or Rényi-type functional of that distribution, including $\chi^2(Q_x \,\|\, P_a) = \int_0^1 p_a(x)^{-1}\, dx - 1 = \pi^2/8 - 1$.* (b) *$p_a(\frac{1}{2}) = 2/(\pi\sqrt{1 - a^2/4})$ is strictly increasing in a; hence for $a_1 \neq a_2$ there is $\eta > 0$ such that $m_{a_1}(\frac{1}{2}, h) \neq m_{a_2}(\frac{1}{2}, h)$ for all $0 < h < \eta$. The finite-scale mass field is therefore not design-invariant.*

Interpretation. Even complete value histograms are insufficient: they record which Jacobian or density values occur, but not where. Any functional of the law of J_x or of $p_a(X)$ is therefore identical across this family, while finite-sample neighborhood occupancy can differ (proofs in section B.3).

The theorem gives a structural comparison in the design-normalized coordinate x; shared-deployment comparisons elsewhere in the paper use a common task coordinate, metric, and deployment law. The coarea formula identifies the spatial information lost by these summaries.

Under the regularity conditions of section B.3.2, let $k = \dim \mathcal{S}$, $d = \dim \mathcal{Z}$, ϱ be the source-space density, crit the critical set, and J_{T_θ} the normal Jacobian. For almost every z, the coarea formula (Evans & Gariepy, 2015) gives the density p_θ of P_θ as

$$p_\theta(z) = \int_{T_\theta^{-1}(z) \setminus \mathrm{crit}} \frac{\varrho(q)}{J_{T_\theta}(q)}\, d\mathcal{H}^{k-d}(q), \qquad m_\theta(z, h) = \int_{\mathbb{B}_h(z)} p_\theta(u)\, du,$$

where $\mathcal{H}^{k-d}$ is Hausdorff measure along the fiber. The density at z aggregates source weight and inverse-Jacobian factors over its fiber; the mass field integrates the resulting density over task neighborhoods. A Jacobian-value histogram loses the fiber correspondence, and a density-value histogram loses spatial placement. Near a fold, density can be unbounded while $m_\theta(z, h) \leq 1$, motivating the field-level form of theorem 1(a).

Evidence for finite-scale non-rigidity. The design-normalized 2R witness has identical density-value laws yet a 4.2% profile difference at $h = 0.1$ (fig. 3(a), section B.3). The separate 23-design

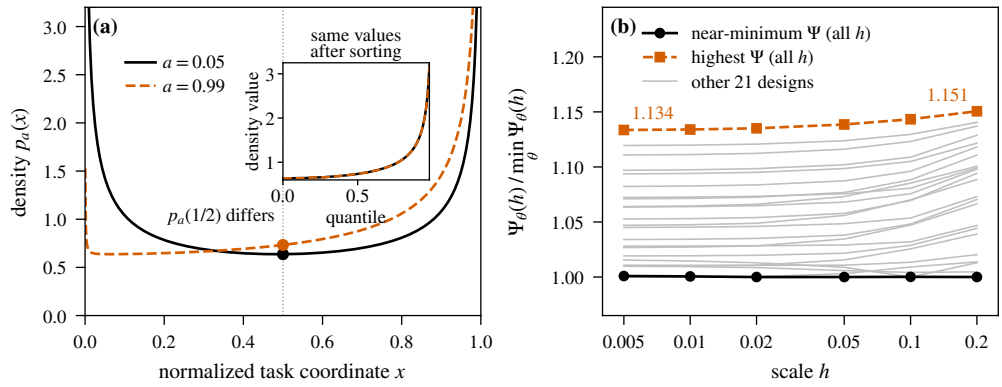

Figure 3: Rigid summaries versus finite-scale coverage. (a) Design-normalized 2R designs $a = 0.05$ and $a = 0.99$ have different densities but the same sorted density-value curve (inset, $X \sim \mathrm{Unif}[0,1]$); small-scale mass near $x = \frac{1}{2}$ differs (endpoint singularities clipped). (b) In the separate 23-design common-task 3R family, profiles are normalized by $\min_\theta \Psi_\theta(h)$ per scale; $(0.48, 0.42, 0.10)$ stays within 0.1% of the minimum and $(0.36, 0.33, 0.31)$ is maximal throughout (designs A, B of fig. 4); the family ratio exceeds one by 13–15% across scales.

common-task 3R family shows 13–15% profile spread across scales (fig. 3(b)), with the corresponding prediction-risk comparisons in fig. 2.

Thus global value distributions need not determine finite-scale coverage, and realizable common-task sources can differ in coverage. We next ask whether acquisition can change this field, at what cost, and how to combine sources before training.

5 From Coverage to Acquisition and Allocation

5.1 Identifying the Acquisition Channel

Task matching sets the training covariate law to the fixed deployment law Q. By the relationality principle, all designs then induce the same statistical experiment, removing the generator-dependent risk gap at a fixed label budget. The remaining issue is implementation cost.

Assume $Q \ll P_\theta$, let $w = dQ/dP_\theta$, and define $M_\theta := \operatorname{ess\,sup}_{P_\theta} w$, the smallest valid constant envelope. If $M_\theta < \infty$, exact rejection sampling draws proposals from P_θ, accepts Z with probability $w(Z)/M_\theta$, and labels only accepted proposals. Write N_{prop} for the proposal budget. To connect this cost to finite-scale coverage, define the deployment profile $\Psi_Q(h) := \int dQ(z)/m_Q(z,h)$. Define also the ball-mass domination coefficient $D_\theta(h) := \operatorname{ess\,sup}_{z\sim Q} m_Q(z,h)/m_\theta(z,h)$, the essential supremum of the deployment-to-source neighborhood-mass ratio.

Proposition 1 (Task-matched acquisition and cost sandwich)**.** *[$Q \ll P_\theta$; finite envelope where stated] Assume $Q \ll P_\theta$.* (a) *If $M_\theta < \infty$, then for any estimator that reads only the accepted samples, the number of accepted samples is* $\mathrm{Binomial}(N_{\mathrm{prop}}, 1/M_\theta)$ *and, given that number, the accepted labeled samples are independent draws from the task-matched experiment; hence the rejection-sampling risk is the* $\mathrm{Binomial}(N_{\mathrm{prop}}, 1/M_\theta)$*-mixture of task-matched risks, including the zero-sample term.* (b) *If $m_Q(z,h) > 0$ and $m_\theta(z,h) > 0$ for Q-a.e. z, then $D_\theta(h) \le M_\theta$ and $\Psi_\theta(h) \le D_\theta(h)\,\Psi_Q(h)$ in the extended reals, and if moreover $\Psi_Q(h) < \infty$,*

$$\frac{\Psi_\theta(h)}{\Psi_Q(h)} \;\le\; D_\theta(h) \;\le\; M_\theta. \tag{6}$$

Interpretation. Part (a) separates two resource frames: at a fixed accepted-label count, task-matched risk is design-invariant; at a fixed proposal budget, the accepted count has mean $N_{\mathrm{prop}}/M_\theta$, so exact rejection requires M_θ proposals per acceptance on average. Part (b) shows that the profile and ball-mass ratios lower-bound, but do not upper-bound, exact-rejection cost (proofs in section B.4.1).

Under exact rejection sampling, each proposal is accepted in $\mathbb{B}_h(z)$ with probability $m_Q(z,h)/M_\theta$, so the accepted in-ball count is $\mathrm{Binomial}(N_{\mathrm{prop}}, m_Q(z,h)/M_\theta)$. For zero-fallback uniform-ball NW under homoscedastic noise, this gives the fixed-proposal identities of corollaries 5 and 6;

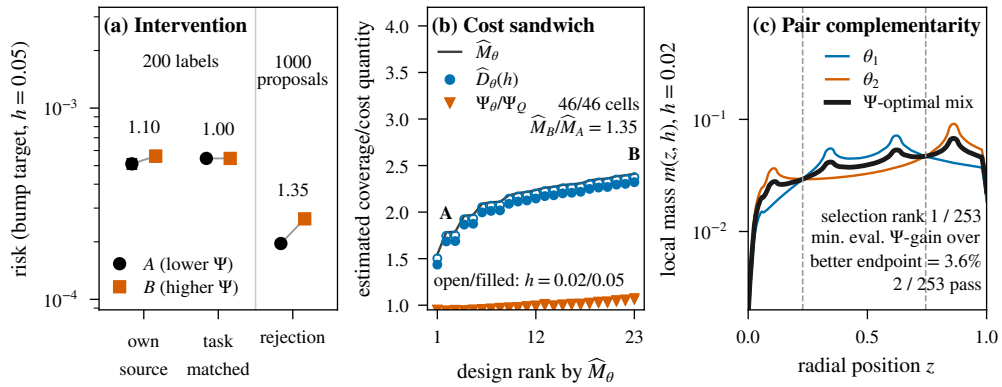

Figure 4: Acquisition intervention and allocation. (a) For $A = (0.48, 0.42, 0.10)$ and $B = (0.36, 0.33, 0.31)$ at $h = 0.05$, task matching removes the own-source gap at $n = 200$ labels; the modeled $N_{\text{prop}} = 1000$ thinning arm with candidate $\widehat{M}_\theta$ reveals proposal cost. Displayed ratios are B/A within each resource frame. (b) For 23 designs and $h = 0.02, 0.05$, $\Psi_\theta/\Psi_Q \le \widehat{D}_\theta(h) \le \widehat{M}_\theta$ holds in 46/46 cells, with $\widehat{M}_B/\widehat{M}_A = 1.35$. (c) The selection-rank-1 pair, $\theta_1 = (0.48, 0.33, 0.19)$ and $\theta_2 = (0.48, 0.45, 0.07)$, is shown at $h = 0.02$. Dashed lines mark endpoint crossings, and the selection-derived Ψ-optimal mixture raises mass in complementary regions without pointwise dominance (section C.4).

m_Q/M_θ is a success-probability field, not the mass field of a new law. The three-link density-ratio boundary analysis is in section B.4.1.

Intervention evidence. In the NW experiment on the 23-design common-support family (section C.2), task matching removes the fixed-label gap; for A, B the modeled fixed-proposal thinning-risk ratio is close to $\widehat{M}_B/\widehat{M}_A = 1.35$ (fig. 4(a)). With grid estimate $\widehat{D}_\theta(h)$ and candidate envelope $\widehat{M}_\theta$, the numerical analogue of eq. (6) holds in 46/46 design–scale cells, and $\widehat{M}_\theta$ is boundary-determined across this design family (fig. 4(b)). The thinning calculation uses a convergence-checked numerical candidate $\widehat{M}_\theta$; exact-rejection certification requires the almost-sure bound $w \le \widehat{M}_\theta$ under P_θ. Thus direct task matching removes design dependence at fixed labels, while source-dependent rejection converts it into proposal complexity.

5.2 Pre-Training Source Allocation

A second decision is how to split a fixed collection budget across sources before training. Given K sources with mass fields $m_1, \dots, m_K$ at a fixed scale h, let $\alpha \in \Delta_K := \{\alpha \in \mathbb{R}_+^K : \sum_i \alpha_i = 1\}$ choose the source of each example independently. Then $m_\alpha = \sum_i \alpha_i m_i$ is affine in α and can be evaluated before model fitting; Q enters only through the objective. Under strict finite-scale common support—for some $\eta_h > 0$, $\eta_h \le m_i(z,h) \le 1$ for Q-a.e. z and every i— define $\Psi(\alpha) := \int dQ/m_\alpha$ and the marginal coverage value $A_i(\alpha) := \int m_i/m_\alpha^2 \, dQ$. A source is active at α when $\alpha_i > 0$. If e_i is the i-th vertex of Δ_K, then $\Gamma_{b\to i} := \Psi(e_b) - A_i(e_b)$ is the directional derivative at e_b toward e_i; source i is locally actionable relative to b exactly when $\Gamma_{b\to i} < 0$.

Proposition 2 (Convex profile allocation and actionability). *[strict finite-scale common support]* (a) *Mixture weights average mass fields, $m_\alpha = \sum_i \alpha_i m_i$, and the local count is exactly* $\text{Binomial}(n, m_\alpha(z,h))$*; hence Ψ is convex and C^1 on Δ_K with $\partial_i \Psi = -A_i(\alpha)$ and $\sum_i \alpha_i A_i(\alpha) = \Psi(\alpha)$.* (b) *$\alpha^\star$ is optimal iff $A_i(\alpha^\star) = \Psi(\alpha^\star)$ for every active source and $A_i(\alpha^\star) \le \Psi(\alpha^\star)$ for every inactive source. A source vertex e_b is optimal iff no inactive source offers a strictly larger marginal coverage value ($\Gamma_{b\to i} \ge 0$ for all i); pointwise dominance is sufficient but not necessary.* (c) *The Frank–Wolfe (FW) gap $g_\Psi(\alpha) := \max_i A_i(\alpha) - \Psi(\alpha)$ satisfies $\Psi(\alpha) - \min_{\Delta_K} \Psi \le g_\Psi(\alpha)$, and $g_\Psi(\alpha) = 0$ iff α is optimal.*

Interpretation. The budget can be allocated before training by a convex program over the mass fields. At an optimum, active sources share the same marginal coverage value (b), while the FW gap (c) upper-bounds suboptimality. For two sources, write $m_{12,\alpha} := \alpha m_1 + (1-\alpha) m_2$ and $\Psi_{12}(\alpha) := \Psi(\alpha e_1 + (1-\alpha) e_2)$ for the restriction to the segment $\alpha \in [0,1]$. The mixture gap is

then exact (corollary 7):

$$\alpha\,\Psi(e_1) + (1-\alpha)\,\Psi(e_2) - \Psi_{12}(\alpha) = \alpha(1-\alpha)\int \frac{(m_1 - m_2)^2}{m_1 m_2\, m_{12,\alpha}}\, dQ.$$

The gain therefore depends on the cross-source arrangement of local coverage. Equal endpoint profiles with mass fields differing on a set of positive Q-measure make every interior mixture strictly better than either endpoint. Pointwise crossing alone does not guarantee improvement over the better endpoint. Proofs and convexity details are in section B.4.2.

Allocation evidence. Proposition 2 concerns fixed-scale Ψ. We compare its allocation rule with a finite-n rule that minimizes the target-free finite-(n, h) objective $\mathcal{J}_{n,h}(\alpha) := \int [B^2(1 - m_\alpha)^n + \sigma^2 g_n(m_\alpha)]\, dQ$. This objective is convex on Δ_K when $B^2 \geq \frac{3}{2}\sigma^2$, as in our experiments (proposition 8). At fixed h, it is the allocation-dependent part of the certificate and equals the exact zero-fallback NW risk for $f \equiv B$ under homoscedastic variance σ^2 (section C.5). Uniformly accurate mass-field approximations admit explicit error bounds for both Ψ and $\mathcal{J}_{n,h}$ (proposition 9).

We summarize performance by a family-balanced risk-ratio score: the equal-weight average of the within-bank median allocation-to-baseline risk ratios for the bump and smooth target banks. Across the 8 (h, n) settings, the score relative to uniform mixing ranges from 0.953 to 0.966 for the Ψ rule and from 0.955 to 0.971 for the finite-n rule. Both rules score below one against uniform mixing in every setting and against each single-source baseline in at least 7/8 settings. Numerical Wasserstein-1 optimization over the full source-mixture simplex returns the scalar-selected source (section C.3). In this family, these comparisons favor allocating regression data by finite-scale sample availability rather than Wasserstein proximity. Numerical-stability and regret bounds are reported in section C.3.

For complementarity, we rank all 253 source pairs on four randomized quasi-Monte Carlo (RQMC) selection scrambles and evaluate the top five on four disjoint scrambles. Two pairs meet the pre-specified two-scale criteria. The selection-rank-1 pair (fig. 4(c)) attains minimum evaluation gains over the better endpoint of $3.6\%/3.1\%$ for the selection-derived Ψ-optimal mixture at $h = 0.02/0.05$. For the same pair at $h = 0.02$, separately optimized constant-target finite-n mixtures yield minimum evaluation gains of $20.4\%/10.6\%$ over the better endpoint at $n = 100/200$. Averaged over the evaluation scrambles, reduced empty-neighborhood loss accounts for $98\%/73\%$ of the corresponding gain (section C.4).

6 Coverage Signals Beyond the Controlled NW Setting

To assess whether the coverage score remains informative beyond exact NW, we study 8 synthetic seven-joint arm-like variants sharing an acquisition law and deployment region (section D). Three target fields and $n \in \{200, 800\}$ yield six kernel ridge regression (KRR) settings, scored before fitting by $\Psi(h^\star)$. The label-free scale $h^\star = 0.10\,\mathrm{m}$ is the smallest candidate whose preliminary mass estimates have minimum at least 10^{-4} and median expected count at least three at reference budget 200, across variants and coverage-grid points (section C.5). For the primary comparisons, KRR uses one configuration selected on a separate calibration field. The score–risk association is positive in all 6/6 settings, with 4/6 paired-bootstrap 95% intervals excluding zero. At $n = 800$, the primary correlations are 0.71–0.79; all three intervals lie above zero under both the primary and expanded-grid configurations (section D). The numerically Ψ-optimal allocation, which places all mass on one variant, has lower mean KRR risk than uniform mixing in 5/6 settings under the primary configuration, with one interval excluding zero. At $n = 200$, we also study a downstream decision: among 256 fixed candidates, the learner selects the one with smallest predicted value. Normalized regret is its excess true target value over the candidate minimum, divided by the candidate-set target range. All three paired-bootstrap 95% intervals for the median normalized-regret difference (uniform mixing minus the Ψ-optimal allocation) include zero; decision quality depends jointly on candidate-local prediction error and the decision margin, the target-value gap from the optimum to the best suboptimal candidate (section B.5).

With the forward map fixed and the static acquisition law varied, median risk ratios of the Ψ-optimal mixture to equal mixing over the six target–budget cells are 0.92 for NW, 0.87 for KRR, 0.99 for random-feature ridge (RF), and 1.00 for a multilayer perceptron (MLP). The MLP response varies across targets and does not show a consistent improvement under the pre-specified criteria; the ideal Q-matched control has lower aggregate point estimates than equal mixing for all four learners (section D.1). The KRR results support pre-training regression-risk screening beyond NW: at the

larger budget, $\Psi(h^\star)$ remains informative across both calibrated configurations. The acquisition-law study shows that gains from reallocating coverage can depend on the learner and target.

7 Related Work

Covariate-shift theory: source laws are given. Covariate shift fixes the conditional law while source and deployment marginals differ (Shimodaira, 2000). Global analyses use divergences or density ratios (Ben-David et al., 2010; Mansour et al., 2009; Cortes et al., 2010); local analyses use transfer exponents (Kpotufe & Martinet, 2021) or likelihood-ratio-controlled KRR (Ma et al., 2023). Closest to our setting is Pathak et al. (2022): their reciprocal ball-probability functional $\int P(\mathbb{B}_h(z))^{-1}\, dQ(z)$ equals $\Psi_\theta(h)$ for $P = P_\theta$, and their analysis provides nonasymptotic NW bounds and matching rates over source–target classes. We instead derive $P_\theta = (T_\theta)_{\#}\mu$ from a constrained generator, obtain a finite-n certificate with design-uniform constants and controlled-family matching, and retain the full field $m_\theta(\cdot, h)$ for rigidity, acquisition, and allocation.

Data selection, source mixtures, and generator control. Given-source adaptation combines fixed sources or predictors using target samples, discrepancy bounds, or validation (Crammer et al., 2008; Mansour et al., 2008; Hoffman et al., 2018; Faw et al., 2020). Large-scale mixture and selection methods choose source weights from pilot runs, fitted performance laws, proxy or validation objectives, online or deployment signals, or dataset representations (Hashimoto, 2021; Ye et al., 2025; Liu et al., 2025; Li et al., 2025; Magnusson et al., 2025; Xie et al., 2023a; Fan et al., 2024; Tan et al., 2026; Chen et al., 2025; Jiang et al., 2025; Haghtalab et al., 2022; Chen et al., 2026; Zhang et al., 2025; Xie et al., 2023b); robot-imitation curation likewise uses downstream performance or datamodel signals (Hejna et al., 2025; Agia et al., 2025; Dass et al., 2026). Generator-adaptation methods tune simulation parameters from target data or downstream task signals (Veeravasarapu et al., 2017; Ruiz et al., 2019; Mishra et al., 2022). Wasserstein distance under the task metric and maximum mean discrepancy with a task-metric kernel provide distribution-alignment objectives, while core-set methods select covering subsets from an unlabeled candidate pool (Vallender, 1974; Gretton et al., 2012; Sener & Savarese, 2018). Our decision variable is instead a mixture over source laws realizable by specified generators and acquisition protocols; its weights are chosen before model fitting from finite-scale mass fields, which also enter the risk certificate.

Design and forward-map geometry. Our allocation analysis draws on optimal design over design measures, including equivalence and vertex-direction results, nonparametric design densities, Frank–Wolfe certificates, and push-forward design (Kiefer & Wolfowitz, 1960; Whittle, 1973; Kiefer, 1974; Wynn, 1970; Wu & Wynn, 1978; Fedorov, 1972; Silvey, 1980; Pukelsheim, 1993; Cheng et al., 1998; Hall et al., 1997; Frank & Wolfe, 1956; Jaggi, 2013; Butler et al., 2020). Here, however, the mixture is over mechanism-realizable source laws and the objective is finite-scale reciprocal occupancy under a fixed deployment law. In robotics, manipulability and workspace metrics characterize forward-map geometry (Yoshikawa, 1985; Gosselin & Angeles, 1991; Park & Brockett, 1994; van den Doel & Pai, 1996; Dong et al., 2013); we treat the forward map as a data-generating map, with the coarea formula connecting its geometry to the induced source law, while theorem 2 shows that Jacobian- and density-value distributions need not determine the finite-scale mass field.

8 Scope and Conclusion

Scope. We study the upstream sampling layer of mechanism-induced covariate shift: a structured generator induces the training marginal, while the deployment law, task metric, and conditional law are fixed. This isolates how generator-side choices shape finite-budget local coverage before learner-side feedback is available. Our controlled experiments instantiate this generator-to-coverage view with analytically specified forward maps into one- and two-dimensional task spaces. For task matching, proposal complexity is measured in source acquisitions, the resource consumed by rejection-based collection.

Conclusion. The finite-scale local-mass field provides the generator-side link between finite-budget learning and pre-training data decisions under mechanism-induced covariate shift. It exposes where source mass lies around deployment points, connects this local availability to finite-budget risk, and obeys exact relations under task matching, rejection sampling, and source mixing. It therefore serves as a common pre-training interface for assessing where data will be available, how coverage constrains task-matching cost, and how a collection budget can be allocated across realizable sources.

AI USE STATEMENT

Generative AI tools were used to aid and polish the writing of this paper. The authors reviewed all AI-assisted text and take responsibility for the final content.

REPRODUCIBILITY STATEMENT

The appendix provides experimental settings and implementation details supporting the reported results. Code and scripts needed to reproduce the experiments will be made publicly available upon acceptance.

REFERENCES

Christopher Agia, Rohan Sinha, Jingyun Yang, Rika Antonova, Marco Pavone, Haruki Nishimura, Masha Itkina, and Jeannette Bohg. CUPID: Curating data your robot loves with influence functions. In *Proceedings of The 9th Conference on Robot Learning (CoRL)*, volume 305 of *Proceedings of Machine Learning Research*, pp. 2907–2932, 2025.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1–2):151–175, 2010. doi: 10.1007/s10994-009-5152-4.

Troy Butler, John D. Jakeman, and Timothy Wildey. Optimal experimental design for prediction based on push-forward probability measures. *Journal of Computational Physics*, 416:109518, 2020. doi: 10.1016/j.jcp.2020.109518.

Mayee F. Chen, Michael Y. Hu, Nicholas Lourie, Kyunghyun Cho, and Christopher Ré. Aioli: A unified optimization framework for language model data mixing. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/71cc3acb2578bdd1e52bf9f85bbb1af0-Abstract-Conference.html.

Zhiliang Chen, Gregory Kang Ruey Lau, Chuan-Sheng Foo, and Bryan Kian Hsiang Low. DUET: Optimizing LLM training data mixtures via noisy feedback from unseen, downstream evaluation tasks. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Ming-Yen Cheng, Peter Hall, and D. Michael Titterington. Optimal design for curve estimation by local linear smoothing. *Bernoulli*, 4(1):3–14, 1998.

Corinna Cortes, Yishay Mansour, and Mehryar Mohri. Learning bounds for importance weighting. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pp. 442–450, 2010.

Koby Crammer, Michael Kearns, and Jennifer Wortman. Learning from multiple sources. *Journal of Machine Learning Research*, 9:1757–1774, 2008.

Shivin Dass, Alaa Khaddaj, Logan Engstrom, Aleksander Madry, Andrew Ilyas, and Roberto Martín-Martín. DataMIL: Selecting data for robot imitation learning with datamodels. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Hui Dong, Zhijiang Du, and Gregory S. Chirikjian. Workspace density and inverse kinematics for planar serial revolute manipulators. *Mechanism and Machine Theory*, 70:508–522, 2013.

Lawrence C. Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, revised edition, 2015.

Simin Fan, Matteo Pagliardini, and Martin Jaggi. DoGE: Domain reweighting with generalization estimation. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12895–12915, 2024.

Matthew Faw, Rajat Sen, Karthikeyan Shanmugam, Constantine Caramanis, and Sanjay Shakkottai. Mix and match: An optimistic tree-search approach for learning models from mixture distributions. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.

Valerii V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1–2):95–110, 1956. doi: 10.1002/nav.3800030109.

Clément Gosselin and Jorge Angeles. A global performance index for the kinematic optimization of robotic manipulators. *Journal of Mechanical Design*, 113(3):220–226, 1991. doi: 10.1115/1.2912772.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL https://jmlr.org/papers/v13/gretton12a.html.

László Győrfi, Michael Kohler, Adam Krzyżak, and Harro Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer, New York, 2002. doi: 10.1007/b97848.

Nika Haghtalab, Michael I. Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, 2022.

Peter Hall, J. Steve Marron, Michael H. Neumann, and D. Michael Titterington. Curve estimation when the design density is low. *The Annals of Statistics*, 25(2):756–770, 1997. doi: 10.1214/aos/1031833672.

Tatsunori Hashimoto. Model performance scaling with multiple data sources. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4107–4116, 2021.

Joey Hejna, Chethan Anand Bhateja, Yichen Jiang, Karl Pertsch, and Dorsa Sadigh. ReMix: Optimizing data mixtures for large scale imitation learning. In *Proceedings of The 8th Conference on Robot Learning (CoRL)*, volume 270 of *Proceedings of Machine Learning Research*, pp. 145–164, 2025.

Judy Hoffman, Mehryar Mohri, and Ningshan Zhang. Algorithms and theory for multiple-source adaptation. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, pp. 8256–8266, 2018.

Martin Jaggi. Revisiting Frank–Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28 of *Proceedings of Machine Learning Research*, pp. 427–435, 2013.

Yiding Jiang, Allan Zhou, Zhili Feng, Sadhika Malladi, and J. Zico Kolter. Adaptive data optimization: Dynamic sample selection with scaling laws. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

Jack Kiefer. General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 2(5):849–879, 1974. doi: 10.1214/aos/1176342810.

Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi: 10.4153/CJM-1960-030-4.

Samory Kpotufe and Guillaume Martinet. Marginal singularity and the benefits of labels in covariate-shift. *The Annals of Statistics*, 49(6):3299–3323, 2021. doi: 10.1214/21-AOS2084.

Yuan Li, Zhengzhong Liu, and Eric Xing. Data mixing optimization for supervised fine-tuning of large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 35419–35437, 2025.

Qian Liu, Xiaosen Zheng, Niklas Muennighoff, Guangtao Zeng, Longxu Dou, Tianyu Pang, Jing Jiang, and Min Lin. RegMix: Data mixture as regression for language model pre-training. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

Cong Ma, Reese Pathak, and Martin J. Wainwright. Optimally tackling covariate shift in RKHS-based nonparametric regression. *The Annals of Statistics*, 51(2):738–761, 2023. doi: 10.1214/23-AOS2268.

Ian Magnusson, Nguyen Tai, Ben Bogin, David Heineman, Jena D. Hwang, Luca Soldaini, Akshita Bhagia, Jiacheng Liu, Dirk Groeneveld, Oyvind Tafjord, Noah A. Smith, Pang Wei Koh, and Jesse Dodge. DataDecide: How to predict best pretraining data with small experiments. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 42487–42502, 2025.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation with multiple sources. In *Advances in Neural Information Processing Systems 21*, pp. 1041–1048, 2008.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Multiple source adaptation and the Rényi divergence. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 367–374, 2009.

Samarth Mishra, Rameswar Panda, Cheng Perng Phoo, Chun-Fu Richard Chen, Leonid Karlinsky, Kate Saenko, Venkatesh Saligrama, and Rogério Schmidt Feris. Task2Sim: Towards effective pre-training and transfer from synthetic data. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9184–9194, 2022. doi: 10.1109/CVPR52688.2022.00898.

Elizbar A. Nadaraya. On estimating regression. *Theory of Probability and Its Applications*, 9(1): 141–142, 1964. doi: 10.1137/1109020.

Frank C. Park and Roger W. Brockett. Kinematic dexterity of robotic mechanisms. *The International Journal of Robotics Research*, 13(1):1–15, 1994. doi: 10.1177/027836499401300101.

Reese Pathak, Cong Ma, and Martin J. Wainwright. A new similarity measure for covariate shift with applications to nonparametric regression. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17517–17530, 2022.

Friedrich Pukelsheim. *Optimal Design of Experiments*. Wiley, New York, 1993.

Nataniel Ruiz, Samuel Schulter, and Manmohan Chandraker. Learning to simulate. In *International Conference on Learning Representations*, 2019.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations (ICLR)*, 2018.

Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.

Samuel D. Silvey. *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Chapman & Hall, London, 1980.

Haoru Tan, Sitong Wu, Yanfeng Chen, Jun Xia, Ruobing Xie, Bin Xia, Xingwu Sun, and Xiaojuan Qi. Fast data mixture optimization via gradient descent. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/01d845baaa31b1043e5f280592d57dc4-Abstract-Conference.html.

S. S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786, 1974. doi: 10.1137/1118101.

Kees van den Doel and Dinesh K. Pai. Performance measures for robot manipulators: A unified approach. *The International Journal of Robotics Research*, 15(1):92–111, 1996. doi: 10.1177/027836499601500106.

V. S. R. Veeravasarapu, Constantin A. Rothkopf, and Visvanathan Ramesh. Adversarially tuned scene generation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6441–6449, 2017. doi: 10.1109/CVPR.2017.682.

Geoffrey S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

Peter Whittle. Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society. Series B (Methodological)*, 35(1):123–130, 1973. doi: 10.1111/j.2517-6161.1973.tb00944.x.

Chien-Fu Wu and Henry P. Wynn. The convergence of general step-length algorithms for regular optimum design criteria. *The Annals of Statistics*, 6(6):1273–1285, 1978. doi: 10.1214/aos/1176344373.

Henry P. Wynn. The sequential generation of D-optimum experimental designs. *The Annals of Mathematical Statistics*, 41:1655–1664, 1970. doi: 10.1214/aoms/1177696809.

Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy Liang, Quoc V. Le, Tengyu Ma, and Adams Wei Yu. DoReMi: Optimizing data mixtures speeds up language model pretraining. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pp. 69798–69818, 2023a.

Sang Michael Xie, Shibani Santurkar, Tengyu Ma, and Percy Liang. Data selection for language models via importance resampling. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pp. 34201–34227, 2023b.

Jiasheng Ye, Peiju Liu, Tianxiang Sun, Jun Zhan, Yunhua Zhou, and Xipeng Qiu. Data mixing laws: Optimizing data mixtures by predicting language modeling performance. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

Tsuneo Yoshikawa. Manipulability of robotic mechanisms. *The International Journal of Robotics Research*, 4(2):3–9, 1985. doi: 10.1177/027836498500400201.

Mozhi Zhang, Howe Tissue, Lu Wang, and Xipeng Qiu. Domain2Vec: Vectorizing datasets to find the optimal data mixture without training. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 76139–76158, 2025.

A Notation, Assumptions, and Theorem Map

This appendix collects the principal symbols (section A.1), explains the bracketed assumptions attached to the formal statements (section A.2), and summarizes how the main results fit together (section A.3).

A.1 Symbol table

Table 1 collects the principal symbols used throughout the main text and appendix, grouped by role.

A.2 Assumptions and scope

The bracketed phrases in the main statements summarize the assumptions under which each result applies.

> [uniform constants; fixed design]: the bound holds for every design with numerical constants independent of the design; design dependence enters only through m_θ.
>
> [controlled 1D; $n \geq n_0$]: the controlled one-dimensional family of section 3 on $[0, 1]$, with source densities in $[1/C, 1/c]$ as induced by a uniform source law and monotone C^1 bijections satisfying $0 < c \leq |T'_\theta| \leq C$, Gaussian noise with $\sigma > 0$, $\Lambda, B > 0$, and deployment density in $[q_{\min}, q_{\max}]$ with $q_{\min} > 0$. The result applies for $n \geq n_0$, with constants shared across the family.

Table 1: Principal symbols used throughout the paper.

Symbol	Type	Meaning
θ	design	index of a structured data generator
$\mathcal{S}$	space	source (configuration) space; a measurable space in the statistical setup and a compact $C^{r_{\mathrm{reg}}}$ manifold or flat box in the geometric sections
μ, ϱ	law, density	acquisition law on the source space and its density with respect to volume
T_θ	map	forward map $\mathcal{S} \to \mathcal{Z}$
$\mathcal{Z}$, $d_{\mathcal{Z}}$, d	space, metric, integer	task space, its metric, and $d = \dim \mathcal{Z}$
$P_\theta = (T_\theta)_{\#}\mu$	law	source distribution induced in task space
Q	law	deployment distribution, fixed across the designs being compared
$\mathbb{B}_h(z)$	set	closed task-space ball of radius h around z
$m_\theta(z,h) = P_\theta(\mathbb{B}_h(z))$, m_θ	field	finite-scale local mass; m_θ denotes the field $(z,h) \mapsto m_\theta(z,h)$
$m_Q(z,h) = Q(\mathbb{B}_h(z))$	field	finite-scale mass field of the deployment law
$\Psi_Q(h) = \int dQ(z)/m_Q(z,h)$	scalar	coverage profile of the deployment law
$N_h(z)$	r.v.	number of the n training covariates in $\mathbb{B}_h(z)$; $N_h(z) \sim \mathrm{Binomial}(n, m_\theta(z,h))$
$g_n(m) = \mathbb{E}[1/N; N > 0]$, $N \sim \mathrm{Binomial}(n,m)$	function	expected reciprocal count on a nonempty ball; $g_0 \equiv 0$
$\Psi_\theta(h) = \int dQ(z)/m_\theta(z,h)$	scalar	deployment-weighted finite-scale coverage profile
p_θ, J_{T_θ}, crit	density, Jacobian, set	task-space source density under (Co1)–(Co3) of section B.3, when it exists; normal Jacobian $\sqrt{\det(DT_\theta DT_\theta^*)}$; critical set
f, $\mathcal{F}$, Λ, β, B	target	regression function and Hölder class with constant Λ, exponent β, and sup-norm bound B
ε, σ^2	noise	conditionally centered noise with conditional variance at most σ^2
n, h, $\hat{f}_{n,h}$	integer, scalar, estimator	sample size, neighborhood scale, and zero-fallback uniform-ball Nadaraya–Watson (NW) estimator
$\mathcal{C}_\theta(n)$	scalar	optimized finite-sample certificate in eq. (3)
ξ, Φ_+, $Q_{1/2}$	auxiliary	quantities used in the controlled-family lower bound and defined in section B.2
κ_1, κ_2, n_0	constants	family-uniform matching constants and sample-size threshold
a, x, J_x, p_a	2R	second link length, normalized coordinate, normalized Jacobian, and density of x in the two-link construction
L_j	3R	link lengths in the normalized three-link family ($\sum_j L_j = 1$)
$w = dQ/dP_\theta$, M_θ	acquisition	deployment-to-source likelihood ratio and its essential supremum (the smallest valid constant rejection envelope)
$D_\theta(h)$	acquisition	finite-scale ball-mass domination coefficient $\operatorname{ess\,sup}_{z\sim Q} m_Q(z,h)/m_\theta(z,h)$
$\widehat{D}_\theta(h)$, $\widehat{M_\theta}$	numerical	numerical grid estimate of $D_\theta(h)$ and candidate rejection envelope obtained from the numerical density-ratio scan; exact rejection theory uses M_θ
N_{prop}, N_{acc}	acquisition	proposal budget and number of accepted proposals
K, $\alpha \in \Delta_K$, m_α	allocation	number of sources, mixture weights, and mixed mass field
$A_i(\alpha)$, $\Gamma_{b\to i}$	allocation	marginal coverage value $\int m_i(z,h)/m_\alpha(z,h)^2\, dQ(z)$ and vertex directional derivative $\Psi(e_b) - A_i(e_b)$
$r_n(m)$, $\mathcal{J}_{n,h}(\alpha)$	allocation	$r_n(m) = B^2(1-m)^n + \sigma^2 g_n(m)$ and the finite-(n,h) mass-field objective $\mathcal{J}_{n,h}(\alpha) = \int r_n(m_\alpha(z,h))\, dQ(z)$
g_Ψ, $g_{\mathcal{J}}$, $\hat{g}_\Psi$, $\hat{g}_{\mathcal{J}}$	allocation	objective-specific Frank–Wolfe (FW) gaps and their numerical counterparts
$\widetilde{m}_i$, δ_m	allocation	approximate mass field and its uniform approximation error in proposition 9
η_h	scalar	uniform positive-mass level at scale h

[design-normalized 2R]: the two-link family is compared in a design-dependent normalized coordinate. The result is a structural counterexample and does not constitute a comparison under one shared deployment law.

[strict finite-scale common support]: at the fixed scale h, there exists $\eta_h > 0$ such that $\eta_h \le m_i(z,h) \le 1$ for Q-a.e. z and every source.

[$Q \ll P_\theta$; finite envelope where stated]: the deployment law is absolutely continuous with respect to the source law. Finiteness of M_θ is required only for statements that invoke a finite-envelope rejection sampler.

A.3 Theorem map

The formal results share the statistical setup of section 2. Section B.1 formalizes the relationality principle: with the conditional model and deployment law fixed, the generator affects the statistical experiment through its induced source law. This motivates the finite-scale mass field m_θ and its scalar profile Ψ_θ as the common objects used throughout the paper.

For prediction risk, section B.2 proves the finite-sample certificate in theorem 1(a), develops fixed-design lower bounds on the controlled family, and combines them to obtain the controlled-family sharpness statement in theorem 1(b). For geometry, section B.3 proves theorem 2, while the coarea representation eq. (10) explains how forward-map fibers and Jacobians induce the task-space source density and finite-scale mass.

For acquisition and allocation, section B.4 proves the task-matching and rejection-cost result proposition 1 and the convex allocation result proposition 2; the latter yields the exact mixture-gap statement corollary 7, and proposition 9 additionally quantifies the stability of the allocation objectives under uniformly accurate mass-field approximations. Complementary boundary results are given by the learner-independent lower bounds proposition 10 and corollary 8 in section B.6, while lemma 11 in section B.5 explains why lower regression risk need not imply a better downstream argmin decision.

B Proofs

Throughout, $\sup$ and $\inf$ over risks are taken in $[0,+\infty]$, and all integrals of nonnegative functions are understood in the extended reals. We write $\mathbb{E}$ for expectation and $\mathbb{P}$ for probability.

B.1 Relational Equivalence

This subsection formalizes the relationality principle of section 2. With the conditional model, target class, estimator class, sample size, and deployment law fixed, a design affects the statistical experiment only through its induced task-space source law.

Statistical model. Let

$$E = (P, K_\varepsilon, \ell_Q, \mathcal{F}, n)$$

denote a statistical model. Here P is the task-space source law, $K_\varepsilon(\cdot \mid z)$ is the conditional noise kernel,

$$Y \mid Z = z \sim f(z) + \varepsilon, \qquad \varepsilon \sim K_\varepsilon(\cdot \mid z), \qquad \mathbb{E}[\varepsilon \mid Z = z] = 0,$$

and the same conditional noise kernel is used across designs; heteroscedasticity in z is allowed. The deployment loss is

$$\ell_Q(g,f) := \int (g-f)^2 \, dQ \in [0,+\infty],$$

$\mathcal{F}$ is the target class, and n is the sample size.

For $f \in \mathcal{F}$, write $K_{f,\varepsilon}(\cdot \mid z)$ for the conditional law of $f(z)+\varepsilon$ and

$$S_{E,f} := (P \otimes K_{f,\varepsilon})^{\otimes n}$$

for the sample law. Thus $S_{E,f}$ is determined by (P, f, K_ε, n); Q enters only through the loss. An estimator $\hat{f}_n$ is a jointly measurable map from samples to functions on $\mathcal{Z}$ and does not depend on

the design. Define

$$R_E(\hat{f}_n, f) := \mathbb{E}_{S_{E,f}}\Big[\ell_Q(\hat{f}_n, f)\Big].$$

Assume $\mathcal{F}$ is equipped with a σ-algebra for which $f \mapsto R_E(\hat{f}_n, f)$ is measurable for every estimator. For a prior π on $\mathcal{F}$, define the integrated risk $r_\pi^E(\hat{f}_n) := \int R_E(\hat{f}_n, f)\, d\pi(f)$ and the Bayes risk $r_\pi^*(E) := \inf_{\hat{f}_n} r_\pi^E(\hat{f}_n)$. For the Hölder classes used here, the sup-norm Borel σ-algebra satisfies this measurability requirement, because the risk is built from pointwise operations and integrals.

Proposition 3 (Same-marginal equivalence)**.** *Let $E_j = (P_j, K_\varepsilon, \ell_{Q_j}, \mathcal{F}, n)$, $j = 1, 2$. If $P_1 = P_2$ and $Q_1 = Q_2$, then*

$$R_{E_1}(\hat{f}_n, f) = R_{E_2}(\hat{f}_n, f)$$

for every estimator $\hat{f}_n$ and every $f \in \mathcal{F}$. Consequently, the minimax risks, the integrated risks of every fixed estimator under every prior, and the Bayes risks coincide.

Proof. The equality $P_1 = P_2$, together with the common K_ε and n, gives $S_{E_1,f} = S_{E_2,f}$ for every f, so every estimator has the same law under both models. The equality $Q_1 = Q_2$ gives $\ell_{Q_1} = \ell_{Q_2}$. Hence the risks agree pointwise in $(\hat{f}_n, f)$, and the remaining functionals are suprema, integrals, or infima of this pointwise equality over identical index sets. □

Equality of the source laws alone is not sufficient if the deployment laws differ: the sample laws then coincide, but the losses need not. The relationality principle therefore concerns the induced source law relative to a common deployment law.

Corollary 1 (Source-space transport)**.** *Let two designs have source spaces and acquisition laws $(\mathcal{S}^{(1)}, \mu_1)$ and $(\mathcal{S}^{(2)}, \mu_2)$, with forward maps $T_i : \mathcal{S}^{(i)} \to \mathcal{Z}$. Suppose there is a measurable map $\psi : \mathcal{S}^{(1)} \to \mathcal{S}^{(2)}$ such that*

$$\mu_2 = \psi_\# \mu_1, \qquad T_2 \circ \psi = T_1.$$

Then $P_2 = P_1$. With a common task metric, $m_2(z, h) = m_1(z, h)$ for every (z, h); with a common deployment law, $\Psi_2(h) = \Psi_1(h)$ as extended reals; and with common $(K_\varepsilon, \ell_Q, \mathcal{F}, n)$ all risk functionals coincide by proposition 3. The map ψ need not be bijective.

Proof. For every measurable $A \subset \mathcal{Z}$,

$$(T_2)_\#(\psi_\#\mu_1)(A) = \mu_1\big(\psi^{-1}(T_2^{-1}(A))\big) = \mu_1\big((T_2 \circ \psi)^{-1}(A)\big) = (T_1)_\#\mu_1(A).$$

Thus $P_2 = P_1$. Evaluating this equality on task-space balls of the common metric gives the mass-field identity, and integrating the pointwise equal integrands against the common deployment law gives the profile identity. □

The folding of the two-link radial map and the torus automorphism of the three-link map in section B.3 are examples: although the source-space parameterizations differ, they induce the same task-space source law and hence the same statistical problem under the common model ingredients above.

Proposition 4 (Reparametrization correspondence)**.** *Let $\varphi : \mathcal{Z} \to \mathcal{Z}$ be a measurable bijection with measurable inverse. Given*

$$E_1 = (P_1, K_\varepsilon^{(1)}, \ell_{Q_1}, \mathcal{F}_1, n),$$

define

$$P_2 := \varphi_\# P_1, \qquad Q_2 := \varphi_\# Q_1, \qquad \mathcal{F}_2 := \{g \circ \varphi^{-1} : g \in \mathcal{F}_1\},$$

and

$$K_\varepsilon^{(2)}(\cdot \mid w) := K_\varepsilon^{(1)}(\cdot \mid \varphi^{-1}(w)).$$

Equip $\mathcal{F}_2$ with the σ-algebra transported from $\mathcal{F}_1$ by $\Phi_\mathcal{F} : g \mapsto g \circ \varphi^{-1}$. If $\mathcal{F}_1$ carries its sup-norm Borel σ-algebra and $\mathcal{F}_2$ the subspace sup-norm topology, then $\Phi_\mathcal{F}$ is a sup-norm isometry onto $\mathcal{F}_2$, hence a Borel isomorphism, and the transported σ-algebra is the Borel one.

For an estimator $\hat{f}_1$, define

$$(\Phi_{\mathrm{est}}\hat{f}_1)\big(w; (W_i, Y_i)_{i=1}^n\big) := \hat{f}_1\big(\varphi^{-1}(w); (\varphi^{-1}(W_i), Y_i)_{i=1}^n\big).$$

Then Φ_{est} is a bijection between the estimator sets, $\Phi_\mathcal{F}$ is a bimeasurable bijection $\mathcal{F}_1 \to \mathcal{F}_2$, and for every $f_1 \in \mathcal{F}_1$ with $f_2 := \Phi_\mathcal{F} f_1$:

(a) $R_{E_2}(\Phi_{\mathrm{est}}\hat{f}_1, f_2) = R_{E_1}(\hat{f}_1, f_1)$*;*

(b) the minimax risks coincide;

(c) for every prior π_1 *and* $\pi_2 := (\Phi_{\mathcal{F}})_{\#}\pi_1$*, the integrated risks of corresponding estimators and the Bayes risks coincide.*

Proof. Let $W_i := \varphi(Z_i^{(1)})$. Then (W_i) are i.i.d. from P_2. Because φ is a bimeasurable bijection, conditioning on $W_i = w$ is equivalent to conditioning on $Z_i^{(1)} = \varphi^{-1}(w)$, so

$$Y_i \mid W_i = w \sim f_2(w) + \varepsilon, \qquad \varepsilon \sim K_\varepsilon^{(2)}(\cdot \mid w).$$

Hence $(W_i, Y_i)_{i=1}^n$ has sample law S_{E_2,f_2}.

The transported estimator outputs $\hat{f}_1 \circ \varphi^{-1}$. Using

$$\int g \, d(\varphi_{\#}Q_1) = \int g \circ \varphi \, dQ_1$$

for nonnegative measurable g gives

$$\ell_{Q_2}(\Phi_{\mathrm{est}}\hat{f}_1, f_2) = \ell_{Q_1}(\hat{f}_1, f_1),$$

and taking expectations proves (a). Joint measurability of $\Phi_{\mathrm{est}}\hat{f}_1$ follows from that of $\hat{f}_1$ composed with the measurable coordinate map, and the inverse correspondence $\hat{f}_1(z; (Z_i, Y_i)) := \hat{f}_2(\varphi(z); (\varphi(Z_i), Y_i))$ is obtained by replacing φ^{-1} with φ, so Φ_{est} is bijective. For (b), by (a) and bijectivity of $\Phi_{\mathcal{F}}$ the two sets of risks coincide as subsets of $[0, +\infty]$, hence so do their suprema; bijectivity of Φ_{est} then gives equal infima. Part (c) follows from (a) and the pushforward identity for π_2, with Bayes risks obtained by taking infima over corresponding estimators. □

An important distinction is that $\mathcal{F}_2$ need not equal $\mathcal{F}_1$: a non-isometric task-space reparametrization generally changes a Hölder class. The proposition therefore establishes equivalence between transported statistical problems; unless φ is an isometry of the task metric (or the metric is transported with it), it does not give invariance of m_θ or Ψ_θ at a fixed numerical radius. This is why the cross-design comparisons in the paper use a common task coordinate, metric, target class, and deployment law. The invariance in corollary 1 holds precisely because the task space and its metric are unchanged.

Corollary 2 (Task-matched acquisition)**.** *If every design is acquired with the same induced task-space law* $P = Q$*, then all designs define the same statistical model*

$$(Q, K_\varepsilon, \ell_Q, \mathcal{F}, n).$$

Thus the risk of any fixed estimator is design-independent, as are the minimax and Bayes risks.

Proof. Apply proposition 3 with $P_1 = P_2 = Q$ and $Q_1 = Q_2 = Q$. □

This equivalence concerns the fixed-n learning problem in which the learner observes only (Z, Y) and all designs share the conditional response and loss. Acquisition feasibility and proposal, interaction, labeling, or computation cost can still depend on the design; these effects are treated in section B.4.1. Importance weighting changes the estimator rather than the source law, so the equivalence above does not imply equality of importance-weighted risks across designs. Finally, if $Y \mid Z$ changes across designs, the common conditional kernel assumption fails and $S_{E,f}$ is no longer determined by (P, f, K_ε, n) with design-free f and K_ε. The equivalence therefore applies under the common-conditional covariate-shift model of section 2.

B.2 Certificate and Matching

This subsection proves theorem 1. We first derive the finite-sample NW certificate by an exact pointwise decomposition, then establish fixed-design lower bounds on the controlled one-dimensional family, and finally match the two at a common design-independent scale.

B.2.1 THE CERTIFICATE

Fix $h > 0$ and write $m(z) := m_\theta(z, h)$ and $N(z) := N_h(z) \sim \text{Binomial}(n, m(z))$. The zero-fallback uniform-ball NW estimator is

$$\hat{f}(z) = N(z)^{-1} \sum_{i: Z_i \in \mathbb{B}_h(z)} Y_i \quad \text{on } \{N(z) > 0\},$$

and $\hat{f}(z) = 0$ on $\{N(z) = 0\}$. Conditionally on $Z_{1:n}$, assume $\varepsilon_1, \ldots, \varepsilon_n$ are independent with $\mathbb{E}[\varepsilon_i \mid Z_{1:n}] = 0$ and $\mathbb{E}[\varepsilon_i^2 \mid Z_{1:n}] \le \sigma^2$; the i.i.d. conditional-response model of section 2 satisfies this condition.

Proof of theorem 1(a). **Step 1 (pointwise decomposition).** Fix z and let $I := \{i : Z_i \in \mathbb{B}_h(z)\}$ and $N := |I|$. On $\{N > 0\}$,

$$\hat{f}(z) - f(z) = A + \mathcal{N},$$

where

$$A := N^{-1} \sum_{i \in I} \big(f(Z_i) - f(z)\big), \qquad \mathcal{N} := N^{-1} \sum_{i \in I} \varepsilon_i.$$

The term A is $Z_{1:n}$-measurable, while the noise assumptions give

$$\mathbb{E}[\mathcal{N} \mid Z_{1:n}] = 0, \qquad \mathbb{E}[\mathcal{N}^2 \mid Z_{1:n}] = N^{-2} \sum_{i \in I} \mathbb{E}[\varepsilon_i^2 \mid Z_{1:n}] \le \frac{\sigma^2}{N}.$$

Hence the cross term vanishes and, on $\{N > 0\}$,

$$\mathbb{E}[(\hat{f}(z) - f(z))^2 \mid Z_{1:n}] \le A^2 + \frac{\sigma^2}{N}.$$

On $\{N = 0\}$, $(\hat{f}(z) - f(z))^2 = f(z)^2 \le B^2$. Taking expectations therefore yields

$$\mathbb{E}[(\hat{f}(z) - f(z))^2] \le B^2(1 - m(z))^n + \mathbb{E}[A^2; N > 0] + \sigma^2 g_n(m(z)). \tag{7}$$

Step 2 (bias). On $\{N > 0\}$ every $i \in I$ satisfies $d_{\mathcal{Z}}(Z_i, z) \le h$. Since $f \in \mathcal{F}$,

$$|f(Z_i) - f(z)| \le \Lambda h^\beta, \qquad |A| \le \Lambda h^\beta$$

sample by sample, and therefore

$$\mathbb{E}[A^2; N > 0] \le \Lambda^2 h^{2\beta} \mathbb{P}(N > 0) \le \Lambda^2 h^{2\beta}.$$

Step 3 (integration). All terms are nonnegative and measurable, so Tonelli's theorem applied to eq. (7) gives eq. (4). The source law enters this bound only through the local masses $m(z)$; the numerical constants are 1 and are independent of the design. When $m(z) = 0$, the neighborhood is empty almost surely and eq. (7) reduces exactly to B^2, so the certificate also covers zero-mass neighborhoods. □

A slightly sharper decomposition keeps the nonempty-ball probability:

$$\mathbb{E}[A^2; N > 0] \le \Lambda^2 h^{2\beta} \big(1 - (1 - m)^n\big).$$

The simpler uniform bias bound above is sufficient for the results below.

Lemma 1 (Scalar bound on the noise term)**.** *For $N \sim \text{Binomial}(n, m)$, $m \in (0, 1]$:*

$$g_n(m) \le \frac{2}{(n+1)m} \le \frac{2}{nm}.$$

Hence

$$\sigma^2 \int g_n(m_\theta)\, dQ \le \frac{2\sigma^2}{n} \Psi_\theta(h),$$

and eq. (4) *implies*

$$\sup_{f \in \mathcal{F}} \mathbb{E}\|\hat{f} - f\|^2 \le \Lambda^2 h^{2\beta} + \frac{2\sigma^2}{n} \Psi_\theta(h) + B^2 \int (1 - m_\theta)^n\, dQ.$$

The tighter bound $g_n(m) \le 1/(nm)$ is false in general.

Proof. Using $1/k \le 2/(k+1)$ and $\binom{n}{k}/(k+1) = \binom{n+1}{k+1}/(n+1)$,

$$\begin{aligned} g_n(m) &\le \frac{2}{(n+1)m} \sum_{k=1}^{n} \binom{n+1}{k+1} m^{k+1}(1-m)^{n-k} \\ &= \frac{2}{(n+1)m}\Big[1 - (1-m)^{n+1} - (n+1)m(1-m)^n\Big] \\ &\le \frac{2}{(n+1)m}. \end{aligned}$$

The bound cannot in general be strengthened to $g_n(m) \le 1/(nm)$:

$$g_2(m) = 2m - \tfrac{3}{2}m^2, \qquad g_2(0.8) = 0.64 > 0.625 = \frac{1}{2 \cdot 0.8}.$$

□

If the noise is homoscedastic, $\mathbb{E}[\varepsilon_i^2 \mid Z_{1:n}] \equiv \sigma^2$, then Step 1 gives $\mathbb{E}[\mathcal{N}^2 \mid Z_{1:n}] = \sigma^2/N$ exactly. Thus the target-independent NW variance contribution is

$$F_{n,\theta}(h) = \sigma^2 \int g_n(m_\theta)\, dQ.$$

Lemma 1 gives a finite-n upper bound on this quantity; the next corollary identifies its fixed-scale asymptotic limit.

Corollary 3 (Fixed-scale profile limit)**.** *For each fixed $m \in (0,1]$, $ng_n(m) \to 1/m$ as $n \to \infty$. Consequently, if $\sigma > 0$ and $m_\theta(z,h) \ge \eta_h > 0$ for Q-a.e. z, then*

$$\frac{nF_{n,\theta}(h)}{\sigma^2} \to \Psi_\theta(h).$$

Proof. Let $B_1, B_2, \ldots$ be i.i.d. $\mathrm{Bernoulli}(m)$ on one probability space and set $N_n := \sum_{i \le n} B_i$. Then $N_n \sim \mathrm{Binomial}(n,m)$ for each n, and $N_n/n \to m$ almost surely by the strong law of large numbers.

On $\{N_n \ge nm/2\}$, $n/N_n \le 2/m$ and $n/N_n \to 1/m$ almost surely, so dominated convergence gives

$$\mathbb{E}[n/N_n; N_n \ge nm/2] \to 1/m.$$

On $\{0 < N_n < nm/2\}$, $n/N_n \le n$, while the Chernoff bound gives $\mathbb{P}(N_n < nm/2) \le e^{-nm/8}$. Hence this contribution is at most $ne^{-nm/8} \to 0$, proving $ng_n(m) \to 1/m$.

Under the uniform positive-mass condition, $ng_n(m_\theta(z,h)) \to 1/m_\theta(z,h)$ for Q-a.e. z, and lemma 1 gives the uniform bound $ng_n(m_\theta(z,h)) \le 2/\eta_h$. Dominated convergence over Q therefore yields

$$\int ng_n(m_\theta)\, dQ \to \int \frac{dQ}{m_\theta} = \Psi_\theta(h).$$

□

The uniform positive-mass condition is not required for the certificate itself. It ensures $\Psi_\theta(h) < \infty$ in lemma 1, justifies the fixed-scale limit above, and, together with the controlled density bounds below, permits the upper and lower bounds to be compared at a common scale. The matching argument below uses only the upper control on g_n from lemma 1.

B.2.2 Fixed-design lower bounds on the controlled family

Setting. Let $\mathcal{Z} = [0,1]$. The source law P_θ has density $p_\theta \in [c', C']$, where $c' = 1/C$ and $C' = 1/c$, as for a uniform source law on $[0,1]$ pushed through a monotone C^1 bijection satisfying $0 < c \le |T_\theta'| \le C$. The deployment law has density $q \le q_{\max}$, and the matching statement additionally assumes $q \ge q_{\min} > 0$. The noise variables are independent $\mathcal{N}(0, \sigma^2)$ with $\sigma > 0$, independent of Z, and $\mathcal{F}$ is the Hölder class of section 2 with $0 < \beta \le 1$, $\Lambda > 0$, and $B > 0$.

Fix a design-independent geometric scale $h_0 \in (0, 1]$ and write

$$m(z, h) = P_\theta([z - h, z + h]), \qquad Q_{1/2}(z, h) = Q([z - h/2, z + h/2]).$$

Let $b : \mathbb{R} \to [0, 1]$ satisfy $b(0) = 1$, $\operatorname{supp} b \subset [-1, 1]$, and be β-Hölder with constant C_b. Define

$$b_{\min}^2 := \min_{|u| \le 1/2} b(u)^2 > 0, \qquad C_b' := 2^{1-\beta} C_b, \qquad a := \frac{\Lambda h^\beta}{C_b'},$$

and the Kullback–Leibler (KL) parameter

$$\xi(z, h) := \frac{n\, m(z, h)\, a^2}{\sigma^2} = \frac{n\, m(z, h) \Lambda^2 h^{2\beta}}{C_b'^2 \sigma^2}, \qquad \Phi_+(\xi) := \left(1 - \frac{\sqrt{\xi}}{2}\right)_+.$$

For a target f, let $\mathbb{P}_f^n$ denote the joint law of $(Z_i, Y_i)_{i \le n}$, and use $\mathbb{P}_i^n := \mathbb{P}_{f_i}^n$ and $\mathbb{P}_\omega^n := \mathbb{P}_{f_\omega}^n$ below, with $\mathbb{E}_i, \mathbb{E}_\omega$ the corresponding expectations. Finally set

$$\bar{h} := \min\left\{h_0, \frac{1}{4}, \left(\frac{BC_b'}{\Lambda}\right)^{1/\beta}\right\}.$$

Whenever a pairwise KL divergence below is at most $\xi/2$, Pinsker's inequality gives $\mathrm{TV} \le \min(1, \sqrt{\xi}/2)$, where TV denotes the total variation distance between the two sample laws; hence

$$1 - \mathrm{TV} \ge \Phi_+(\xi) \ge 0.$$

Lemma 2 (Local two-point bound)**.** *Fix $z_0 \in (0, 1)$ and $0 < h \le \min\{h_0, z_0, 1 - z_0\}$ with $a \le B$. Then*

$$\inf_{\hat{f}} \max_{i \in \{0,1\}} \mathbb{E}_i \|\hat{f} - f_i\|_{L^2(Q)}^2 \ge \frac{b_{\min}^2}{4C_b'^2} \Lambda^2 h^{2\beta} Q_{1/2}(z_0, h) \Phi_+(\xi(z_0, h))$$

for $f_0 \equiv 0$ and $f_1 = a\, b((\cdot - z_0)/h)$; if $\xi \le 1$, the constant is $b_{\min}^2/(8C_b'^2)$. The design enters only through ξ.

Proof. First, $f_1 \in \mathcal{F}$ because

$$|f_1(z) - f_1(z')| \le aC_b |z - z'|^\beta / h^\beta \le \Lambda |z - z'|^\beta, \qquad |f_1| \le a \le B.$$

Moreover,

$$\|f_1 - f_0\|_{L^2(Q)}^2 \ge a^2 b_{\min}^2 Q_{1/2}(z_0, h).$$

For the two Gaussian regression models,

$$\mathrm{KL}(\mathbb{P}_1^n \| \mathbb{P}_0^n) = \frac{na^2}{2\sigma^2} \int b\left(\frac{z - z_0}{h}\right)^2 p_\theta(z)\, dz \le \frac{na^2 m(z_0, h)}{2\sigma^2} = \frac{\xi(z_0, h)}{2}.$$

Let $R_i := \mathbb{E}_i \|\hat{f} - f_i\|_{L^2(Q)}^2$. The parallelogram identity gives, for every u,

$$\|u - f_0\|^2 + \|u - f_1\|^2 \ge \frac{1}{2} \|f_1 - f_0\|^2.$$

Integrating against $\mathbb{P}_0^n \wedge \mathbb{P}_1^n$ yields

$$R_0 + R_1 \ge \frac{1}{2} \|f_1 - f_0\|^2 (1 - \mathrm{TV}).$$

Since the maximum is at least the average and $1 - \mathrm{TV} \ge \Phi_+(\xi)$, the result follows. □

Lemma 3 (Hölder sums with interior-disjoint supports)**.** *Let $g_1, \ldots, g_J : [0, 1] \to \mathbb{R}$ be β-Hölder with constant L_0, with supports that are closed intervals with pairwise disjoint interiors (adjacent intervals may share an endpoint), each g_j vanishing on the boundary of its support. Then $\sum_j g_j$ is β-Hölder with constant at most $2^{1-\beta} L_0$.*

Proof. Write $g := \sum_{j=1}^{J} g_j$. At most one summand is nonzero at any point. Fix $z < z'$. If both points lie in one support, the Hölder bound gives $|g(z) - g(z')| \le L_0|z - z'|^\beta$. If neither point lies in the interior of any support, then $g(z) = g(z') = 0$.

It remains to treat the case in which the two points do not lie in one common support and at least one lies in a support interior. If z lies in the interior of a support, let z_1 be the right boundary point of that support; otherwise set $z_1 := z$. If z' lies in the interior of a support, let z_2 be the left boundary point of that support; otherwise set $z_2 := z'$. Because the support interiors are pairwise disjoint and $z < z'$, we have $z_1 \le z_2$. Since every g_j vanishes on the boundary of its support, $g(z_1) = g(z_2) = 0$, and z, z_1 (respectively z', z_2) lie in one support or coincide, so, with $s := |z - z_1|$, $t := |z' - z_2|$, and $s + t \le |z - z'|$,

$$|g(z) - g(z')| \le |g(z) - g(z_1)| + |g(z') - g(z_2)| \le L_0(s^\beta + t^\beta).$$

The subadditivity bound $s^\beta + t^\beta \le 2^{1-\beta}(s + t)^\beta \le 2^{1-\beta}|z - z'|^\beta$ gives the claim. □

Lemma 4 (Assouad)**.** *Let $h \le \min(h_0, 1/4)$, $J := \lfloor 1/(2h) \rfloor$, $z_j := (2j-1)h$, $I_j := [z_j - h, z_j + h]$, $b_j := b((\cdot - z_j)/h)$, and $f_\omega := \sum_j \omega_j a b_j$ for $\omega \in \{0,1\}^J$. If*

$$\mathrm{TV}(\mathbb{P}^n_\omega, \mathbb{P}^n_{\omega^{(j)}}) \le 1 - \tau_j$$

for all j, ω, where $\omega^{(j)}$ flips coordinate j, then

$$\max_\omega \mathbb{E}_\omega \|\hat{f} - f_\omega\|^2_{L^2(Q)} \ge \frac{1}{8} \sum_j \|a b_j\|^2_{L^2(Q)} \tau_j.$$

Proof. The intervals I_j have disjoint interiors and their shared endpoints are Q-null, so

$$\|\hat{f} - f_\omega\|^2 \ge \sum_j \|\hat{f}\mathbf{1}_{I_j} - \omega_j a b_j\|^2.$$

Define the test

$$\hat{\omega}_j := \mathbf{1}\{\|\hat{f}\mathbf{1}_{I_j} - a b_j\| \le \|\hat{f}\mathbf{1}_{I_j}\|\},$$

which depends only on $\hat{f}|_{I_j}$. If $\hat{\omega}_j \ne \omega_j$, the triangle inequality gives, in both directions,

$$\|\hat{f}\mathbf{1}_{I_j} - \omega_j a b_j\| \ge \frac{1}{2}\|a b_j\|.$$

Averaging over the uniform prior on ω therefore yields

$$\max_\omega \mathbb{E}_\omega \|\hat{f} - f_\omega\|^2 \ge \sum_j \frac{\|a b_j\|^2}{4} \cdot \frac{1}{2}\left[\bar{\mathbb{P}}_{j,0}(\hat{\omega}_j = 1) + \bar{\mathbb{P}}_{j,1}(\hat{\omega}_j = 0)\right],$$

where $\bar{\mathbb{P}}_{j,s}$ averages $\mathbb{P}^n_\omega$ over $\omega_j = s$. The bracket is at least

$$1 - \mathrm{TV}(\bar{\mathbb{P}}_{j,0}, \bar{\mathbb{P}}_{j,1}) \ge \tau_j$$

by joint convexity of total variation. □

Applying lemma 3 to $g_j = a b_j$, which has Hölder constant $a C_b / h^\beta$, shows that $f_\omega \in \mathcal{F}$ whenever $a \le B$. Moreover, $\mathbb{P}^n_\omega$ and $\mathbb{P}^n_{\omega^{(j)}}$ differ only on the j-th bump, so their KL divergence is at most $\xi_j/2$, where $\xi_j := \xi(z_j, h)$. Thus $\tau_j := \Phi_+(\xi_j)$ is admissible. The resulting packing bound retains the individual local masses $m(z_j, h)$ rather than reducing them to the scalar profile Ψ.

Proposition 5 (Packing lower bound)**.** *For $h \le \min(h_0, 1/4)$ and $a \le B$,*

$$\inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E}\|\hat{f} - f\|^2_{L^2(Q)} \ge c_5 \Lambda^2 h^{2\beta} \sum_{j=1}^{J} Q_{1/2}(z_j, h) \Phi_+(\xi_j),$$

where $c_5 = b^2_{\min}/(8C'^2_b)$. If $q \ge q_{\min}$ and $\xi_j \le 1$ for all j, then

$$\sum_j Q_{1/2}(z_j, h) \ge q_{\min}/4$$

and the lower bound is at least

$$\frac{b^2_{\min} q_{\min}}{64 C'^2_b} \Lambda^2 h^{2\beta}.$$

Proof. Combine lemma 4 with

$$\|ab_j\|^2 \geq a^2 b_{\min}^2 Q_{1/2}(z_j, h)$$

and $\tau_j = \Phi_+(\xi_j) \geq 1/2$ when $\xi_j \leq 1$. Finally,

$$Jh \geq \left(\frac{1}{2h} - 1\right) h \geq \frac{1}{4}$$

for $h \leq 1/4$. □

Matching scale. We now choose a design-independent scale at which the Hölder and variance terms have the same order. Let $U(h)$ denote the right side of eq. (4). On the controlled family,

$$c'h \leq m(z, h) \leq 2C'h,$$

so lemma 1 and $(1-m)^n \leq e^{-nm}$ give

$$U(h) \leq \Lambda^2 h^{2\beta} + \frac{2\sigma^2}{nc'h} + B^2 e^{-nc'h}.$$

With $\bar{h}$ as in the Setting, define

$$A_0 := \frac{C_b'^2 \sigma^2}{2C'\Lambda^2}, \qquad h_n := \left(\frac{A_0}{n}\right)^{1/(2\beta+1)},$$

so that

$$nh_n\Lambda^2 h_n^{2\beta} = \frac{C_b'^2\sigma^2}{2C'}.$$

Let

$$n_{\text{geom}} := \left\lceil \frac{A_0}{\bar{h}^{2\beta+1}} \right\rceil,$$

so that $n \geq n_{\text{geom}}$ if and only if $h_n \leq \bar{h}$, which ensures $h_n \leq h_0$, $h_n \leq 1/4$, and $a_{h_n} := \Lambda h_n^\beta / C_b' \leq B$. Choose n_{fb} so that

$$B^2 e^{-nc'h_n} \leq \Lambda^2 h_n^{2\beta}$$

for $n \geq n_{\text{fb}}$; such a threshold exists because the left side decays exponentially in $n^{2\beta/(2\beta+1)}$, whereas the right side decays polynomially. Set

$$n_0 := \max\{n_{\text{geom}}, n_{\text{fb}}\}.$$

Proposition 6 (Two-sided bound at the matching scale)**.** *On the controlled family with $q \geq q_{\min}$ and $n \geq n_0$,*

$$\kappa_1 \Lambda^2 h_n^{2\beta} \leq \inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E}\|\hat{f} - f\|_{L^2(Q)}^2 \leq \kappa_2 \Lambda^2 h_n^{2\beta},$$

where

$$\kappa_1 = \frac{b_{\min}^2 q_{\min}}{64 C_b'^2}, \qquad \kappa_2 = 2 + \frac{4C'}{c' C_b'^2},$$

and neither constant depends on the design.

Proof. For $n \geq n_0$,

$$\xi_j \leq \frac{2C' n h_n \Lambda^2 h_n^{2\beta}}{C_b'^2 \sigma^2} = 1$$

for every j, so proposition 5 at h_n gives the lower bound. For the upper bound, the definition of h_n gives

$$\frac{\sigma^2}{nh_n} = \frac{2C'\Lambda^2 h_n^{2\beta}}{C_b'^2},$$

and therefore

$$U(h_n) \leq \Lambda^2 h_n^{2\beta}\left(1 + \frac{4C'}{c' C_b'^2}\right) + B^2 e^{-nc'h_n} \leq \kappa_2 \Lambda^2 h_n^{2\beta}.$$

□

Proof of theorem 1(b). Write R^*_θ for the minimax risk. The upper bound follows from theorem 1(a) by taking the infimum over $0 < h \le \bar{h}$. For the lower bound, proposition 6 gives

$$R^*_\theta \ge \kappa_1 \Lambda^2 h_n^{2\beta}.$$

Since $h_n \le \bar{h}$ is feasible,

$$\mathcal{C}_\theta(n) \le U(h_n) \le \kappa_2 \Lambda^2 h_n^{2\beta},$$

so $\Lambda^2 h_n^{2\beta} \ge \mathcal{C}_\theta(n)/\kappa_2$ and hence

$$R^*_\theta \ge \frac{\kappa_1}{\kappa_2} \mathcal{C}_\theta(n).$$

□

The constants κ_1, κ_2 depend on the controlled-family bounds c, C, giving a per-design equivalence up to family-uniform constants. For any two designs in the family,

$$\frac{\kappa_1}{\kappa_2} \frac{\mathcal{C}_{\theta_1}(n)}{\mathcal{C}_{\theta_2}(n)} \le \frac{R^*_{\theta_1}}{R^*_{\theta_2}} \le \frac{\kappa_2}{\kappa_1} \frac{\mathcal{C}_{\theta_1}(n)}{\mathcal{C}_{\theta_2}(n)}. \tag{8}$$

Because h_n, κ_1, and κ_2 are shared across the controlled family, proposition 6 together with $R^*_\theta \le \mathcal{C}_\theta(n) \le U(h_n)$ gives

$$\kappa_1 \Lambda^2 h_n^{2\beta} \le R^*_\theta \le \mathcal{C}_\theta(n) \le \kappa_2 \Lambda^2 h_n^{2\beta}$$

for every design in the family. Hence $\mathcal{C}_{\theta_1}(n)/\mathcal{C}_{\theta_2}(n) \in [\kappa_1/\kappa_2, \kappa_2/\kappa_1]$ for every pair in the family. Equation (8) therefore gives a family-uniform constant-factor comparison of minimax risks; fine-grained design comparisons use the evaluable certificate and empirical risk tracking. Both sides retain the local mass field: the upper bound through $\int g_n(m_\theta)\, dQ$ and the lower bound through $\sum_j Q_{1/2} \Phi_+(\xi_j)$. The scalar profile enters only through the scalar upper consequence of lemma 1.

B.3 Rigidity and Coarea

This subsection provides the geometric results used in sections 4 and 6. We first prove theorem 2 for the design-normalized two-link family, then derive the coarea representation eq. (10) and verify its assumptions for the model families used in the paper.

B.3.1 The two-link family

Setting. Let $L_1 = 1$ and $L_2 = a \in (0, 1)$. The elbow angle $q \sim \mathrm{Unif}[0, 2\pi)$ is folded by $q \mapsto \min(q, 2\pi - q)$ onto $[0, \pi]$. The folded law is $\mathrm{Unif}[0, \pi]$: for Borel $A \subset [0, \pi]$, the preimage $A \cup (2\pi - A)$ consists of two sets of Lebesgue measure $|A|$ meeting at most at endpoints, so the pushforward has mass $2|A|/(2\pi) = |A|/\pi$. The radial map is

$$r(q) = |1 + ae^{iq}| = \sqrt{1 + a^2 + 2a\cos q},$$

which is strictly decreasing on $[0, \pi]$. The normalized coordinate is

$$x(q) = \frac{r(q) - (1 - a)}{2a} \in [0, 1],$$

with

$$J_x = \left|\frac{dx}{dq}\right| = \frac{J_r}{2a}, \qquad J_r = \left|\frac{dr}{dq}\right| = \frac{a \sin q}{r}.$$

The normalized deployment law is $Q_x = \mathrm{Unif}[0, 1]$.

Lemma 5 (Argument change of variables)**.** *Let* $\delta(q) = \arg(1 + ae^{iq}) \in [0, \pi/2)$ *and* $\phi(q) := q - \delta(q)$. *For* $q \in [0, \pi]$*:* (i) $r'(q) = -a \sin\phi(q)$*, so* $J_r = a \sin\phi$*;* (ii) $\sin\delta = a \sin\phi$ *and* $\delta = \arcsin(a \sin\phi)$*;* (iii) $q(\phi) = \phi + \arcsin(a \sin\phi)$ *is a strictly increasing bijection* $[0, \pi] \to [0, \pi]$ *with*

$$\frac{dq}{d\phi} = 1 + \frac{a\cos\phi}{\sqrt{1 - a^2 \sin^2\phi}} \ge 1 - a > 0.$$

Proof. Since $\mathrm{Re}(1 + ae^{iq}) \geq 1 - a > 0$, δ is well defined and smooth. Moreover, $0 \leq \delta \leq q$, because the argument of a sum of two vectors with arguments 0 and q lies between them; hence $\phi \in [0, q]$, $\phi(0) = 0$, and $\phi(\pi) = \pi$. Differentiating

$$r(q)e^{i\delta(q)} = 1 + ae^{iq}$$

and multiplying by $e^{-i\delta}$ gives

$$r' + ir\delta' = iae^{i\phi},$$

whose real part gives (i). Taking the imaginary part of $r = e^{-i\delta} + ae^{i\phi}$ yields $\sin\delta = a\sin\phi$. Since $\delta \in [0, \pi/2)$ and $a\sin\phi \in [0, 1)$, the principal branch gives (ii).

For (iii), differentiation gives

$$\frac{dq}{d\phi} = 1 + \frac{a\cos\phi}{\sqrt{1 - a^2\sin^2\phi}}.$$

Since

$$1 - a^2\sin^2\phi - \cos^2\phi = (1 - a^2)\sin^2\phi \geq 0,$$

we have $\sqrt{1 - a^2\sin^2\phi} \geq |\cos\phi|$, and therefore

$$\left|\frac{a\cos\phi}{\sqrt{1 - a^2\sin^2\phi}}\right| \leq a$$

whatever the sign of $\cos\phi$. Hence $dq/d\phi \geq 1 - a > 0$; together with continuity, strict monotonicity, and the endpoint values, this proves the bijection. □

Proof of theorem 2(a). Let ψ be bounded and measurable. By lemma 5(i),

$$J_x(q) = \frac{1}{2}\sin\phi(q),$$

so

$$\mathbb{E}[\psi(J_x)] = \frac{1}{\pi}\int_0^\pi \psi\left(\frac{1}{2}\sin\phi(q)\right)\, dq.$$

Substituting $q = q(\phi)$ using lemma 5(iii) gives

$$\mathbb{E}[\psi(J_x)] = \frac{1}{\pi}\int_0^\pi \psi\left(\frac{1}{2}\sin\phi\right)\left[1 + \frac{a\cos\phi}{\sqrt{1 - a^2\sin^2\phi}}\right] d\phi.$$

The second term is antisymmetric under $\phi \mapsto \pi - \phi$, while $\psi(\frac{1}{2}\sin\phi)$ is symmetric; the integrand is bounded by $\|\psi\|_\infty(1 + a)$, so the antisymmetric term integrates to zero. Thus

$$\mathbb{E}[\psi(J_x)] = \frac{1}{\pi}\int_0^\pi \psi\left(\frac{1}{2}\sin\phi\right) d\phi = \mathbb{E}\left[\psi\left(\frac{1}{2}\sin U\right)\right],$$

where $U \sim \mathrm{Unif}[0, \pi]$. Taking ψ to be indicators gives

$$J_x \overset{d}{=} \frac{1}{2}\sin U.$$

For $s > -1$, monotone approximation and the Wallis–Beta integral give

$$\mathbb{E}[(2J_x)^s] = \frac{1}{\pi}\int_0^\pi \sin^s\phi\, d\phi = \frac{\Gamma((s+1)/2)}{\sqrt{\pi}\,\Gamma((s+2)/2)};$$

in particular $\mathbb{E}[2J_x] = 2/\pi$ and $\mathbb{E}[(2J_x)^2] = 1/2$. For $s \leq -1$, the moment $\mathbb{E}[(2J_x)^s]$ and the integral $\pi^{-1}\int_0^\pi \sin^s\phi\, d\phi$ both diverge to $+\infty$ because $\sin^s\phi \sim \phi^s$ near 0; the Gamma-function expression is used only for $s > -1$.

Density-value rigidity. The density of x is

$$p_a(x) = \frac{1}{\pi}\left|\frac{dq}{dx}\right| = \frac{1}{\pi J_x(q(x))} = \frac{2r(x)}{\pi\sqrt{1-c(x)^2}},$$

where

$$r(x) = (1-a) + 2ax, \qquad c(x) = \frac{r(x)^2 - 1 - a^2}{2a}.$$

It has integrable inverse-square-root singularities at the endpoints. For nonnegative measurable H, the identities

$$p_a(x(q)) = \frac{1}{\pi J_x(q)}, \qquad |dx| = J_x(q)\,dq$$

on $(0,\pi)$ (the endpoints are null) give

$$\int_0^1 H(p_a(x))\,dx = \int_0^\pi H\big((\pi J_x(q))^{-1}\big)\,J_x(q)\,dq = \pi\,\mathbb{E}\big[J_x H((\pi J_x)^{-1})\big], \tag{9}$$

in the extended reals. The right-hand side depends only on the law of J_x, so the law of $p_a(X)$ for $X \sim \mathrm{Unif}[0,1]$ is independent of a.

In particular, for $t < 2$,

$$\int_0^1 p_a(x)^t\,dx = \pi^{1-t}\mathbb{E}[J_x^{1-t}]$$

is finite and design-invariant, whereas for $t \geq 2$ it is $+\infty$ for every a. At $t = -1$,

$$\int_0^1 p_a(x)^{-1}\,dx = \pi^2\mathbb{E}[J_x^2] = \frac{\pi^2}{8},$$

so

$$\chi^2(Q_x\|P_a) = \int_0^1 p_a(x)^{-1}\,dx - 1 = \frac{\pi^2}{8} - 1.$$

Thus every well-defined power- or Rényi-type functional stated in theorem 2 is design-invariant; in particular, $\int_0^1 p_a(x)^2\,dx = +\infty$. □

Proof of theorem 2(b). At $x = \frac{1}{2}$, $r = 1$ and $c = -a/2$, hence

$$p_a\left(\frac{1}{2}\right) = \frac{2}{\pi\sqrt{1-a^2/4}},$$

which is strictly increasing in a. Therefore $p_{a_1}(\frac{1}{2}) \neq p_{a_2}(\frac{1}{2})$ when $a_1 \neq a_2$. Since p_a is continuous near $\frac{1}{2}$ and

$$m_a\left(\frac{1}{2}, h\right) = \int_{1/2-h}^{1/2+h} p_a(x)\,dx \qquad (0 < h < 1/2),$$

we have

$$\lim_{h\downarrow 0}\frac{m_a(\frac{1}{2},h)}{2h} = p_a\left(\frac{1}{2}\right).$$

Hence there exists $\eta > 0$ such that

$$m_{a_1}\left(\frac{1}{2}, h\right) \neq m_{a_2}\left(\frac{1}{2}, h\right)$$

for all $0 < h < \eta$. □

The raw Jacobian satisfies $J_r = a(2J_x)$, so its scale identifies a; the rigidity result therefore concerns the normalized-coordinate Jacobian, which is the common coordinate of the finite-scale comparison. Because the coordinate x is design-dependent, this is a design-normalized structural comparison rather than a comparison under one shared deployment law.

For the finite-scale numerical witness used in the main text, define

$$\Psi_a(h) = \int_0^1 \frac{dx}{m_a(x,h)}$$

with interval masses truncated to $[0,1]$. Deterministic quadrature gives

$$\Psi_{0.05}(0.1) = 6.193498, \qquad \Psi_{0.99}(0.1) = 6.451552,$$

a difference of about 4.2%, and at $h = 0.05$ gives 12.3298 and 12.5995, respectively, with estimated numerical error below 10^{-6}; an independent recomputation agrees to ten digits. Thus, for every pair $a_1 \neq a_2$, the normalized-Jacobian and density-value laws coincide while the mass fields differ for all sufficiently small h, and the displayed designs also have separated profiles at $h = 0.05$ and $h = 0.1$: no functional of the normalized-Jacobian law or of the density-value distribution determines finite-scale coverage.

B.3.2 COAREA REPRESENTATION OF THE SOURCE DENSITY

Assumptions. Let $\mathcal{S}$ be a compact $C^{r_{\mathrm{reg}}}$ Riemannian manifold of dimension $k \geq d$ (or a compact region of $\mathbb{R}^k$ with the Euclidean metric), with

$$r_{\mathrm{reg}} \geq \max(k - d + 1, 1).$$

If $\mathcal{S}$ has boundary, assume T_θ extends $C^{r_{\mathrm{reg}}}$ to a neighborhood. Write

$$d\mu = \varrho \, d\mathrm{vol}, \qquad T_\theta : \mathcal{S} \to \mathcal{Z} \subset \mathbb{R}^d.$$

The normal Jacobian is

$$J_{T_\theta}(q) = \sqrt{\det\big(DT_\theta(q) DT_\theta(q)^*\big)},$$

where the adjoint is taken with respect to the source metric and the Euclidean target metric. In coordinates with source metric matrix $\mathbf{G}_{\mathcal{S}}$,

$$J_{T_\theta} = \sqrt{\det(DT_\theta \, \mathbf{G}_{\mathcal{S}}^{-1} DT_\theta^\top)},$$

and in flat coordinates $J_{T_\theta} = \sqrt{\det(DT_\theta DT_\theta^\top)}$. The model families considered below use flat source coordinates. Define

$$\mathrm{crit} := \{q : \operatorname{rank} DT_\theta(q) < d\} = \{J_{T_\theta} = 0\}.$$

- (Co1) $\mathrm{vol}(\mathrm{crit}) = 0$, hence $\mu(\mathrm{crit}) = 0$. This is not automatic from analyticity (a constant map is analytic and everywhere rank-deficient) and is verified for the model families below. For a connected, globally flat, real-analytic source domain, one full-rank configuration is sufficient: a nonzero maximal minor is a real-analytic function whose zero set is Lebesgue-null and contains crit.
- (Co2) $Q \ll \mathrm{Leb}^d$. A bounded deployment density is imposed only in results that explicitly require it.
- (Co3) $T_\theta \in C^{r_{\mathrm{reg}}}$ with $r_{\mathrm{reg}} \geq \max(k - d + 1, 1)$. By the Morse–Sard theorem, $T_\theta(\mathrm{crit})$ is Lebesgue-null in $\mathbb{R}^d$.

Conditions (Co1) and (Co3) control different null sets: (Co1) removes the contribution of crit to the source integral, whereas (Co3) makes its image Lebesgue-null. The theorem below first establishes $P_\theta \ll \mathrm{Leb}^d$; together with (Co3) this gives $P_\theta(T_\theta(\mathrm{crit})) = 0$, while (Co2) gives $Q(T_\theta(\mathrm{crit})) = 0$.

For $z \notin T_\theta(\mathrm{crit})$, and hence for almost every z,

$$p_\theta(z) = \int_{T_\theta^{-1}(z) \setminus \mathrm{crit}} \frac{\varrho(q)}{J_{T_\theta}(q)} \, d\mathcal{H}^{k-d}(q), \qquad m_\theta(z,h) = \int_{\mathbb{B}_h(z)} p_\theta(u) \, du, \tag{10}$$

where $\mathcal{H}^{k-d}$ is the $(k-d)$-dimensional Hausdorff measure on the fiber $T_\theta^{-1}(z)$, with counting measure when $k = d$.

Theorem 3 (Source density). *Under (Co1)–(Co3), $P_\theta \ll \mathrm{Leb}^d$ with density given by eq.* (10) *for $z \notin T_\theta(\mathrm{crit})$ (a Lebesgue-full, hence P_θ-full and Q-full, set); for $k = d$ in flat coordinates the fiber integral is the branch sum $p_\theta(z) = \sum_{q \in T_\theta^{-1}(z)} \varrho(q)/|\det DT_\theta(q)|$; $m_\theta(z,h) = \int_{\mathbb{B}_h(z)} p_\theta$, and $T_\theta(\mathrm{crit})$ carries neither source-distribution nor deployment mass.*

Proof. For measurable $A \subset \mathcal{Z}$,

$$P_\theta(A) = \int_{T_\theta^{-1}(A)\setminus \mathrm{crit}} \varrho \, d\mathrm{vol} + \int_{T_\theta^{-1}(A)\cap \mathrm{crit}} \varrho \, d\mathrm{vol}.$$

The second term is at most $\mu(\mathrm{crit}) = 0$ by (Co1). On $\mathcal{S} \setminus \mathrm{crit}$, $J_{T_\theta} > 0$. With

$$g := \frac{\varrho}{J_{T_\theta}} \mathbf{1}_{T_\theta^{-1}(A)\setminus \mathrm{crit}},$$

the coarea formula gives

$$\int g J_{T_\theta} \, d\mathrm{vol} = \int_{\mathbb{R}^d} \left(\int_{T_\theta^{-1}(z)} g \, d\mathcal{H}^{k-d} \right) dz = \int_A p_\theta(z) \, dz.$$

Hence $P_\theta \ll \mathrm{Leb}^d$ with density p_θ. Taking $A = T_\theta(\mathrm{crit})$ and using (Co3) gives $P_\theta(T_\theta(\mathrm{crit})) = 0$. For $k = d$, $\mathcal{H}^0$ is counting measure and $J_{T_\theta} = |\det DT_\theta|$. □

Lemma 6 (Local fold contribution, $k = d = 1$). *Suppose ϱ is continuous and positive near q_0, $T'_\theta(q_0) = 0$, and $T''_\theta(q_0) \neq 0$. Let $z_0 := T_\theta(q_0)$, let U be a sufficiently small source interval around q_0 (one-sided when q_0 is an endpoint of the source interval), and define the local branch contribution*

$$p_{\theta,U}(z) := \sum_{q \in U \cap T_\theta^{-1}(z)} \frac{\varrho(q)}{|T'_\theta(q)|}.$$

Then, as $z \to z_0$ from the attainable side $(z - z_0)T''_\theta(q_0) > 0$,

$$p_{\theta,U}(z) \sim c\,|z - z_0|^{-1/2}, \qquad c = \frac{\varrho(q_0)}{\sqrt{2|T''_\theta(q_0)|}}$$

when q_0 is an endpoint of the source interval, and with twice this constant when q_0 is interior.

Proof. Taylor expansion gives

$$T_\theta(q) - z_0 = \frac{1}{2} T''_\theta(q_0)(q - q_0)^2 + o((q - q_0)^2)$$

and

$$T'_\theta(q) = T''_\theta(q_0)(q - q_0) + o(|q - q_0|).$$

Hence, on each local inverse branch within U on the attainable side of z_0,

$$|q - q_0| = \sqrt{\frac{2|z - z_0|}{|T''_\theta(q_0)|}} \, (1 + o(1)),$$

and therefore

$$|T'_\theta(q)| = \sqrt{2|T''_\theta(q_0)|\,|z - z_0|} \, (1 + o(1)).$$

Since $\varrho(q) = \varrho(q_0) + o(1)$, each local branch contributes

$$\frac{\varrho(q_0)}{\sqrt{2|T''_\theta(q_0)|}} |z - z_0|^{-1/2} (1 + o(1)).$$

There is one such local branch for an endpoint fold and two for an interior nondegenerate fold, which gives the stated constants. □

For nearby regular values z, the one-dimensional branch formula of theorem 3 decomposes the full density $p_\theta(z)$ into the local contribution $p_{\theta,U}(z)$ and contributions from preimages outside U. A regular inverse branch outside U contributes a locally bounded density term whenever the source density is locally bounded along that branch. Other critical preimages can contribute additional singular terms; in particular, another nondegenerate fold mapping to the same critical value contributes another $|z - z_0|^{-1/2}$ term whose coefficient must be added. Thus the lemma identifies the contribution of the fold at q_0, not necessarily the entire leading asymptotic of p_θ without further isolation assumptions.

The fold calculation illustrates why the mass field is more stable than pointwise density: a nondegenerate fold contributes an inverse-square-root density singularity, whereas $m_\theta(z, h) \le 1$ always remains finite. Conversely, vanishing pointwise density is compatible with a positive finite-scale lower bound $m_\theta(z, h) \ge \eta_h$ on a common-support family. This is why the principal object and the certificate are stated through m_θ, and why the density-regular matching result proposition 6 is a separate regime from the field-level certificate.

Verification of (Co1) for the model families. Planar chains. For

$$F(q) = \sum_{j=1}^{m} L_j e^{iU_j}, \qquad U_j = q_1 + \cdots + q_j, \qquad L_j > 0, \quad m \ge 2,$$

the configuration $q^* = (0, \pi/2, 0, \dots)$ gives

$$\frac{\partial F}{\partial q_1} = (-L_{\text{tail}}, L_1), \qquad \frac{\partial F}{\partial q_2} = (-L_{\text{tail}}, 0),$$

where $L_{\text{tail}} := \sum_{j \ge 2} L_j > 0$. The corresponding 2×2 minor equals $L_1 L_{\text{tail}} > 0$, providing a full-rank configuration.

For the two-link radial map, $k = d = 1$ and $\text{crit} = \{0, \pi\}$. Moreover,

$$r''(0) = -\frac{a}{1+a}, \qquad r''(\pi) = \frac{a}{1-a},$$

so both critical points are endpoint folds, consistent with the inverse-square-root singularities of p_a.

For the three-link planar map $F_L : \mathbb{T}^3 \to \mathbb{R}^2$ ($k = 3$, $d = 2$, real-analytic), the same full-rank configuration $(0, \pi/2, 0)$ gives the nonzero minor

$$L_1(L_2 + L_3) > 0.$$

The radial law is then obtained from the rotation-invariant planar density by the polar pushforward

$$p_{R,L}(r) = 2\pi r\, \bar{p}_{F,L}(r),$$

using the rotation invariance induced by joint-uniform angles, rather than by applying the theorem to the radial map $\|F_L\|$, which is singular at the origin.

Seven-joint sagittal projection. For the seven-joint family of section D, the source is the connected flat joint-limit box

$$\prod_{j=1}^{7} [q_j^-, q_j^+],$$

and $T_\theta(q) = (x_{\text{tool}}, z_{\text{tool}})$ is the sagittal projection of the tool point of the fixed modified Denavit–Hartenberg (DH) chain. Each matrix entry is a polynomial in sines and cosines of the joint coordinates, so T_θ is real-analytic and (Co3) holds.

For (Co1), take

$$q_0 = (0, 0, 0, -\pi/2, 0, \pi/2, 0),$$

which lies strictly inside the joint box. Interval arithmetic gives the uniform lower bound

$$\left|\det DT_\theta(q_0)_{(q_2, q_4)}\right| > 0.1379$$

for each of the 8 fixed variants, where the subscript (q_2, q_4) selects the two columns of the differential corresponding to those joint coordinates. For the idealized twists exactly equal to $\pm\pi/2$, this minor reduces to

$$d_3(d_5 + 0.088) + a_4(d_{\text{tool}} + 0.0245),$$

where $d_3, d_5, a_4, d_{\text{tool}}$ are the modified-DH length parameters listed in table 7; the expression is positive for every variant. The fixed twists differ from $\pm\pi/2$ by approximately 3×10^{-8}, changing the minor by less than 10^{-15}. Thus (Co1) holds for the 8 fixed variants. Numerical agreement with an independent implementation is summarized in section C.6.

Corollary 4 (Link-order invisibility and the source of common-support variation)**.** *For the three-link map with i.i.d. uniform joints, the map $(q_1, q_2, q_3) \mapsto (U_1, U_2, U_3)$ is a unimodular integer lower-triangular automorphism of $\mathbb{T}^3$ and preserves Haar measure, so (U_j) are i.i.d. uniform and the end-point law depends on the link lengths only through the multiset $\{L_j\}$ (a Pearson random walk, i.e., a sum of independent random link vectors); fibers of permuted designs correspond under the automorphism, not pointwise in the original coordinates, and the radial laws coincide.*

On a common-support family, variation of the mass field across designs is represented by the design dependence of the fiber integral in eq. (10)*. The fiber representation explains how design dependence enters the mass field; the size of the variation is empirical, with the observed 13–15% spread reported in section C.1.*

B.4 Acquisition and Allocation

This subsection provides the acquisition and allocation results used in section 5. We first characterize exact rejection sampling and its finite-scale cost bounds, including the structure of the rejection envelope for the three-link family. We then prove the convex source-allocation results, the exact mixture-gap identity, and the convexity properties of the finite-n mass-field objective.

B.4.1 Thinning, envelopes, and the cost sandwich

General setting. Assume $Q \ll P_\theta$. Let

$$w := \frac{dQ}{dP_\theta}$$

be a measurable version and define

$$M_\theta := \operatorname*{ess\,sup}_{P_\theta} w \in [1, +\infty].$$

Since $\int w \, dP_\theta = 1$, necessarily $M_\theta \geq 1$. Part (b) of proposition 1 is interpreted in the extended reals and does not require $M_\theta < \infty$.

Sampling model for exact rejection. For part (a) of proposition 1 and the thinning corollaries, assume additionally $M_\theta < \infty$ and take a version satisfying $0 \leq w \leq M_\theta$ P_θ-a.e. Proposals $(Z_i)_{i \leq N_{\text{prop}}}$ are i.i.d. from P_θ, and auxiliary variables (U_i) are i.i.d. $\text{Unif}[0,1]$ and independent of the proposals. Define

$$I_i^{\text{acc}} := \mathbf{1}\left\{U_i \leq \frac{w(Z_i)}{M_\theta}\right\}, \qquad \mathcal{I}_{\text{acc}} := \{i : I_i^{\text{acc}} = 1\}, \qquad N_{\text{acc}} := |\mathcal{I}_{\text{acc}}|.$$

Acceptance depends only on (Z_i, U_i). Accepted proposals are then labeled according to

$$Y_i \mid Z_i \sim f(Z_i) + \varepsilon,$$

conditionally independently given all (Z, U); equivalently, $I_i^{\text{acc}} \perp Y_i \mid Z_i$.

An estimator family $(\hat{f}_N)_{N \geq 0}$ receives only N_{acc} and the accepted labeled samples in proposal order, not rejected proposals, acceptance times, or other process metadata. The estimator $\hat{f}_0$ is the no-sample output; for zero-fallback NW, $\hat{f}_0 \equiv 0$ and its risk is $\int f^2 \, dQ$. Write $R_{\text{TM}}(\hat{f}; N)$ for the task-matched risk with N i.i.d. Q-samples and $R_{\text{RJ}}(\hat{f}; N_{\text{prop}})$ for the rejection-sampling risk.

Lemma 7 (Binomial thinning)**.** *If $N \sim \text{Binomial}(n, p)$ and $K \mid N \sim \text{Binomial}(N, q)$, then $K \sim \text{Binomial}(n, pq)$. Consequently $\mathbb{E}[(1-q)^N] = (1-pq)^n$ and $\mathbb{E}[g_N(q)] = g_n(pq)$.*

Proof. The probability-generating function satisfies

$$\mathbb{E}[s^K] = \mathbb{E}[(1 - q + qs)^N] = (1 - pq + pqs)^n,$$

which is the generating function of $\mathrm{Binomial}(n, pq)$. For the second identity,

$$\mathbb{E}[g_N(q)] = \mathbb{E}[\mathbb{E}[1/K; K > 0 \mid N]] = \mathbb{E}[1/K; K > 0] = g_n(pq).$$

□

Proof of proposition 1(a). **Acceptance rate.** The pairs $(Z_i, I_i^{\mathrm{acc}})$ are i.i.d. and

$$\mathbb{P}(I_i^{\mathrm{acc}} = 1) = \mathbb{E}\left[\frac{w(Z_i)}{M_\theta}\right] = \frac{1}{M_\theta}.$$

Hence

$$N_{\mathrm{acc}} \sim \mathrm{Binomial}(N_{\mathrm{prop}}, 1/M_\theta).$$

If $M_\theta = 1$, then $w = 1$ a.e., $Q = P_\theta$, $N_{\mathrm{acc}} = N_{\mathrm{prop}}$ a.s., and the claim is trivial. Assume henceforth $M_\theta > 1$, so that every index set has positive probability.

Accepted law. For measurable $B \subset \mathcal{Z}$,

$$\begin{aligned}\mathbb{P}(Z_i \in B \mid I_i^{\mathrm{acc}} = 1) &= \frac{\mathbb{E}[\mathbf{1}_B(Z_i) w(Z_i)/M_\theta]}{1/M_\theta} \\ &= \int_B w \, dP_\theta = Q(B).\end{aligned}$$

For an index set S with $|S| = k$ and measurable $B_i \subset \mathcal{Z}$, $i \in S$,

$$\mathbb{P}(\mathcal{I}_{\mathrm{acc}} = S,\ Z_i \in B_i\ \forall i \in S) = \prod_{i \in S} \frac{Q(B_i)}{M_\theta} \left(1 - \frac{1}{M_\theta}\right)^{N_{\mathrm{prop}} - k}.$$

Dividing by $\mathbb{P}(\mathcal{I}_{\mathrm{acc}} = S) > 0$ shows that, conditional on $\mathcal{I}_{\mathrm{acc}} = S$, the accepted covariates are i.i.d. from Q. Averaging over all sets S of size k gives the same law for the ordered accepted sequence conditional on $N_{\mathrm{acc}} = k$.

Because acceptance depends only on (Z, U), conditional on the accepted covariates the labels have the task-matched conditional law and remain conditionally independent: there is no label-selection bias. Thus, under the accepted-sample-only convention, the estimator input on $\{N_{\mathrm{acc}} = k\}$ has exactly the law of the k-sample task-matched experiment.

Mixture over accepted sample size. The tower property gives

$$R_{\mathrm{RJ}}(\hat{f}; N_{\mathrm{prop}}) = \sum_{k=0}^{N_{\mathrm{prop}}} \mathbb{P}(\mathrm{Binomial}(N_{\mathrm{prop}}, 1/M_\theta) = k)\, R_{\mathrm{TM}}(\hat{f}; k),$$

including the $k = 0$ term $\ell_Q(\hat{f}_0, f)$.

Larger valid envelopes. If the exact w is used with any constant $M_\theta \le \bar{M} < \infty$, then

$$\mathbb{P}(Z \in B \mid \mathrm{accept}) = \frac{\mathbb{E}[\mathbf{1}_B w/\bar{M}]}{\mathbb{E}[w/\bar{M}]} = Q(B),$$

with acceptance rate $1/\bar{M}$. Hence the same argument applies with $\bar{M}$ in place of M_θ. The exact M_θ is the smallest valid constant envelope and therefore gives the largest acceptance rate. □

Conditional on N_{acc}, the rejection experiment is the task-matched statistical experiment of proposition 3; proposition 1(a) is this equivalence mixed over the random accepted sample size. The exact distributional identity requires the exact likelihood ratio w and an implementation constant $M_{\mathrm{impl}} \ge M_\theta$.

The three-link numerical experiments instead use the approximation

$$\hat{w}_\theta := \frac{q_R}{\hat{p}_{R,\theta}},$$

the exact radial deployment density divided by the numerically integrated radial source density, together with candidate envelopes $\widehat{M}_\theta$. The accepted law is then only approximately Q: finite-grid

envelope checks, accepted-CDF comparisons, and Monte Carlo checks provide numerical evidence but do not certify the global envelope condition for the exact ratio. We therefore distinguish throughout between the theoretical quantities (w, M_θ) and the numerical quantities $(\hat{w}_\theta, \widehat{M}_\theta, M_{\mathrm{impl}})$; the corresponding numerical checks are described in section C.2.

Corollary 5 (Thinned-field identity for constant-target NW)**.** *For $s : \mathcal{Z} \to [0,1]$ measurable and $n \geq 0$ define*

$$\mathfrak{R}_n^{\mathrm{const}}[s] := B^2 \int (1-s)^n \, dQ + \sigma^2 \int g_n(s) \, dQ$$

($g_0 \equiv 0$, so $\mathfrak{R}_0^{\mathrm{const}} = B^2$). For uniform-ball NW with zero fallback, constant target $f \equiv B$, and homoscedastic noise,

$$R_{\mathrm{RJ}}(N_{\mathrm{prop}}) = \mathbb{E}_{N_{\mathrm{acc}}} \mathfrak{R}_{N_{\mathrm{acc}}}^{\mathrm{const}}[m_Q] = \mathfrak{R}_{N_{\mathrm{prop}}}^{\mathrm{const}}[m_Q/M_\theta].$$

Proof. The first equality is proposition 1(a) together with the exact constant-target NW decomposition, whose bias term is zero.

For each z, the events "accepted and in $\mathbb{B}_h(z)$" are independent across proposals with probability

$$\int_{\mathbb{B}_h(z)} \frac{w}{M_\theta} \, dP_\theta = \frac{m_Q(z,h)}{M_\theta}.$$

Thus the accepted in-ball count is

$$\mathrm{Binomial}\left(N_{\mathrm{prop}}, \frac{m_Q(z,h)}{M_\theta}\right).$$

Applying lemma 7 with $N = N_{\mathrm{acc}}$ gives

$$\mathbb{E}[(1-m_Q)^{N_{\mathrm{acc}}}] = (1 - m_Q/M_\theta)^{N_{\mathrm{prop}}}, \qquad \mathbb{E}[g_{N_{\mathrm{acc}}}(m_Q)] = g_{N_{\mathrm{prop}}}(m_Q/M_\theta).$$

Integrating with respect to Q gives the result by Tonelli's theorem. When $m_Q(z,h) = 0$, the count is zero almost surely and the formula remains valid without defining an in-ball conditional law. □

The quantity m_Q/M_θ is a per-proposal success-probability field. In general it is not the ball-mass field of a probability law, so the right side of corollary 5 should not be interpreted as the ordinary NW risk under a new sampling distribution.

Corollary 6 (Thinned-field identity for a general target)**.** *Let f be bounded, the noise homoscedastic with variance σ^2, and $M_\theta \leq \bar{M} < \infty$ a valid constant envelope used with the exact w. For z with $m_Q(z,h) > 0$, let $\mu_f(z,h)$ and $v_f(z,h)$ denote the mean and variance of $f(U)$ for*

$$U \sim Q \quad \textit{conditional on} \quad U \in \mathbb{B}_h(z),$$

and set

$$s(z) := \frac{m_Q(z,h)}{\bar{M}}.$$

Then, for uniform-ball NW with zero fallback,

$$\begin{aligned} R_{\mathrm{RJ}}(N_{\mathrm{prop}}) = \int \Big[& f(z)^2 (1-s(z))^{N_{\mathrm{prop}}} \\ & + \big(\mu_f(z,h) - f(z)\big)^2 \big(1 - (1-s(z))^{N_{\mathrm{prop}}}\big) \\ & + \big(v_f(z,h) + \sigma^2\big) g_{N_{\mathrm{prop}}}(s(z)) \Big] \, dQ(z), \end{aligned}$$

where at points with $m_Q(z,h) = 0$ the last two terms vanish and the ball moments may be defined arbitrarily.

Proof. Fix z. Each proposal is accepted and lies in $\mathbb{B}_h(z)$ independently with probability $s(z)$, so the accepted in-ball count satisfies

$$K \sim \mathrm{Binomial}(N_{\mathrm{prop}}, s(z)).$$

By the proof of proposition 1(a), conditional on K the accepted in-ball locations are i.i.d. from Q restricted to $\mathbb{B}_h(z)$, with labels $f(U_j) + \varepsilon_j$.

On $\{K = k > 0\}$, the NW output is $\bar{f}_k + \bar{\varepsilon}_k$, where $\bar{f}_k$ is the mean of k i.i.d. copies of $f(U)$. Hence

$$\mathbb{E}[(\hat{f} - f(z))^2 \mid K = k] = (\mu_f - f(z))^2 + \frac{v_f + \sigma^2}{k}.$$

On $\{K = 0\}$ the squared error is $f(z)^2$. Averaging over K and integrating with respect to Q by Tonelli's theorem gives the displayed identity. For $f \equiv B$, $\mu_f = B$ and $v_f = 0$, recovering corollary 5. □

This identity is used for the fixed-proposal bump-target arm in section C.2. That experiment also reports a descriptive mean-count comparison: the same functional evaluated at the fixed count

$$\bar{n}_\theta = \mathrm{round}(N_{\mathrm{prop}}/\widehat{M}_\theta)$$

with the unthinned field m_Q (for the constant target, $\mathfrak{R}^{\mathrm{const}}_{\bar{n}_\theta}[m_Q]$). Because this comparison replaces a random sample size by a rounded deterministic count inside a nonlinear functional, its difference from the random-count risk is not a pure Jensen gap; it is reported only descriptively.

Proof of proposition 1(b). If $M_\theta < \infty$, then for every task ball

$$m_Q(z,h) = \int_{\mathbb{B}_h(z)} w \, dP_\theta \le M_\theta \, m_\theta(z,h),$$

so

$$\frac{m_Q}{m_\theta} \le M_\theta \qquad Q\text{-a.e.},$$

and hence $D_\theta(h) \le M_\theta$. If $M_\theta = +\infty$, this inequality is immediate. Therefore

$$D_\theta(h) \le M_\theta$$

in the extended reals.

By definition of $D_\theta(h)$,

$$\frac{1}{m_\theta} \le \frac{D_\theta(h)}{m_Q} \qquad Q\text{-a.e.}$$

in $(0, +\infty]$. Integrating gives

$$\Psi_\theta(h) \le D_\theta(h)\, \Psi_Q(h)$$

in the extended reals, without any uniform lower-mass assumption. Since $m_Q \le 1$, $\Psi_Q(h) \ge 1$. If $\Psi_Q(h) < \infty$, division yields the ratio bound in eq. (6). A uniform lower bound $m_Q(z,h) \ge \eta_h > 0$ a.e. is a sufficient, but not necessary, condition for finiteness, giving $\Psi_Q(h) \le 1/\eta_h$. □

Thus $D_\theta(h)$ is a finite-scale ball-mass domination coefficient, whereas M_θ is the exact density-ratio envelope. The sandwich uses $D_\theta(h)$ as a lower bound on M_θ; no small-scale limit relating the two is required.

Structure of the rejection cost for the three-link family.

Setting. Let $L_j > 0$ with $\sum_j L_j = 1$, and let the joints be i.i.d. uniform. The end-point map is

$$F_L = \sum_j L_j e^{iU_j}.$$

The joint-uniform law makes P_L rotation invariant, so its planar density has a radial version $\bar{p}_{F,L}$ with $p_{F,L}(x) = \bar{p}_{F,L}(\|x\|_2)$ a.e., and the radial density is

$$p_{R,L}(r) = 2\pi r \, \bar{p}_{F,L}(r)$$

as in corollary 4. The deployment law is uniform on the unit disk, whose radial density is

$$q_R(r) = 2r.$$

Common support corresponds to

$$\max_j L_j \le \frac{1}{2}.$$

Formally, define

$$P_L := (F_L)_{\#}\mathrm{Unif}(\mathbb{T}^3), \qquad \mathrm{rad}(x) := \|x\|_2 \quad (x \in \mathbb{R}^2),$$

and

$$P_{R,L} := (\mathrm{rad})_{\#}P_L, \qquad Q_R := (\mathrm{rad})_{\#}Q.$$

If $Z_L \sim P_L$, write

$$R_L := \mathrm{rad}(Z_L),$$

and let p_{1-R_L} denote the density of $1 - R_L$. The exact rejection cost is

$$M_L := \left\| \frac{dQ}{dP_L} \right\|_{L^\infty(P_L)} \in [1, +\infty]$$

in the extended reals. Finiteness of M_L is a separate condition required only when a finite-envelope sampler is invoked.

Lemma 8 (Common support implies domination)**.** *If* $\max_j L_j \le 1/2$ *then* $Q \ll P_L$.

Proof. The reachable set is the closed unit disk. By Morse–Sard, the critical values of F_L are Lebesgue-null. For every interior regular value x, the fiber $F_L^{-1}(x)$ is a nonempty compact one-dimensional regular submanifold of positive length. Since the source density $(2\pi)^{-3}$ is positive, theorem 3 gives

$$p_{F,L}(x) = \int_{F_L^{-1}(x)} \frac{\varrho}{J_{T_\theta}}\, d\mathcal{H}^1 > 0.$$

Thus $p_{F,L} > 0$ Lebesgue-a.e. on the unit disk, so the absolutely continuous deployment law is dominated by P_L. When $\max_j L_j = 1/2$, the origin may be a critical value, but it lies in the same Lebesgue-null critical-value set. □

Proposition 7 (Reduction, boundary limit, center anchor, interior bottleneck)**.** (a) *Under area-uniform deployment, wherever* $\bar{p}_{F,L}(r) > 0$, $w_L(r) = q_R(r)/p_{R,L}(r) = 1/(\pi \bar{p}_{F,L}(r))$, *and under* $Q \ll P_L$, $M_L = 1/(\pi\, \mathrm{ess\,inf}_{r\in(0,1)} \bar{p}_{F,L}(r))$ *with* $1/0 = +\infty$*: the worst rejection cost sits where the end-point density is thinnest.* (b) *For every positive* L *with* $\sum_j L_j = 1$, $\lim_{r\uparrow 1} q_R(r)/p_{R,L}(r) = 4\pi\sqrt{L_1L_2L_3}$ *(a radial-density-ratio limit); under common support it is the outer-boundary limit of the exact rejection ratio.* (c) *For a strict triangle (*$\max_j L_j < 1/2$*),* $\bar{p}_{F,L}(0) = 1/(4\pi^2\,\mathrm{Area})$ *with* Area *the area of the triangle with sides* (L_1, L_2, L_3)*; hence* $w_L(0^+) = 4\pi\,\mathrm{Area} \le \pi\sqrt{L_1L_2L_3}$*, one quarter of the boundary limit, with equality iff equilateral.* (d) *The full-family identity* $M_L = 4\pi\sqrt{L_1L_2L_3}$ *is false: any Radon–Nikodym ratio has* $\mathrm{ess\,sup} \ge 1$*, while e.g.* $L = (0.495, 0.495, 0.01)$ *has* $4\pi\sqrt{L_1L_2L_3} \approx 0.622 < 1$*. More generally, boundary dominance (*$M_L = 4\pi\sqrt{L_1L_2L_3}$*) can hold only if* $L_1L_2L_3 \ge 1/(16\pi^2)$. (e) *On the nonempty relatively open set* $\mathcal{U}_{\mathrm{int}} := \{L : \max_j L_j < 1/2,\ L_1L_2L_3 < 1/(16\pi^2)\}$, $M_L \ge 1 > b_L := 4\pi\sqrt{L_1L_2L_3}$*, and for every* $c \in (b_L, 1)$ *there is* $\delta > 0$ *with* $P_{R,L}(\{r \le 1 - \delta : w_L(r) > c\}) > 0$*: a positive-mass interior region carries the bottleneck.*

Proof. **(a)** The coarea representation gives $P_{R,L} \ll dr$, while $q_R(r) = 2r > 0$ on $(0, 1)$ gives $Q_R \sim dr$. Together with $Q_R \ll P_{R,L}$, the three measures $Q_R, P_{R,L}, dr$ are mutually absolutely continuous on $(0, 1)$, so their essential extrema use the same null sets. Rotational symmetry reduces the disk ratio to the radial ratio, and cancelling the common factor $2r$ gives part (a); this cancellation is specific to area-uniform deployment.

(b) Let

$$\phi_j := U_j - U_1.$$

By the unimodular coordinate change, (ϕ_2, ϕ_3) is uniform on $\mathbb{T}^2$. The radial map

$$r(\phi) = |L_1 + L_2 e^{i\phi_2} + L_3 e^{i\phi_3}|$$

attains its maximum 1 only at $\phi = 0$, and

$$1 - r = \frac{1}{2}\phi^\top A_L \phi + o(|\phi|^2),$$

where

$$A_L = \operatorname{diag}(L_2, L_3) - vv^\top, \qquad v = (L_2, L_3)^\top.$$

The matrix determinant lemma gives

$$\det A_L = L_2 L_3 (1 - L_2 - L_3) = L_1 L_2 L_3 > 0,$$

so the maximum is nondegenerate.

By the Morse lemma there is a local C^1 chart y such that

$$1 - r = \frac{1}{2}|y|^2$$

and whose Jacobian at the origin is

$$J(0) = \frac{1}{\sqrt{\det A_L}}.$$

For $t = 1 - r \downarrow 0$, the level-set coarea formula ($k = 2$, $d = 1$) gives

$$p_{1-R_L}(t) = \frac{1}{4\pi^2} \int_{|y|=\sqrt{2t}} \frac{J(y)}{|y|}\, ds \longrightarrow \frac{1}{4\pi^2}\, J(0)\, 2\pi = \frac{1}{2\pi\sqrt{L_1 L_2 L_3}}.$$

Therefore

$$\lim_{r \uparrow 1} \frac{q_R(r)}{p_{R,L}(r)} = \frac{2}{1/(2\pi\sqrt{L_1 L_2 L_3})} = 4\pi\sqrt{L_1 L_2 L_3}.$$

The density-ratio limit holds for every positive L summing to one and needs no triangle condition; its rejection-sampling interpretation additionally requires $Q \ll P_L$, as in lemma 8. For $\max_j L_j > 1/2$ only the density-ratio statement applies.

(c) Rotation fixes the origin and preserves density values, so

$$\bar{p}_{F,L}(0) = p_G(0), \qquad G(\phi_2, \phi_3) = L_1 + L_2 e^{i\phi_2} + L_3 e^{i\phi_3},$$

a $k = d = 2$ map with source density $1/(4\pi^2)$. For a strict triangle, $G = 0$ has exactly two mirror solutions and each is regular, with

$$|\det DG| = L_2 L_3 |\sin(\phi_3 - \phi_2)| = 2\,\mathrm{Area}.$$

The branch-sum formula therefore gives

$$\bar{p}_{F,L}(0) = \frac{1}{4\pi^2} \cdot 2 \cdot \frac{1}{2\,\mathrm{Area}} = \frac{1}{4\pi^2\,\mathrm{Area}}.$$

Equivalently, the same value follows from the $\mathbb{T}^3$ fiber integral: the two fibers have Hausdorff length $2\pi\sqrt{3}$, the normal Jacobian is $2\sqrt{3}\,\mathrm{Area}$ by Cauchy–Binet, and the source density is $1/(8\pi^3)$.

Heron's formula together with

$$(a + b - c)(b + c - a)(c + a - b) \le abc$$

(put $x_j = \frac{1}{2} - L_j > 0$, so that $L_j = x_k + x_l \ge 2\sqrt{x_k x_l}$) gives

$$\mathrm{Area} \le \frac{\sqrt{L_1 L_2 L_3}}{4},$$

with equality exactly in the equilateral case.

(d) Because every Radon–Nikodym derivative integrates to one,

$$\int w\, dP = 1 \quad \Longrightarrow \quad \operatorname{ess\,sup} w \ge 1.$$

Hence if

$$4\pi\sqrt{L_1 L_2 L_3} < 1,$$

the boundary limit cannot equal the global essential supremum. The example and necessary condition in the statement follow immediately.

(e) The set $\mathcal{U}_{\text{int}}$ is nonempty; for example,

$$L^{\epsilon} = \left(\frac{1-\epsilon}{2}, \frac{1-\epsilon}{2}, \epsilon\right)$$

belongs to it for sufficiently small ϵ, and the set is relatively open. On $\mathcal{U}_{\text{int}}$, lemma 8 gives $Q \ll P_L$, hence

$$w_L = \frac{dQ_R}{dP_{R,L}}, \qquad \int w_L \, dP_{R,L} = 1.$$

If $M_L < 1$, the integral would be strictly below one, so

$$M_L \geq 1 > b_L.$$

For $c \in (b_L, 1)$, choose $\delta > 0$ such that $w_L < c$ on $(1-\delta, 1)$, possible by the boundary limit. If $P_{R,L}(w_L > c) = 0$, then

$$1 = \int w_L \, dP_{R,L} \leq c < 1,$$

a contradiction. Therefore the set $\{w_L > c\}$ has positive $P_{R,L}$-mass and lies in $\{r \leq 1-\delta\}$. □

Part (e) establishes a positive-mass interior bottleneck without requiring the global envelope to be attained at an interior point or to be finite. In the 23-design intervention study, the numerical candidate envelope is

$$\widehat{M_L} = \max\left\{4\pi\sqrt{L_1 L_2 L_3},\ \text{nonboundary radial-grid maximum of } q_R/\hat{p}_{R,L}\right\}.$$

For every common-support design in that study, the nonboundary numerical maximum does not exceed the boundary term, so the candidate envelope is boundary-determined; at the three anchor designs, $[0.48, 0.42, 0.10] \to 1.7842$, $[0.36, 0.33, 0.31] \to 2.4116$, and $[0.48, 0.45, 0.07] \to 1.5452$, the values of $4\pi\sqrt{L_1 L_2 L_3}$. This numerical observation is specific to the evaluated design set; proposition 7(e) identifies a relatively open regime $\mathcal{U}_{\text{int}}$ in which an interior bottleneck dominates. The local boundary asymptotic is identified by the link product and boundary Hessian, whereas the global acquisition cost depends on the thinnest coverage over the full task space. Numerical details are given in section C.2.

B.4.2 Allocation

Assumptions. We use the following conditions.

(U0) There are $K < \infty$ sources with task-space laws $P_1, \ldots, P_K$ on a common metric task space. For fixed $n \geq 1$ and $h > 0$,

$$m_i(z, h) := P_i(\mathbb{B}_h(z))$$

is measurable.

(U1) Strict finite-scale common support: for some $\eta_h > 0$,

$$\eta_h \leq m_i(z, h) \leq 1$$

for Q-a.e. z and every source i. This is a mass condition at the fixed scale h, stronger than common topological support.

(U2) Mixture acquisition: independently for each example, source i is selected with probability α_i, followed by one draw from that source.

(U3) The conditional response is shared across sources.

(U4) All convexity statements below are for fixed (n, h). Nonnegative averages of convex objectives over any fixed collection of numerical mass-field approximations remain convex.

(U5) Without (U1), the extended-valued functional

$$\Psi(\alpha) := \int \frac{dQ}{m_\alpha} \in [1, +\infty]$$

remains convex on Δ_K. The C^1, first-order optimality, and Frank–Wolfe statements below use (U1).

Lemma 9 (Mixing)**.** *Under (U0)(U2) the marginal covariate law is $P_\alpha = \sum_i \alpha_i P_i$, so $m_\alpha(z,h) = \sum_i \alpha_i m_i(z,h)$ is affine in α and the in-ball count is exactly* $\mathrm{Binomial}(n, m_\alpha(z,h))$.

Proof. By the law of total probability, the marginal covariate law is $P_\alpha = \sum_i \alpha_i P_i$. For fixed (z,h), the indicators $\mathbf{1}\{Z_j \in \mathbb{B}_h(z)\}$ are therefore i.i.d. Bernoulli with success probability $m_\alpha(z,h)$, giving the stated binomial count. □

Proof of proposition 2(a). By lemma 9, m_α is affine in α and satisfies $m_\alpha \geq \eta_h > 0$. Since $t \mapsto 1/t$ is convex on $(0,\infty)$, affine composition and integration imply that Ψ is convex, with values in $[1, 1/\eta_h]$.

The partial derivative of the integrand is

$$-\frac{m_i}{m_\alpha^2},$$

whose absolute value is bounded by $1/\eta_h^2$. Dominated convergence therefore permits differentiation under the integral:

$$\partial_i \Psi = -A_i(\alpha).$$

Finally, the identity

$$\sum_i \alpha_i A_i(\alpha) = \int \frac{m_\alpha}{m_\alpha^2}\, dQ = \Psi(\alpha)$$

holds pointwise in the integrand. □

Proof of proposition 2(b). The feasible directions from e_b are

$$\sum_{i \neq b} \gamma_i (e_i - e_b), \qquad \gamma_i \geq 0.$$

For each vertex direction,

$$D_{e_i - e_b}\Psi(e_b) = \int \frac{m_b - m_i}{m_b^2}\, dQ = \Psi(e_b) - A_i(e_b) = \Gamma_{b\to i}.$$

A differentiable convex function is globally minimal at a point iff all feasible directional derivatives there are nonnegative. The general first-order optimality condition for convex programs on the simplex therefore gives part (b), and the multiplier equals $\Psi(\alpha^\star)$ by the identity in part (a).

Pointwise dominance $m_b \geq m_i$ a.e. for every i implies $\Gamma_{b\to i} \geq 0$, hence optimality of e_b, but is not necessary. For example, let

$$\mathcal{Z} = \{z_1, z_2\}, \qquad Q = \mathrm{Unif}\{z_1, z_2\},$$

and

$$m_b = (0.8, 0.2), \qquad m_i = (0.9, 0.1).$$

These masses can be realized by point-mass laws at a sufficiently small scale. The fields cross, but

$$\Gamma_{b\to i} = \frac{1}{2}\left[\frac{0.8 - 0.9}{0.64} + \frac{0.2 - 0.1}{0.04}\right] = \frac{75}{64} > 0,$$

so e_b is optimal (and unique by the next observation).

If $\Gamma_{b\to i} > 0$ for every $i \neq b$, then for any $\alpha \neq e_b$,

$$\nabla\Psi(e_b)^\top(\alpha - e_b) = \sum_{i \neq b} \alpha_i \Gamma_{b\to i} > 0,$$

and convexity yields $\Psi(\alpha) > \Psi(e_b)$, so e_b is the unique optimum. Conversely, if some $\Gamma_{b\to i} < 0$, then e_b is not optimal and moving a sufficiently small amount of mass toward source i strictly improves the objective. This is the sense in which source i is locally actionable relative to b: the directional condition shows that b is not optimal and that a small two-source perturbation improves the objective, but it does not by itself determine the support or interiority of the global optimizer. □

Corollary 7 (Exact mixture gap and equal-profile complementarity)**.** *Under (U0)–(U2), fix h and two sources $1, 2$. For $\alpha \in [0,1]$ write $m_{12,\alpha}(z,h) := \alpha m_1(z,h) + (1-\alpha)m_2(z,h)$ and $\Psi_{12}(\alpha) := \Psi(\alpha e_1 + (1-\alpha)e_2)$. Then, with the arguments (z,h) suppressed in integrands,*

$$\alpha\Psi(e_1) + (1-\alpha)\Psi(e_2) - \Psi_{12}(\alpha) = \alpha(1-\alpha)\int \frac{(m_1-m_2)^2}{m_1 m_2\, m_{12,\alpha}}\, dQ. \tag{11}$$

Consequently, if $\Psi(e_1) = \Psi(e_2)$ and $Q\{m_1(\cdot,h) \neq m_2(\cdot,h)\} > 0$, then $\Psi_{12}(\alpha) < \Psi(e_1) = \Psi(e_2)$ for every $0 < \alpha < 1$; moreover $\Gamma_{1\to 2} < 0$, $\Gamma_{2\to 1} < 0$, and Ψ_{12} has a unique minimizer in $(0,1)$.

Proof. For $a, b > 0$ and $\alpha \in [0,1]$, direct algebra gives

$$\frac{\alpha}{a} + \frac{1-\alpha}{b} - \frac{1}{\alpha a + (1-\alpha)b} = \frac{\alpha(1-\alpha)(a-b)^2}{ab[\alpha a + (1-\alpha)b]}.$$

Applying this identity pointwise with $a = m_1(z,h)$ and $b = m_2(z,h)$ and integrating against Q gives eq. (11); all terms are finite under (U1).

For $0 < \alpha < 1$, the integrand on the right side is positive exactly where $m_1 \neq m_2$. Hence the positive-measure condition makes the gap strictly positive, and equal endpoint values give

$$\Psi_{12}(\alpha) < \Psi(e_1).$$

By pointwise algebra from $\Gamma_{b\to i} = \int (m_b - m_i)/m_b^2\, dQ$, the directional derivatives satisfy

$$\Gamma_{1\to 2} = \Psi(e_2) - \Psi(e_1) - \int \frac{(m_1-m_2)^2}{m_1^2 m_2}\, dQ, \qquad \Gamma_{2\to 1} = \Psi(e_1) - \Psi(e_2) - \int \frac{(m_1-m_2)^2}{m_2^2 m_1}\, dQ,$$

so equal endpoint values make both strictly negative. Finally, differentiation under the integral, justified by (U1), gives

$$\Psi_{12}''(\alpha) = 2\int \frac{(m_1-m_2)^2}{m_{12,\alpha}^3}\, dQ > 0$$

on $(0,1)$. Hence Ψ_{12} is strictly convex, and because every interior value lies below both endpoint values, its unique minimizer lies in $(0,1)$. □

Remarks. (a) When $\Psi(e_1) \neq \Psi(e_2)$, eq. (11) still gives strict improvement over the chord, but an interior mixture need not beat the better endpoint. The directional identities give the exact criterion: source 2 is actionable relative to source 1 iff

$$\int \frac{(m_1-m_2)^2}{m_1^2 m_2}\, dQ > \Psi(e_2) - \Psi(e_1).$$

(b) For the symmetric two-point example $\mathcal{Z} = \{z_1, z_2\}$, $Q = \mathrm{Unif}\{z_1, z_2\}$, $m_1 = (0.8, 0.2)$, and $m_2 = (0.2, 0.8)$, we obtain $\Psi(e_1) = \Psi(e_2) = 3.125$, $\Psi_{12}(\frac{1}{2}) = 2$, and $\Gamma_{1\to 2} = \Gamma_{2\to 1} = -225/32$. (c) The uniqueness statement concerns only the two-source segment; it does not imply uniqueness of weights when $K > 2$.

Lemma 10 (Frank–Wolfe gap)**.** *Let $F : \Delta_K \to \mathbb{R}$ be convex and C^1 and*

$$g_F(\alpha) := \max_{s\in\Delta_K} \nabla F(\alpha)^\top(\alpha - s) = \sum_i \alpha_i \partial_i F(\alpha) - \min_i \partial_i F(\alpha) \geq 0.$$

Then for every optimal $\alpha^\star$,

$$F(\alpha) - F(\alpha^\star) \leq \nabla F(\alpha)^\top(\alpha - \alpha^\star) \leq g_F(\alpha),$$

and $g_F(\alpha) = 0$ iff α is optimal.

Proof. A linear function on the simplex attains its maximum at a vertex, which gives the closed form of g_F. The first-order convexity inequality gives the first bound, and maximizing its right-hand side over the simplex gives the second. The condition $g_F(\alpha) = 0$ is exactly the first-order optimality condition on the simplex. □

Proof of proposition 2(c). Apply lemma 10 with $F = \Psi$ and $\partial_i \Psi = -A_i(\alpha)$. Then

$$g_\Psi(\alpha) = \max_i A_i(\alpha) - \Psi(\alpha).$$

At a vertex,

$$g_\Psi(e_b) = \max_i(-\Gamma_{b\to i}) \vee 0,$$

consistent with part (b). □

Proposition 8 (Monotonicity and convexity of the combined risk)**.** *Let $r_n(m) = B^2(1-m)^n + \sigma^2 g_n(m)$ on $[0,1]$. (i) For every $n \geq 1$, r_n is nonincreasing iff $B^2 \geq \sigma^2$. (ii) r_1 is affine; for $n \geq 2$, r_n is convex iff $B^2 \geq \frac{3}{2}\sigma^2$.*

Proof. In the Bernstein basis

$$b_{k,n}(m) = \binom{n}{k} m^k (1-m)^{n-k},$$

write

$$r_n = \sum_{k=0}^{n} \ell_k b_{k,n}, \qquad \ell_0 = B^2, \qquad \ell_k = \frac{\sigma^2}{k} \quad (k \geq 1).$$

Then

$$r_n' = n \sum_{k=0}^{n-1} (\Delta \ell_k) b_{k,n-1},$$

with

$$\Delta\ell_0 = \sigma^2 - B^2, \qquad \Delta\ell_k = \sigma^2 \left(\frac{1}{k+1} - \frac{1}{k}\right) \leq 0 \quad (k \geq 1),$$

and

$$r_n'' = n(n-1) \sum_{k=0}^{n-2} (\Delta^2 \ell_k) b_{k,n-2},$$

with

$$\Delta^2\ell_0 = B^2 - \frac{3}{2}\sigma^2, \qquad \Delta^2\ell_k = \frac{2\sigma^2}{k(k+1)(k+2)} \geq 0 \quad (k \geq 1).$$

Sufficiency follows from nonnegativity of the Bernstein basis. For necessity, at $m = 0$ only the $k = 0$ basis function is nonzero, so

$$r_n'(0) = n(\sigma^2 - B^2), \qquad r_n''(0) = n(n-1)\left(B^2 - \frac{3}{2}\sigma^2\right) \quad (n \geq 2).$$

If the respective condition fails, r_n is strictly increasing, respectively strictly concave, near zero. □

Proof of finite-n allocation-objective convexity. Under

$$B^2 \geq \frac{3}{2}\sigma^2,$$

proposition 8 gives convexity of r_n. Since m_α is affine in α, integration gives convexity of

$$\mathcal{J}_{n,h}(\alpha) = \int r_n(m_\alpha(z,h))\, dQ(z)$$

on Δ_K. The same conclusion holds for nonnegative averages over a fixed collection of such objectives, as in (U4). Moreover, $\mathcal{J}_{n,h}$ is C^1 with

$$\partial_i \mathcal{J}_{n,h} = \int r_n'(m_\alpha) m_i \, dQ,$$

so lemma 10 supplies a global first-order suboptimality certificate.

The values used in section C.3, $B = 1$ and $\sigma = 0.1$, satisfy both conditions of proposition 8. The function g_n alone is neither convex nor concave in general: g_2 is strictly concave, while $g_3''(m) = 11m - 9$ changes sign. Convexity of $\mathcal{J}_{n,h}$ therefore comes from the combined fallback-plus-noise functional r_n, with the fallback term compensating for the noise term exactly at the threshold $B^2 = \frac{3}{2}\sigma^2$. □

The results above are population statements involving the true deployment law and mass fields. In section C.3, the numerical optimization uses a fixed discretized objective based on the stated quadrature grid and scrambles; its numerical FW gaps are denoted by $\hat{g}$ to distinguish them from the population gaps. Under homoscedastic noise and the constant target $f \equiv B$, the zero-fallback NW risk is exactly $\mathcal{J}_{n,h}$. For general $|f| \le B$, $\mathcal{J}_{n,h}$ is the allocation-dependent target-free part of the certificate; the Hölder bias term $\Lambda^2 h^{2\beta}$ is independent of α and therefore does not affect the optimizer at fixed h.

Mass-field approximation error. The allocation results above use population mass fields, whereas numerical implementations approximate them from forward samples. The next proposition quantifies the effect of uniform mass-field error with the deployment integral held exact. Numerical quadrature and scramble-level stability are addressed separately below.

Proposition 9 (Stability under mass-field approximation)**.** *Assume (U0)–(U2), with the lower mass bound $m_i(z,h) \ge \eta_h$ from (U1) for Q-a.e. z and every source i. Let $\widetilde{m}_i : \mathcal{Z} \to [0,1]$ be measurable approximations satisfying*

$$\operatorname*{ess\,sup}_{z\sim Q} |\widetilde{m}_i(z,h) - m_i(z,h)| \le \delta_m \qquad \text{for every source } i.$$

For $\alpha \in \Delta_K$, define

$$\widetilde{m}_\alpha := \sum_i \alpha_i \widetilde{m}_i, \qquad \widetilde{\Psi}(\alpha) := \int \frac{dQ(z)}{\widetilde{m}_\alpha(z,h)}, \qquad \widetilde{\mathcal{J}}_{n,h}(\alpha) := \int r_n(\widetilde{m}_\alpha(z,h))\, dQ(z).$$

(a) *If $\delta_m < \eta_h$, then $\widetilde{\Psi}$ is convex and C^1 on Δ_K, and*

$$\sup_{\alpha\in\Delta_K} \left|\widetilde{\Psi}(\alpha) - \Psi(\alpha)\right| \le \Delta_m^{\Psi} := \frac{\delta_m}{\eta_h(\eta_h - \delta_m)}.$$

(b) *For every $\delta_m \ge 0$,*

$$\sup_{\alpha\in\Delta_K} \left|\widetilde{\mathcal{J}}_{n,h}(\alpha) - \mathcal{J}_{n,h}(\alpha)\right| \le \Delta_{m,n}^{\mathcal{J}} := n \max\left\{|B^2 - \sigma^2|, \frac{\sigma^2}{2}\right\} \delta_m.$$

(c) *Let $\widetilde{\alpha}_\Psi \in \Delta_K$ be any candidate and let $\widetilde{g}_\Psi := g_{\widetilde{\Psi}}(\widetilde{\alpha}_\Psi)$ be its Frank–Wolfe gap for the approximate profile objective (lemma 10). If $\delta_m < \eta_h$, then*

$$\Psi(\widetilde{\alpha}_\Psi) - \min_{\Delta_K} \Psi \le 2\Delta_m^{\Psi} + \widetilde{g}_\Psi.$$

If $B^2 \ge \frac{3}{2}\sigma^2$, let $\widetilde{\alpha}_{\mathcal{J}} \in \Delta_K$ be any candidate and $\widetilde{g}_{\mathcal{J}} := g_{\widetilde{\mathcal{J}}_{n,h}}(\widetilde{\alpha}_{\mathcal{J}})$. Then

$$\mathcal{J}_{n,h}(\widetilde{\alpha}_{\mathcal{J}}) - \min_{\Delta_K} \mathcal{J}_{n,h} \le 2\Delta_{m,n}^{\mathcal{J}} + \widetilde{g}_{\mathcal{J}}.$$

Proof. For every $\alpha \in \Delta_K$,

$$|\widetilde{m}_\alpha - m_\alpha| \le \sum_i \alpha_i |\widetilde{m}_i - m_i| \le \delta_m \qquad Q\text{-a.e.}$$

Hence $m_\alpha \ge \eta_h$ and, when $\delta_m < \eta_h$, $\widetilde{m}_\alpha \ge \eta_h - \delta_m$.

For part (a), the proof of proposition 2(a), with $\eta_h - \delta_m$ as the lower mass bound, gives convexity and C^1 regularity of $\widetilde{\Psi}$. Moreover,

$$\left|\frac{1}{\widetilde{m}_\alpha} - \frac{1}{m_\alpha}\right| = \frac{|\widetilde{m}_\alpha - m_\alpha|}{m_\alpha \widetilde{m}_\alpha} \le \frac{\delta_m}{\eta_h(\eta_h - \delta_m)}.$$

Integrating with respect to Q and taking the supremum over α proves (a).

For part (b), write the Bernstein expansion used in the proof of proposition 8 as

$$r_n' = n \sum_{k=0}^{n-1} (\Delta\ell_k) b_{k,n-1}, \qquad \Delta\ell_0 = \sigma^2 - B^2, \qquad \Delta\ell_k = -\frac{\sigma^2}{k(k+1)} \quad (k \ge 1).$$

Since the Bernstein basis is nonnegative and sums to one,

$$|r_n'(m)| \le n \max\left\{|B^2 - \sigma^2|, \frac{\sigma^2}{2}\right\} \qquad (m \in [0,1]),$$

so r_n has the displayed global Lipschitz bound on $[0,1]$. Both m_α and $\widetilde{m}_\alpha$ take values in $[0,1]$, hence

$$|r_n(\widetilde{m}_\alpha) - r_n(m_\alpha)| \le n \max\left\{|B^2 - \sigma^2|, \frac{\sigma^2}{2}\right\} \delta_m$$

pointwise, and integration gives (b).

For part (c), let $\alpha_\Psi^\star$ minimize Ψ. By part (a) and the Frank–Wolfe bound of lemma 10 applied to $F = \widetilde{\Psi}$,

$$\begin{aligned}\Psi(\widetilde{\alpha}_\Psi) &\le \widetilde{\Psi}(\widetilde{\alpha}_\Psi) + \Delta_m^\Psi \\ &\le \widetilde{\Psi}(\alpha_\Psi^\star) + \widetilde{g}_\Psi + \Delta_m^\Psi \\ &\le \Psi(\alpha_\Psi^\star) + \widetilde{g}_\Psi + 2\Delta_m^\Psi,\end{aligned}$$

which proves the profile bound.

If $B^2 \ge \frac{3}{2}\sigma^2$, proposition 8 makes r_n convex, so $\widetilde{\mathcal{J}}_{n,h}$ is convex and C^1 on the simplex by the argument of the finite-n paragraph above with $\widetilde{m}_\alpha$ in place of m_α. Applying the same three-step argument with part (b) gives the second inequality. □

For a fully discretized allocation objective, a uniform quadrature-error bound enters the same three-step argument additively. The selection/evaluation scramble check in section C.3 assesses objective-level numerical stability, rather than estimating δ_m or bounding quadrature error.

B.5 Decision Lemma

This subsection gives the finite-candidate argmin stability bound used to interpret the decision analysis in section 6. It explains why lower $L^2(Q)$ regression risk alone need not imply a better argmin decision.

Lemma 11 (Argmin stability). *Let $\mathcal{Z}_{\mathrm{cand}} \subset \mathcal{Z}$ be a nonempty finite candidate set, $f^* = \min_{\mathcal{Z}_{\mathrm{cand}}} f$, $\mathcal{Z}_{\mathrm{cand}}^\star = \arg\min_{\mathcal{Z}_{\mathrm{cand}}} f$, and $\hat{z} \in \arg\min_{\mathcal{Z}_{\mathrm{cand}}} \hat{f}$. Then*

$$f(\hat{z}) - f^* \le 2\|\hat{f} - f\|_{\infty, \mathcal{Z}_{\mathrm{cand}}}.$$

If

$$\Delta := \min_{z \in \mathcal{Z}_{\mathrm{cand}} \setminus \mathcal{Z}_{\mathrm{cand}}^\star} (f(z) - f^*)$$

(with $\Delta = +\infty$ when $\mathcal{Z}_{\mathrm{cand}}^\star = \mathcal{Z}_{\mathrm{cand}}$) and $\|\hat{f} - f\|_{\infty, \mathcal{Z}_{\mathrm{cand}}} < \Delta/2$, then $\hat{z} \in \mathcal{Z}_{\mathrm{cand}}^\star$. If

$$D_f := \max_{\mathcal{Z}_{\mathrm{cand}}} f - \min_{\mathcal{Z}_{\mathrm{cand}}} f > 0,$$

the normalized regret satisfies

$$\frac{f(\hat{z}) - f^*}{D_f} \le \frac{2\|\hat{f} - f\|_{\infty, \mathcal{Z}_{\mathrm{cand}}}}{D_f}.$$

Proof. Let $e := \|\hat{f} - f\|_{\infty, \mathcal{Z}_{\mathrm{cand}}}$. For any $z^* \in \mathcal{Z}_{\mathrm{cand}}^\star$,

$$f(\hat{z}) \le \hat{f}(\hat{z}) + e \le \hat{f}(z^*) + e \le f^* + 2e,$$

which proves the first inequality. If $e < \Delta/2$ and $\hat{z} \notin \mathcal{Z}_{\mathrm{cand}}^\star$, then $\hat{z} \in \mathcal{Z}_{\mathrm{cand}} \setminus \mathcal{Z}_{\mathrm{cand}}^\star$, so

$$\Delta \le f(\hat{z}) - f^* \le 2e < \Delta,$$

a contradiction. Thus $\hat{z} \in \mathcal{Z}_{\mathrm{cand}}^\star$. When $D_f > 0$, dividing the first inequality by D_f gives the normalized-regret bound. □

The lemma shows that finite-candidate argmin stability depends on candidate-local prediction error and the decision margin Δ, rather than on aggregate $L^2(Q)$ risk alone. The corresponding margins and regret outcomes for the seven-joint study are reported in section D.

B.6 A Learner-Independent Finite-Budget Lower Bound

This subsection isolates the no-observation obstruction underlying the empty-neighborhood term in theorem 1(a). The bounds below hold in the general setting of section 2, require no NW structure, and apply to arbitrary estimators; they provide learner-independent local lower bounds complementary to the controlled-family sharpness result.

Setting. Let $(\mathcal{Z}, d_{\mathcal{Z}})$ be a separable metric space, Q a Borel probability measure on $\mathcal{Z}$, and P_θ the source law induced by the generator. Let

$$\mathcal{F} = \left\{ f : |f| \le B, \quad |f(z) - f(z')| \le \Lambda\, d_{\mathcal{Z}}(z, z')^\beta \right\}, \qquad 0 < \beta \le 1.$$

Samples $(Z_i, Y_i)_{i \le n}$ are i.i.d. with $Z_i \sim P_\theta$ and

$$Y_i = f(Z_i) + \varepsilon_i,$$

where the conditional law of ε_i given $Z_i = z$ is a fixed Markov kernel $K_\varepsilon(\cdot \mid z)$ that does not depend on f, with

$$\mathbb{E}[\varepsilon \mid Z] = 0, \qquad \mathrm{Var}(\varepsilon \mid Z) \le \sigma^2.$$

This is the model of section 2 with the noise kernel held fixed across $f \in \mathcal{F}$.

An estimator $\hat{f}$ is any measurable map from the sample to $L^2(Q)$. Write $\mathbb{E}_f$ for expectation under the sample law induced by target f and the fixed kernel K_ε; risks may take the value $+\infty$. Balls are closed,

$$\mathbb{B}_h(z_0) = \{z : d_{\mathcal{Z}}(z, z_0) \le h\}, \qquad m_\theta(z_0, h) = P_\theta(\mathbb{B}_h(z_0)),$$

as in section 2. For a nonempty $A \subseteq \mathcal{Z}$, write

$$d_{\mathcal{Z}}(z, A) := \inf_{a \in A} d_{\mathcal{Z}}(z, a).$$

We use that $z \mapsto d_{\mathcal{Z}}(z, A)$ is 1-Lipschitz and that

$$|a^\beta - b^\beta| \le |a - b|^\beta, \qquad a, b \ge 0, \quad 0 < \beta \le 1.$$

Proposition 10 (Finite-budget no-observation lower bound)**.** *Fix $z_0 \in \mathcal{Z}$ and $h > 0$. For every $n \ge 1$ and every estimator $\hat{f}$,*

$$\sup_{f \in \mathcal{F}} \mathbb{E}_f \, \|\hat{f} - f\|^2_{L^2(Q)} \;\ge\; \big(1 - m_\theta(z_0, h)\big)^n \, Q\big(\mathbb{B}_{h/2}(z_0)\big) \, \min\{B,\, \Lambda (h/2)^\beta\}^2. \tag{12}$$

Corollary 8 (Positive-Q support deficit)**.** *Let $A_r = \{z : d_{\mathcal{Z}}(z, \operatorname{supp} P_\theta) \ge r\}$. If $Q(\mathcal{Z} \setminus \operatorname{supp} P_\theta) > 0$, then there exists $r > 0$ with $Q(A_r) > 0$, and for every $n \ge 1$ and every estimator $\hat{f}$,*

$$\sup_{f \in \mathcal{F}} \mathbb{E}_f \, \|\hat{f} - f\|^2_{L^2(Q)} \;\ge\; Q(A_r) \, \min\{B,\, \Lambda r^\beta\}^2.$$

Proof of proposition 10. Define

$$u(z) := \big(h - d_{\mathcal{Z}}(z, z_0)\big)_+, \qquad g(z) := \min\{B, \Lambda\, u(z)^\beta\}.$$

Both $d_{\mathcal{Z}}(\cdot, z_0)$ and $t \mapsto (h - t)_+$ are 1-Lipschitz, hence u is 1-Lipschitz. Therefore

$$|\Lambda u(z)^\beta - \Lambda u(z')^\beta| \le \Lambda\, d_{\mathcal{Z}}(z, z')^\beta.$$

Truncation by the constant B does not increase this modulus, so $g \in \mathcal{F}$. By symmetry of $\mathcal{F}$,

$$f_+ := g, \qquad f_- := -g$$

both lie in $\mathcal{F}$, and both vanish on $\{d_{\mathcal{Z}}(\cdot, z_0) \ge h\}$.

Let

$$E := \{d_{\mathcal{Z}}(Z_i, z_0) > h \text{ for every } i \le n\}.$$

Its probability is independent of the target and equals

$$\mathbb{P}(E) = \big(1 - m_\theta(z_0, h)\big)^n.$$

On E, every Z_i lies outside $\mathbb{B}_h(z_0)$, so $f_+(Z_i) = f_-(Z_i) = 0$. Hence under either target

$$Y_i = \varepsilon_i$$

with the same conditional kernel $K_\varepsilon(\cdot \mid Z_i)$, and the two sample laws agree when restricted to E. Let $\mathbb{E}_0[\cdot; E]$ denote integration with respect to this common restricted subprobability law.

For any estimator $\hat{f}$, the parallelogram identity in $L^2(Q)$ gives

$$\|v - g\|^2 + \|v + g\|^2 = 2\|v\|^2 + 2\|g\|^2.$$

Therefore

$$\begin{aligned}\frac{1}{2}\Big(\mathbb{E}_{f_+}\|\hat{f} - f_+\|^2 + \mathbb{E}_{f_-}\|\hat{f} - f_-\|^2\Big) &\geq \frac{1}{2}\mathbb{E}_0[\|\hat{f} - g\|^2 + \|\hat{f} + g\|^2; E] \\ &= \mathbb{E}_0[\|\hat{f}\|^2 + \|g\|^2; E] \\ &\geq \|g\|^2_{L^2(Q)}\,\mathbb{P}(E),\end{aligned}$$

with the convention that an infinite risk satisfies the bound. The larger of the two risks is at least their average.

Finally, on $\mathbb{B}_{h/2}(z_0)$,

$$u(z) \geq h/2,$$

so

$$g(z) \geq \min\{B, \Lambda(h/2)^\beta\}.$$

Thus

$$\|g\|^2_{L^2(Q)} \geq Q(\mathbb{B}_{h/2}(z_0)) \min\{B, \Lambda(h/2)^\beta\}^2.$$

Combining this inequality with the value of $\mathbb{P}(E)$ proves eq. (12). □

Proof of corollary 8. The set $\operatorname{supp} P_\theta$ is closed. By definition of support, every point in $\mathcal{Z} \setminus \operatorname{supp} P_\theta$ has an open P_θ-null neighborhood. Since a separable metric space is second countable, these neighborhoods admit a countable subcover, and hence

$$P_\theta(\operatorname{supp} P_\theta) = 1.$$

Moreover,

$$\mathcal{Z} \setminus \operatorname{supp} P_\theta = \bigcup_{k \geq 1} A_{1/k},$$

with $A_{1/k}$ increasing in k. Therefore $Q(A_r) > 0$ for some $r > 0$ whenever

$$Q(\mathcal{Z} \setminus \operatorname{supp} P_\theta) > 0.$$

Define

$$g(z) := \min\{B, \Lambda\, d_{\mathcal{Z}}(z, \operatorname{supp} P_\theta)^\beta\}.$$

As above, $g \in \mathcal{F}$, and $f_\pm := \pm g \in \mathcal{F}$. Both targets vanish on $\operatorname{supp} P_\theta$. Since $Z_i \in \operatorname{supp} P_\theta$ almost surely, the entire sample has the same law under f_+ and f_-. The parallelogram argument from the proof of proposition 10 therefore gives

$$\max_{\pm} \mathbb{E}_{f_\pm}\|\hat{f} - f_\pm\|^2_{L^2(Q)} \geq \|g\|^2_{L^2(Q)}.$$

On A_r,

$$d_{\mathcal{Z}}(z, \operatorname{supp} P_\theta) \geq r,$$

so

$$\|g\|^2_{L^2(Q)} \geq Q(A_r) \min\{B, \Lambda r^\beta\}^2,$$

which proves the corollary. □

Relation to the main certificate. Both inequalities hold for every estimator after taking the supremum over the stated Hölder class, and therefore yield corresponding minimax lower bounds upon taking the infimum over estimators. Proposition 10 can remain nontrivial under common support, whereas corollary 8 treats a positive-Q support deficit. These are one-sided local lower bounds and do not characterize the optimized certificate of theorem 1(b), nor do they assert that one fixed target is hard for one particular learner.

The same no-observation factor appears in the NW analysis: theorem 1(a) contains the fallback term

$$B^2 \int (1 - m_\theta(z, h))^n \, dQ(z),$$

and for a constant target $f \equiv c$ the exact zero-fallback NW risk contains

$$c^2 \int (1 - m_\theta(z, h))^n \, dQ(z).$$

Thus eq. (12) isolates the same finite-budget no-observation mechanism.

Support deficits and a field-level lower bound. If

$$\mathbb{B}_h(z_0) \cap \operatorname{supp} P_\theta = \varnothing,$$

then $m_\theta(z_0, h) = 0$, and eq. (12) becomes independent of sample size. Corollary 8 gives the corresponding global statement whenever a positive amount of deployment mass lies outside the source support.

Finally, taking the supremum of the right-hand side of eq. (12) over (z_0, h) yields a single learner-independent lower bound determined by the full mass field together with the fixed deployment law and the Hölder-class parameters.

C Experimental Protocols and Full Results

This section gives the experimental protocols and full numerical results supporting sections 3 to 5 and the empirical scope discussed in section 8. The design families, target banks, and numerical settings are collected in section C.6, with target-class checks in table 6. Run configurations and identifiers will be released with the code.

All three-link experiments use the same common-support family, in which every design reaches the full deployment region. Their comparisons therefore isolate finite-scale coverage rather than structural reachability.

C.1 Exact Regression Risk and the Certificate

Common setting. All experiments in this subsection use the same common-support family of 23 three-link designs, with link lengths summing to one and every design reaching the full deployment region. The deployment law is the radial distribution with density $2z$ on $[0, 1]$. We use the Hölder class with $\beta = 1$, $\Lambda = 1$, and $B = 1$, homoscedastic Gaussian noise with $\sigma = 0.1$, and the zero-fallback uniform-ball NW estimator.

For each design, mass fields are estimated on a grid of 2001 task locations using 2^{20} scrambled Sobol samples per seed and eight seeds. The eight estimates are averaged once and then held fixed for all profile and risk calculations below.

Coverage across the design family. For each design we evaluate the mass field at

$$h \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.2\}$$

and compute

$$\Psi_\theta(h) = \int \frac{dQ(z)}{m_\theta(z, h)}$$

by deterministic quadrature on the task grid. This common coverage scan supplies the empirical results used in sections 3 and 4. The separation reported in fig. 3(b) is the per-scale family ratio $\max_\theta \Psi_\theta(h) / \min_\theta \Psi_\theta(h)$, which exceeds one by 13–15%: at every scale, the largest profile in the family exceeds the smallest by 13–15%. Numerical evaluation is also invariant under permutations of the link lengths to below 10^{-8}, consistent with corollary 4.

Exact risk over smooth targets. The pre-specified risk scan uses

$$h \in \{0.02, 0.05\}, \qquad n \in \{50, 200, 1000\},$$

giving six (n, h) cells, together with a fixed bank of 30 random-smooth targets generated in advance. The targets are rescaled using the 4001-point gridwise Lipschitz and maximum-absolute-value checks in table 6.

For every design, cell, and target, we evaluate the exact pointwise NW risk

$$R(z) = p_0(z) f(z)^2 + \big(1 - p_0(z)\big)\big(\mu_B(z) - f(z)\big)^2 + \big(\mathrm{Var}_B(f; z) + \sigma^2\big) g_n(m_\theta(z, h)),$$

where

$$p_0(z) := (1 - m_\theta(z, h))^n,$$

and $\mu_B(z)$ and $\mathrm{Var}_B(f; z)$ are the conditional mean and variance of $f(Z)$ under $P_\theta(\cdot \mid \mathbb{B}_h(z))$. This is the exact decomposition of section B.2.1 before replacing the local bias by its Hölder upper bound. The three terms correspond respectively to zero-fallback loss when the neighborhood is empty, squared local smoothing bias when it is nonempty, and variance from within-ball target variation plus label noise. Here p_0 and g_n are supplied by the fixed mass estimates, and the target terms are evaluated from the fixed target values and conditional ball moments on the task grid.

The pointwise risk is integrated against Q. Figure 2(a) displays the resulting risk averaged over the 30 targets against $\Psi_\theta(h)$, with both quantities shown as percentage deviations from the within-cell family median. The reported tracking statistic is different from this visual average: for each target and cell we compute the Spearman correlation across the 23 designs between exact risk and $\Psi_\theta(h)$, then report the median correlation over targets. This median lies within 0.97–1.00 in all six cells. Conditional on the fixed mass estimates, the exact finite-n calculation contains no response-simulation noise, so no Monte Carlo confidence interval is attached to these values.

Two-design finite-budget calibration. We next isolate the finite-budget effect using the low- and high-profile designs

$$(0.48, 0.42, 0.10) \qquad \text{and} \qquad (0.36, 0.33, 0.31),$$

the constant target $f \equiv B = 1$,

$$h \in \{0.01, 0.02, 0.05, 0.1\}, \qquad n \in \{50, 100, 200, 500, 1000, 2000\}.$$

For a constant target, the exact NW risk reduces to

$$R_{\mathrm{NW}}(n, h) = B^2 \int (1 - m_\theta(z, h))^n \, dQ(z) + \sigma^2 \int g_n(m_\theta(z, h)) \, dQ(z),$$

and therefore depends only on the mass field. Three cells were also checked by direct Monte Carlo.

Across this (n, h) grid, the larger of the two design risks divided by the smaller peaks at 2.27. The largest separation occurs at small budgets, where empty-neighborhood fallback contributes substantially to the risk. At $n = 2000$, the target-independent noise floor

$$F_{n,\theta}(h) = \sigma^2 \int g_n(m_\theta(z, h)) \, dQ(z)$$

satisfies

$$\frac{\big|nF_{n,\theta}(h)/\sigma^2 - \Psi_\theta(h)\big|}{\Psi_\theta(h)} \le 3.2\%$$

for both designs at all four scales, illustrating the fixed-scale profile limit of corollary 3.

Target-specific reversal. To test what the scalar profile loses spatially, we use the comparison pair

$$\theta_{\mathrm{rev},1} = (0.45, 0.42, 0.13), \qquad \theta_{\mathrm{rev},2} = (0.48, 0.45, 0.07)$$

and move the same localized bump target across nine deployment quantiles. The C^1 target is

$$f(z) = \frac{a_{\mathrm{bump}}}{2} \left(1 + \cos \frac{\pi |z - z_k|}{s}\right)$$

for $|z - z_k| \le s$ and zero otherwise, with

$$a_{\text{bump}} = \min\{B, 2\Lambda s/\pi\},$$

so that it satisfies the Hölder constraint. The centers are

$$z_k = \sqrt{u_k}, \qquad u_k \in \{0.1, 0.2, \ldots, 0.9\},$$

the nine deployment quantiles. The main-text panel uses $s = 0.1$, $h = 0.02$, and $n = 50$; the same comparison also evaluates $s = 0.05$ and the remaining cells.

For each center we report the target-dependent excess risk

$$R^{\text{NW}}_{\theta,f}(n, h) - F_{n,\theta}(h),$$

that is, exact NW risk after subtracting the target-independent label-noise floor, computed for each center on the grid. In the displayed cell, $\theta_{\text{rev},1}$ has lower excess risk at the first six centers and $\theta_{\text{rev},2}$ at the last three (fig. 2(b)). Therefore no target-independent scalar ordering of this pair reproduces all nine target-specific outcomes.

Scalar-surrogate sensitivity. As a sensitivity analysis of the coverage scan, we compare $\Psi_\theta(0.05)$ with five simple functions of the link lengths:

$$L_3, \qquad L_1/L_3, \qquad \sum_j L_j^2, \qquad \prod_j L_j, \qquad -\sum_j L_j \log L_j.$$

Their absolute Spearman correlations with $\Psi_\theta(0.05)$ are 0.978–0.991 across the 23 designs. Thus, on this family, aggregate profile ranking can be closely tracked by simple link-length summaries. The motivation for the spatial mass field therefore comes from localized coverage and cross-source complementarity, rather than aggregate ranking alone.

C.2 Intervention on Acquisition

Setup and target. We retain the common-support three-link family and statistical setting of section C.1: 23 designs, radial deployment density $2z$ on $[0, 1]$, the Hölder class with $\beta = 1$, $\Lambda = 1$, and $B = 1$, homoscedastic Gaussian noise with $\sigma = 0.1$, and zero-fallback uniform-ball NW. This experiment uses

$$h \in \{0.02, 0.05\}.$$

The primary target is the smooth compactly supported bump

$$f(z) = a_f\, \varphi\!\left(\frac{z - z_0}{s_f}\right), \qquad \varphi(t) = \begin{cases} \exp\!\left(1 - \frac{1}{1-t^2}\right), & |t| < 1, \\ 0, & |t| \ge 1, \end{cases}$$

with width $s_f = 0.05$, center $z_0 = \sqrt{0.7}$ (deployment quantile 0.7), and

$$a_f = \min\left\{B,\ \frac{\Lambda s_f}{\max|\varphi'|}\right\},$$

so that f satisfies the stated Hölder constraint. A constant target $f \equiv 1$ is used separately to check the numerical rejection implementation.

The comparison pair is

$$A = (0.48, 0.42, 0.10), \qquad B = (0.36, 0.33, 0.31),$$

with A the lower-profile and B the higher-profile design in this family.

Numerical coverage and cost quantities. Mass fields are computed on the 2001-point task grid by the deterministic conditional-distribution calculation used for this experiment, with a grid-halving convergence check. The conditional target moments $\mu_f(z, h)$ and $v_f(z, h)$ required by corollary 6 are evaluated by deterministic quadrature using 4001 points within each truncated ball, in the same task coordinate.

Let $\mathcal{G}_{3\mathrm{R}}$ denote the 2001-point grid and $\hat{m}_Q, \hat{m}_\theta$ the corresponding mass estimates. We use

$$\widehat{D}_\theta(h) := \max_{z \in \mathcal{G}_{3\mathrm{R}}} \frac{\hat{m}_Q(z,h)}{\hat{m}_\theta(z,h)}$$

as a grid estimate of the population coefficient $D_\theta(h)$. We record both the maximizing location and the maximum restricted to $z \in [0.05, 0.95]$ to distinguish an interior bottleneck from one near the task-range boundary.

Let $\hat{p}_{R,\theta}$ be the numerically integrated radial source density and define

$$\hat{w}_\theta(r) := \frac{q_R(r)}{\hat{p}_{R,\theta}(r)}, \qquad q_R(r) = 2r,$$

with q_R the exact radial deployment density. The numerical candidate envelope is

$$\widehat{M}_\theta := \max\left\{4\pi\sqrt{L_1 L_2 L_3},\ \max_r \hat{w}_\theta(r)\right\},$$

where the second term is a refined numerical-grid maximum: 1001 radii on $[0.001, 0.9999]$ are scanned and the three largest candidates are refined to 10^{-6}. Grid convergence and cumulative-distribution checks at three radii provide additional numerical validation.

The experiments use the numerical density-ratio estimate $\hat{w}_\theta$ and candidate envelope $\widehat{M}_\theta$; convergence, distributional, and Monte Carlo validation is reported below. The exact rejection identities use the exact likelihood ratio with a valid finite envelope.

Acquisition arms and resource frames. Three acquisition arms were fixed before evaluation.

Own source. We train unweighted NW on i.i.d. P_θ samples with $n \in \{50, 200\}$, using 200 Monte Carlo replicates and a 201-point evaluation grid.

Task matched. We evaluate NW with $n \in \{50, 200, 1000\}$ i.i.d. samples from Q; the finite-n risk is computed exactly.

Rejection. Proposals are drawn from P_θ and accepted numerically with probability

$$\frac{\hat{w}_\theta}{M_{\mathrm{impl}}}, \qquad M_{\mathrm{impl}} = (1 + 10^{-6})\widehat{M}_\theta,$$

without clipping. The implementation checks $\hat{w}_\theta(Z_i) \le M_{\mathrm{impl}}$ for every evaluated proposal; all evaluated proposals satisfy it. For

$$N_{\mathrm{prop}} \in \{200, 1000, 5000\},$$

the reported modeled rejection curve is the deterministic thinning-risk functional of corollary 6, evaluated with

$$s(z) = \frac{m_Q(z,h)}{\widehat{M}_\theta}.$$

This curve evaluates the thinning functional with the numerical density ratio and candidate envelope; its exact-rejection counterpart uses the exact likelihood ratio with a valid finite envelope.

The constant-target check uses 300 replicates of $N_{\mathrm{prop}} = 1000$ proposals per design at $h = 0.05$. The implemented acceptance probability is $\hat{w}_\theta / M_{\mathrm{impl}}$, where $M_{\mathrm{impl}} = (1+10^{-6})\widehat{M}_\theta$, with nominal acceptance rate $1/M_{\mathrm{impl}}$. Monte Carlo risks and the thinning functional (corollary 5) evaluated at $m_Q(z,h)/M_{\mathrm{impl}}$ use the same 201-point task grid. For both designs, the Monte Carlo mean is within 0.41 estimated standard errors of the thinning value (table 2).

For reference, we also report the mean-count comparison of section B.4.1, which replaces the random accepted sample size by

$$\bar{n}_\theta = \mathrm{round}\left(\frac{N_{\mathrm{prop}}}{\widehat{M}_\theta}\right)$$

and evaluates the same risk functional using the unthinned field m_Q. This is a descriptive comparison only; because the risk functional is nonlinear and the count is rounded, its difference from the random-count curve is not a pure Jensen effect.

Table 2: Constant-target Monte Carlo check of the rejection sampler. Risk means, SEs, and thinning values are in units of 10^{-4}. Acceptance rates pool all 300 000 proposals per design. SE is the sample standard deviation of the 300 replicate risks divided by $\sqrt{300}$. The standardized discrepancy z is the absolute difference between the Monte Carlo mean and thinning value divided by SE. CDF discrepancy is the maximum absolute empirical-CDF difference from $F_Q(r) = r^2$, evaluated at the pooled accepted radii.

Design	Empirical acceptance rate	Nominal $1/M_{\mathrm{impl}}$	MC risk mean $\pm$ SE	Thinning value	z	CDF discrepancy
A	0.561017	0.560460	2.667 ± 0.153	2.729	0.40	0.0016
B	0.414710	0.414668	4.689 ± 0.344	4.580	0.32	0.0034

The experiment therefore contains two distinct resource frames: a fixed number of labeled examples and a fixed number of source proposals. One proposal consists of one source draw and one forward evaluation, and only accepted proposals are labeled. Risk ratios are interpreted within these frames rather than by comparing their absolute risk levels.

As an auxiliary comparison, we also evaluate importance weighting on the own-source samples of the comparison pair, with density-ratio weights $\hat{w}_\theta$ and finite-scale weights m_Q/m_θ, without clipping; effective sample size is reported.

Intervention results. At a fixed label budget, task matching gives identical risk across designs, as predicted by corollary 2 (fig. 4(a)). At a fixed proposal budget, the higher-profile design B has larger modeled thinning risk than A for every tested N_{prop} and at both scales; this was the primary fixed-proposal criterion. The corresponding candidate-envelope ratio is

$$\frac{\widehat{M}_B}{\widehat{M}_A} = 1.35.$$

The numerical analogue of the cost sandwich,

$$\frac{\Psi_\theta(h)}{\Psi_Q(h)} \le \widehat{D}_\theta(h) \le \widehat{M}_\theta,$$

the primary cost criterion, holds in all 46/46 design–scale cells (fig. 4(b)). For every design in the family, the independently evaluated nonboundary grid maximum of $\hat{w}_\theta$ is no larger than the proved boundary term $4\pi\sqrt{L_1L_2L_3}$ (proposition 7(b)), so the candidate envelope is boundary-determined. This is a numerical property of the 23-design common-support family; proposition 7(e) shows that boundary dominance need not hold throughout the continuous family.

The importance-weighting comparison shows no risk improvement over unweighted NW in this experiment.

Overall, the experiment separates the two effects intended by the intervention: direct task matching removes design dependence when the number of labeled examples is fixed, whereas source-dependent rejection transfers that dependence to proposal efficiency.

C.3 Pre-Training Allocation

Selection and evaluation profiles. We compare two pre-training allocation rules on the 23-source common-support three-link family: a fixed-scale Ψ rule and a finite-n rule based on $\mathcal{J}_{n,h}$. To keep allocation choice separate from evaluation, all weights and baselines are selected using four randomized quasi-Monte Carlo (RQMC) *selection* scrambles and are scored on four disjoint *evaluation* scrambles.

Let $m^{\mathrm{sel}}_{i,s}(z,h)$ and $m^{\mathrm{ev}}_{i,t}(z,h)$ denote the mass-field estimates of source i on selection scramble $s \in \{1,\dots,4\}$ and evaluation scramble $t \in \{1,\dots,4\}$. Define the selection objectives

$$\widehat{\Psi}^{\mathrm{sel}}_h(\alpha) := \frac{1}{4}\sum_{s=1}^{4} \int \frac{dQ(z)}{\sum_i \alpha_i m^{\mathrm{sel}}_{i,s}(z,h)},$$

and

$$\widehat{\mathcal{J}}^{\mathrm{sel}}_{n,h}(\alpha) := \frac{1}{4}\sum_{s=1}^{4}\int r_n\left(\sum_i \alpha_i m^{\mathrm{sel}}_{i,s}(z,h)\right)\, dQ(z).$$

Let $\alpha^{\Psi}_{\mathrm{sel},h}$ and $\alpha^{\mathcal{J}}_{\mathrm{sel},n,h}$ denote the corresponding numerical minimizers.

For the disjoint evaluation profiles, define

$$\widehat{\Psi}^{\mathrm{ev}}_h(\alpha) := \frac{1}{4}\sum_{t=1}^{4}\int \frac{dQ(z)}{\sum_i \alpha_i m^{\mathrm{ev}}_{i,t}(z,h)},$$

and

$$\widehat{\mathcal{J}}^{\mathrm{ev}}_{n,h}(\alpha) := \frac{1}{4}\sum_{t=1}^{4}\int r_n\left(\sum_i \alpha_i m^{\mathrm{ev}}_{i,t}(z,h)\right)\, dQ(z),$$

with numerical minimizers $\alpha^{\Psi}_{\mathrm{ev},h}$ and $\alpha^{\mathcal{J}}_{\mathrm{ev},n,h}$. The evaluation minimizers are used for the stability analysis; all reported allocations and baselines are determined from the selection profiles. We suppress the (n,h) dependence of $\alpha^{\mathcal{J}}_{\mathrm{sel}}$ and the h dependence of $\alpha^{\Psi}_{\mathrm{sel}}$ where unambiguous.

Experimental design. The sources are the 23 designs of the common-support three-link family, with radial deployment law

$$dQ(z) = 2z\, dz.$$

We evaluate

$$h \in \{0.02, 0.05\}, \qquad n \in \{50, 100, 200, 1000\},$$

giving eight evaluation cells, with $B = 1$ and $\sigma = 0.1$. The finite-n rule is optimized separately in each (n,h) cell, whereas the fixed-scale Ψ rule is optimized once per scale and reused across the four budgets at that scale.

For each labeled example, mixture acquisition first draws a source index from $\mathrm{Categorical}(\alpha)$ and then one joint-uniform sample from that source. Hence the pooled covariates are i.i.d. from

$$P_\alpha = \sum_i \alpha_i P_i,$$

and the local count is exactly

$$\mathrm{Binomial}(n, m_\alpha(z,h)).$$

The experiment therefore concerns randomized source selection for each example, not fixed per-source quotas.

Allocation objectives and numerical certification. At the population level,

$$\mathcal{J}_{n,h}(\alpha) = \int \left[B^2(1 - m_\alpha(z,h))^n + \sigma^2 g_n(m_\alpha(z,h))\right] dQ(z).$$

For the constant target $f \equiv B$ under homoscedastic variance σ^2, this is the exact zero-fallback NW risk. For general $|f| \le B$, it is the allocation-dependent target-free part of the finite-n certificate and contains no target-specific local-bias information.

Under

$$B^2 \ge \tfrac{3}{2}\sigma^2,$$

$\mathcal{J}_{n,h}$ is convex on the simplex by proposition 8. Its Frank–Wolfe (FW) gap

$$g_{\mathcal{J}}(\alpha) = \max_{s \in \Delta_K} \nabla \mathcal{J}_{n,h}(\alpha)^\top (\alpha - s)$$

therefore upper-bounds global suboptimality (lemma 10).

The finite-n rule minimizes the fixed numerical approximation $\widehat{\mathcal{J}}^{\mathrm{sel}}_{n,h}$ using deterministic sequential least-squares programming (SLSQP) from the uniform start with analytic gradients. After clipping

and renormalization, the selection objective, gradient, FW gap, and simplex feasibility are recomputed at the reported weights. A solution is accepted only if

$$\hat{g}_{\mathcal{J}} \le \max\{10^{-6},\, 10^{-4}\, \widehat{\mathcal{J}}^{\mathrm{sel}}_{n,h}(\alpha^{\mathcal{J}}_{\mathrm{sel},n,h})\},$$

the simplex violation is at most 10^{-10}, and its selection-objective value is no larger than that of uniform mixing or any single-source vertex. Every reported solution satisfies these FW-gap, simplex-feasibility, and baseline-objective criteria. Additional starting points quantify the spread of near-optimal solutions.

The Ψ rule minimizes $\widehat{\Psi}^{\mathrm{sel}}_h$ at the same scale and is accepted under the analogous condition

$$\hat{g}_{\Psi} \le \max\{10^{-6},\, 10^{-4}\, \widehat{\Psi}^{\mathrm{sel}}_h(\alpha^{\Psi}_{\mathrm{sel},h})\}.$$

Unlike the finite-n rule, it depends on h but not on n.

Baselines and target families. We compare against four single allocation baselines: uniform mixing over all 23 sources; the single source minimizing $\mathcal{J}_{n,h}$ in the cell; the single source minimizing Ψ at that scale; and the scalar-selected source

$$(0.48, 0.45, 0.07).$$

The last source is selected by all five link-length rules—smallest L_3, largest L_1/L_3, smallest product, smallest entropy, and largest $\sum_j L_j^2$—so duplicate selections are treated as one baseline.

The confirmatory targets are disjoint from those used for the certificate experiment. They consist of twenty cosine bumps from section C.1, with

$$s \in \{0.05, 0.1\}, \qquad z = \sqrt{u}, \qquad u \in \{0.05, 0.15, \ldots, 0.95\},$$

and a separately generated bank of 30 random-smooth targets. The $u = 0.95$ bump is truncated at $z = 1$, which preserves the target class. The smooth targets are rescaled using the fixed 4001-point gridwise Lipschitz and maximum-absolute-value checks in table 6.

Risk is evaluated deterministically by the exact finite-n NW formula on the evaluation profiles. For mixture acquisition, the corresponding moments are computed under P_α by mixing the unnormalized first and second ball moments before normalization. For each comparator and cell, we form the risk ratio

$$\frac{\text{allocation risk}}{\text{baseline risk}}$$

for every target. We take the median ratio within the bump family and within the smooth family, then average those two medians with equal weight. This is the *family-balanced risk-ratio score*; values below one favor the allocation. A cell is counted as a win only when this score is below one and neither target-family median exceeds 1.01. In every cell of table 3 with score below one, this additional cap is also satisfied, so the win counts can be read off the table.

Allocation results. Four criteria were fixed before the confirmatory run.

1. For the constant target $f \equiv B$, the finite-n allocation's mean evaluation-profile risk must be no larger than uniform mixing in all eight cells. This criterion was met in $8/8$ cells, with median relative gain 14%.
2. The family-balanced score must be below one relative to uniform mixing in at least $7/8$ cells. It was met in 8/8 cells.
3. The family-balanced score must be below one in at least $7/8$ cells against each single-source baseline. The results were 8/8 against the best constant-target source, 7/8 against the best-profile source, and 7/8 against the scalar-selected source; all three criteria were met.
4. Numerical stability is assessed against a 0.5% relative-regret threshold on the discretized evaluation objective, recomputed on the evaluation profiles with the selection-stage allocation held fixed. The threshold is met in seven of eight cells; the remaining cell $n = 1000$, $h = 0.05$ gives 0.55%, while all others are at most 0.16%.

For the stability calculation, let $\hat{g}^{\mathrm{ev}}_{\mathcal{J},n,h}(\alpha)$ be the discretized FW gap of $\widehat{\mathcal{J}}^{\mathrm{ev}}_{n,h}$ at α, and define

$$L^{\mathrm{ev}}_{n,h} := \widehat{\mathcal{J}}^{\mathrm{ev}}_{n,h}(\alpha^{\mathcal{J}}_{\mathrm{ev}}) - \hat{g}^{\mathrm{ev}}_{\mathcal{J},n,h}(\alpha^{\mathcal{J}}_{\mathrm{ev}}).$$

This is an FW lower bound on the minimum of the *discretized evaluation objective*, not the population objective, and is positive in every evaluated cell. For any candidate α, define

$$\mathrm{Reg}^{\mathrm{ub}}_{n,h}(\alpha) := \frac{\widehat{\mathcal{J}}^{\mathrm{ev}}_{n,h}(\alpha) - L^{\mathrm{ev}}_{n,h}}{L^{\mathrm{ev}}_{n,h}}.$$

The reported stability quantity is

$$\mathrm{Reg}^{\mathrm{ub}}_{n,h}(\alpha^{\mathcal{J}}_{\mathrm{sel}}).$$

Additional analyses and objective specificity. A separate two-sided certificate-tightening analysis bounds the worst-cell relative regret on the same discretized objective in

$$[6.02 \times 10^{-4}, 6.72 \times 10^{-4}],$$

approximately 0.0602%–0.0672%, indicating that the original 0.55% upper bound was conservative: the stability miss reported above reflects certificate resolution on that discretized objective.

The two allocation objectives favor different regimes. On the family-balanced heterogeneous-target score, the Ψ-mixture has the lower score in all eight cells; on the constant-target objective $\mathcal{J}_{n,h}$ optimized directly by the finite-n rule, the finite-n allocation is lower in seven of eight cells. The two objectives therefore favor different allocations on heterogeneous versus constant targets, consistent with their definitions. Full cellwise comparisons of both rules with all baseline classes are reported in table 3.

Flat optima and stability of profile values. A separate numerical analysis of the finite-n objective finds a tangent-space Hessian eigenvalue ratio of order

$$4 \times 10^{-7},$$

and near-optimal mixtures can differ by as much as 0.46 in ℓ_1 distance while their profile values differ by less than

$$3 \times 10^{-4}$$

in relative terms. Thus substantially different weight vectors can yield nearly identical profile values. We therefore assess allocations through objective values and risk rather than weight stability alone.

Wasserstein-1 source-mixture allocation. We compare occupancy-based allocation with Wasserstein-1 matching over the same 23-source simplex. Let $\widehat{F}^{\mathrm{sel}}_{i,s}$ be the empirical radial CDF of source i on selection scramble s. Using the one-dimensional CDF representation of W_1 (Vallender, 1974) and $F_Q(r) = r^2$, we minimize

$$\widehat{W}_{\mathrm{sel}}(\alpha) = \frac{1}{4}\sum_{s=1}^{4}\int_0^1 \left|\sum_i \alpha_i \widehat{F}^{\mathrm{sel}}_{i,s}(r) - r^2\right| dr, \qquad \alpha \in \Delta_{23}.$$

The objective uses the same four selection scrambles and is minimized once by linear programming with trapezoidal quadrature on the 2001-point task grid. The returned weights are held fixed across all eight (h, n) settings and evaluated using the same target banks and disjoint evaluation scrambles.

The numerical optimizer assigns all weight to $(0.48, 0.45, 0.07)$. Its selection objective is 0.0690, compared with 0.1129 for uniform mixing and 0.0815 for the next-best source vertex, $(0.48, 0.42, 0.10)$. The solver reports a primal–dual gap of 1.1×10^{-15}; reevaluating the selected allocation on a 4001-point grid changes the objective by 8.4×10^{-8}. The selected source coincides with the scalar baseline, so the scalar column of table 3 also reports comparisons against this Wasserstein allocation. Both occupancy rules score below one in seven of the eight settings, supporting allocation objectives that target the local sample counts appearing in the NW risk analysis.

Table 3: Cell-level allocation summary on the evaluation grid: family-balanced risk-ratio score (allocation over baseline; below one favors the allocation) per cell, and the objective-specific discretized FW gap on the selection profiles ($\hat{g}_\Psi$ for the Ψ-mixture, $\hat{g}_\mathcal{J}$ for the finite-n allocation). The Ψ-mixture depends on h only, so its gap is per scale. The last row records the numerical-stability outcome. The scalar-selected source also minimizes the numerical Wasserstein-1 mixture objective (section C.3).

Cell (h, n)	uniform	best const	best Ψ	scalar	$\hat{g}$	Note
Ψ-mixture (fixed-scale rule; gap $\hat{g}_\Psi$)						
$(0.02, 50)$	0.966	0.993	1.012	1.028	5.8×10^{-7}	loses to best Ψ and scalar
$(0.02, 100)$	0.961	0.975	0.978	0.971	5.8×10^{-7}	
$(0.02, 200)$	0.961	0.965	0.965	0.945	5.8×10^{-7}	
$(0.02, 1000)$	0.958	0.972	0.972	0.959	5.8×10^{-7}	
$(0.05, 50)$	0.959	0.973	0.973	0.961	4.7×10^{-8}	
$(0.05, 100)$	0.953	0.958	0.970	0.958	4.7×10^{-8}	
$(0.05, 200)$	0.953	0.964	0.974	0.964	4.7×10^{-8}	
$(0.05, 1000)$	0.961	0.976	0.976	0.966	4.7×10^{-8}	
Finite-n allocation (finite-n rule; gap $\hat{g}_\mathcal{J}$)						
$(0.02, 50)$	0.969	0.997	1.016	1.032	2.7×10^{-8}	loses to best Ψ and scalar
$(0.02, 100)$	0.971	0.985	0.988	0.980	1.2×10^{-7}	
$(0.02, 200)$	0.971	0.975	0.975	0.955	5.3×10^{-7}	
$(0.02, 1000)$	0.959	0.974	0.974	0.961	3.9×10^{-7}	
$(0.05, 50)$	0.966	0.980	0.980	0.968	2.1×10^{-8}	
$(0.05, 100)$	0.967	0.972	0.984	0.972	4.7×10^{-7}	
$(0.05, 200)$	0.955	0.967	0.976	0.967	7.5×10^{-8}	
$(0.05, 1000)$	0.961	0.977	0.977	0.967	4.8×10^{-7}	discretized evaluation-objective regret bound 0.55% (criterion 0.5%)
Cells won (Ψ-mixture)	8/8	8/8	7/8	7/8	—	vs. finite-n: 8/8
Cells won (finite-n)	8/8	8/8	7/8	7/8	—	stability miss at $n = 1000$, $h = 0.05$
Numerical stability	original worst-cell bound 0.55% at $n = 1000$, $h = 0.05$ (tolerance 0.50%); tightened regret in $[6.02 \times 10^{-4}, 6.72 \times 10^{-4}]$ on the same discretized objective					

C.4 Pair Complementarity Study on the 23-Design Family

Question and selection rule. We ask whether two sources can provide spatially complementary coverage: can a mixture improve on the better single-source endpoint, and does that improvement persist across both allocation scales? The ranking rule, candidate set, and evaluation thresholds were fixed before the evaluation-scramble results were extracted.

For every pair $i < j$ among the 23 designs in table 5 and each

$$h \in \{0.02, 0.05\},$$

define on the four selection scrambles

$$\Psi_{ij}^{\mathrm{sel}}(\alpha) := \frac{1}{4} \sum_{s=1}^{4} \int \frac{dQ(z)}{\alpha m_{i,s}^{\mathrm{sel}}(z, h) + (1 - \alpha) m_{j,s}^{\mathrm{sel}}(z, h)}, \qquad \alpha \in [0, 1].$$

Let

$$\Psi_i := \Psi_{ij}^{\mathrm{sel}}(1), \qquad \Psi_j := \Psi_{ij}^{\mathrm{sel}}(0),$$

and let

$$\alpha_{ij}^{\star} \in \arg \min_{\alpha \in [0,1]} \Psi_{ij}^{\mathrm{sel}}(\alpha),$$

where the bounded scalar solve explicitly compares both endpoints.

The primary selection statistic is

$$G_{ij}^{\mathrm{best}} := \frac{\min(\Psi_i, \Psi_j) - \Psi_{ij}^{\mathrm{sel}}(\alpha_{ij}^{\star})}{\min(\Psi_i, \Psi_j)}.$$

Thus $G_{ij}^{\mathrm{best}} > 0$ measures the relative improvement of the pairwise Ψ-optimal mixture over the better single-source endpoint. As a secondary diagnostic motivated by corollary 7, define the normalized midpoint chord gap

$$J_{ij}^{\mathrm{mid}} := \frac{\frac{1}{2}(\Psi_i + \Psi_j) - \Psi_{ij}^{\mathrm{sel}}(\frac{1}{2})}{\frac{1}{2}(\Psi_i + \Psi_j)}.$$

We also mark a pair as a near-tie at a scale when

$$\frac{|\Psi_i - \Psi_j|}{\frac{1}{2}(\Psi_i + \Psi_j)} \leq 0.02.$$

To test whether complementarity also appears in the finite-budget objective, define

$$\widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}(\alpha) := \frac{1}{4}\sum_{s=1}^{4} \int r_n\big(\alpha m_{i,s}^{\mathrm{sel}}(z,h) + (1-\alpha) m_{j,s}^{\mathrm{sel}}(z,h)\big)\; dQ(z),$$

with

$$\alpha_{ij}^{\mathrm{fin}}(n,h) \in \arg\min_{\alpha\in[0,1]} \widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}(\alpha).$$

The corresponding relative gain over the better endpoint is

$$G_{ij}^{\mathrm{fin}}(n,h) := \frac{\min\{\widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}(0), \widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}(1)\} - \widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}\big(\alpha_{ij}^{\mathrm{fin}}(n,h)\big)}{\min\{\widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}(0), \widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{sel}}(1)\}},$$

for

$$n \in \{50, 100, 200, 1000\}.$$

The dependence of Ψ_i, Ψ_j, and $\alpha_{ij}^{\star}$ on h is suppressed; the (n,h) dependence of $\alpha_{ij}^{\mathrm{fin}}$ is shown explicitly.

Evaluation criteria. For each evaluation scramble $t \in \{1,\ldots,4\}$, define

$$\Psi_{ij}^{\mathrm{ev},t}(\alpha) := \int \frac{dQ(z)}{\alpha m_{i,t}^{\mathrm{ev}}(z,h) + (1-\alpha) m_{j,t}^{\mathrm{ev}}(z,h)},$$

and

$$\widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{ev},t}(\alpha) := \int r_n\big(\alpha m_{i,t}^{\mathrm{ev}}(z,h) + (1-\alpha) m_{j,t}^{\mathrm{ev}}(z,h)\big)\; dQ(z).$$

Evaluation gains $G_{ij}^{\mathrm{best,ev},t}$ and $G_{ij}^{\mathrm{fin,ev},t}(n,h)$ use the same better-endpoint normalizations as above, with $\alpha_{ij}^{\star}$ and $\alpha_{ij}^{\mathrm{fin}}(n,h)$ held fixed at their selection-derived values. No evaluation re-optimization is performed.

Pairs are ranked by

$$\min_h G_{ij}^{\mathrm{best}},$$

the worst-scale Ψ gain on the selection scrambles. The top five pairs are the evaluation candidates. A candidate meets the criteria only if

$$G_{ij}^{\mathrm{best,ev},t} \geq 3\%$$

for every evaluation scramble t at both scales and

$$G_{ij}^{\mathrm{fin,ev},t}(n,h) \geq 1\%$$

for every t, both scales, and $n \in \{100, 200\}$. We report the worst evaluation-scramble values,

$$\min_t G_{ij}^{\mathrm{best,ev},t}, \qquad \min_t G_{ij}^{\mathrm{fin,ev},t}(n,h).$$

The main-text witness is the highest-ranked pair meeting the criteria. Pairs are ranked on the selection scrambles, and the selected top five are evaluated unchanged on four disjoint evaluation scrambles to assess RQMC stability.

Evaluation results. Exactly two pairs meet the full two-stage criteria. The rank-1 candidate,

$$\theta_1 = (0.48, 0.33, 0.19), \qquad \theta_2 = (0.48, 0.45, 0.07),$$

is the witness used in fig. 4(c). The rank-2 pair,

$$(0.48, 0.30, 0.22) \quad \text{and} \quad (0.48, 0.45, 0.07),$$

also meets them; the remaining three top-five candidates do not.

For the rank-1 witness, the minimum evaluation-scramble improvement over the better endpoint is 3.6% at $h = 0.02$ and 3.1% at $h = 0.05$ for the selection-derived Ψ mixture. At $h = 0.02$, the separately optimized finite-n mixtures achieve minimum evaluation gains of 20.4% and 10.6% at $n = 100$ and $n = 200$, respectively. Full top-five results are given in table 4. For the displayed top five, the selection value G^{best} and the minimum evaluation value $\min_t G^{\mathrm{best},\mathrm{ev},t}$ agree to the precision shown in the table.

Across all 253 pairs the worst-scale, worst-evaluation Ψ gain $\min_h \min_t G_{ij}^{\mathrm{best},\mathrm{ev},t}$ has median 0.01%, with 90th and 99th percentiles of 1.3% and 2.8%. Sixty-two pairs are near-ties at both scales; none meets the full two-stage criteria. Both passing pairs include the short-third-link design $(0.48, 0.45, 0.07)$. At $h = 0.02$, the rank-1 witness is a numerical near-tie analogue of the equal-profile setting in corollary 7; at $h = 0.05$, the endpoint profiles are unequal, corresponding to Remark (a) after that corollary.

Table 4: Top-five candidates in the selection ranking. G^{best} and J^{mid} are computed on the selection scrambles; the two $\min_t$ columns are minima over the four evaluation scrambles, with the last column reported at $n = 100/200$. Gains are fractions, so 0.036 corresponds to 3.6%.

rank	pair	h	G^{best}	$\min_t G^{\mathrm{best},\mathrm{ev},t}$	J^{mid}	tie	$\min_t G^{\mathrm{fin},\mathrm{ev},t}$
1	$(0.48, 0.33, 0.19) + (0.48, 0.45, 0.07)$	0.02	0.036	0.036	0.043	yes	0.204 / 0.106
		0.05	0.031	0.031	0.042	no	0.029 / 0.023
2	$(0.48, 0.30, 0.22) + (0.48, 0.45, 0.07)$	0.02	0.035	0.035	0.050	no	0.212 / 0.103
		0.05	0.030	0.030	0.050	no	0.030 / 0.023
3	$(0.48, 0.36, 0.16) + (0.48, 0.45, 0.07)$	0.02	0.031	0.031	0.032	yes	0.119 / 0.098
		0.05	0.028	0.028	0.031	yes	0.024 / 0.020
4	$(0.48, 0.27, 0.25) + (0.48, 0.45, 0.07)$	0.02	0.032	0.032	0.053	no	0.193 / 0.094
		0.05	0.028	0.028	0.053	no	0.028 / 0.022
5	$(0.45, 0.33, 0.22) + (0.48, 0.45, 0.07)$	0.02	0.032	0.032	0.050	no	0.228 / 0.106
		0.05	0.028	0.028	0.048	no	0.027 / 0.016

Finite-n gain decomposition. For the rank-1 witness and each (n, h), on evaluation scramble t let $b_t(n, h) \in \{i, j\}$ be an endpoint attaining

$$\min\{\widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{ev},t}(0), \widehat{\mathcal{J}}_{ij,n,h}^{\mathrm{ev},t}(1)\},$$

that is, the better of the two endpoints. Define

$$m_{\mathrm{best}}^{(t)}(z, h) := m_{b_t(n,h),t}^{\mathrm{ev}}(z, h)$$

and

$$m_{\mathrm{mix}}^{(t)}(z, h) := \alpha_{ij}^{\mathrm{fin}}(n, h) m_{i,t}^{\mathrm{ev}}(z, h) + \big(1 - \alpha_{ij}^{\mathrm{fin}}(n, h)\big) m_{j,t}^{\mathrm{ev}}(z, h).$$

With (z, h) suppressed inside the integrals, the finite-n objective gain is

$$\Delta_{n,h}^{(t)} := \int \Big[r_n(m_{\mathrm{best}}^{(t)}) - r_n(m_{\mathrm{mix}}^{(t)}) \Big] \, dQ.$$

Positive $\Delta_{n,h}^{(t)}$ therefore means that the selection-derived mixture improves on the better endpoint.

This gain separates exactly into the empty-neighborhood contribution

$$\Delta_{\mathrm{empty}}^{(t)} := B^2 \int \Big[(1 - m_{\mathrm{best}}^{(t)})^n - (1 - m_{\mathrm{mix}}^{(t)})^n \Big] \, dQ$$

and the variance contribution

$$\Delta_{\mathrm{var}}^{(t)} := \sigma^2 \int \left[g_n(m_{\mathrm{best}}^{(t)}) - g_n(m_{\mathrm{mix}}^{(t)}) \right] dQ.$$

Let

$$\bar{\Delta}_{\mathrm{empty}} := \frac{1}{4} \sum_{t=1}^{4} \Delta_{\mathrm{empty}}^{(t)}, \qquad \bar{\Delta}_{\mathrm{var}} := \frac{1}{4} \sum_{t=1}^{4} \Delta_{\mathrm{var}}^{(t)}.$$

We report the fraction

$$\frac{\bar{\Delta}_{\mathrm{empty}}}{\bar{\Delta}_{\mathrm{empty}} + \bar{\Delta}_{\mathrm{var}}}.$$

At $h = 0.02$, the empty-neighborhood contribution accounts for 100%, 98%, and 73% of the gain at $n = 50, 100, 200$, respectively; the variance contribution dominates at $n = 1000$. At $h = 0.05$, the empty-neighborhood share is 95% at $n = 50$, and the variance contribution dominates from $n = 100$ onward. Thus the finite-n benefit changes mechanism with budget: avoiding empty neighborhoods dominates at small n, while variance reduction dominates once neighborhoods are more reliably populated.

Spatial concentration of the mixture gap. The nonnegative integrand in the exact mixture-gap identity eq. (11), evaluated on each selection scramble and then averaged, integrates to the unnormalized midpoint chord gap

$$\frac{1}{2}(\Psi_i + \Psi_j) - \Psi_{ij}^{\mathrm{sel}}\left(\frac{1}{2}\right).$$

For the witness this gap is 1.023 at $h = 0.02$ and 0.401 at $h = 0.05$; J_{ij}^{mid} is the same quantity normalized by $\frac{1}{2}(\Psi_i + \Psi_j)$.

At $h = 0.02$, the selection-average endpoint mass fields cross at

$$z = 0.228 \qquad \text{and} \qquad z = 0.744.$$

The 10% and 20% of deployment mass with the largest mixture-gap integrand values contribute 26% and 44% of the total gap, respectively. Thus the complementarity is spatially concentrated rather than uniformly distributed across task space.

Family-level allocation summary. Figure 5 summarizes the family-level allocation results of section C.3, separately from the pair-selection analysis. The figure reports the family-balanced score of the fixed-scale Ψ-mixture against uniform mixing and the three single-source baselines across all eight (h, n) cells.

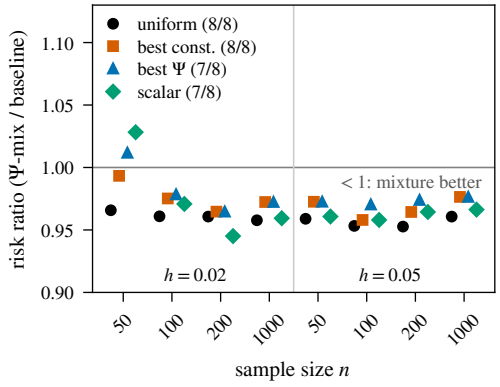

Figure 5: Allocation risk ratios across all eight (n, h) cells and all baseline classes: family-balanced score of the Ψ-mixture returned by the fixed-scale rule; below one favors the mixture; legend counts are cells won (section C.3).

C.5 Scale, Surrogate, and Computation

This subsection clarifies the scale choices used in the empirical studies, the relation between the scalar profile and the finite-n mass-field objective, and the cost of evaluating the resulting allocation problems. The analyses below use the same forward samples and mass-field estimates as the preceding experiments.

Scale selection. The common-support profile study uses

$$h \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.2\},$$

while the allocation experiments and pair study use

$$h \in \{0.02, 0.05\}.$$

For the seven-joint study, the candidate scales were

$$h \in \{0.02, 0.03, 0.04, 0.05, 0.08, 0.10\}\,\mathrm{m}.$$

The selected scale $h^\star = 0.10\,\mathrm{m}$ is the smallest candidate satisfying, on the preliminary mass-field estimates,

$$\min_\theta \min_{z \in \mathcal{G}_{\mathrm{Panda}}} \widehat{m}_\theta(z,h) \geq \tau_m, \qquad \tau_m = 10^{-4},$$

and

$$\mathrm{median}_{(\theta,z)}\left[n_{\mathrm{ref}} \widehat{m}_\theta(z,h)\right] \geq 3, \qquad n_{\mathrm{ref}} = 200,$$

where the median is taken jointly over all variant–grid-point pairs. The first condition avoids numerically vanishing local coverage on the evaluation grid; the second requires at least three expected nearby samples at the median variant–grid-point pair under the reference budget. The achieved median is 3.92 at $h^\star$. The acquisition-law study inherits this scale without retuning, while the remaining seven-joint scales are used only for profile sensitivity.

For comparison, the design-normalized 2R witness uses deterministic quadrature at fixed scales. On the 23-design family, rankings at $h = 0.02$ and $h = 0.05$ have Spearman correlation 0.995 for Ψ and 0.94 for the finite-n objective at $n = 200$.

Exact certificate, scalar surrogate, and finite-n objective. At fixed scale h, theorem 1(a) gives the certificate

$$\mathcal{C}_\theta(n,h) := \Lambda^2 h^{2\beta} + \sigma^2 \int g_n(m_\theta)\, dQ + B^2 \int (1-m_\theta)^n\, dQ,$$

so that

$$\mathcal{C}_\theta(n) = \inf_{0<h\leq \bar{h}} \mathcal{C}_\theta(n,h).$$

Both design-dependent terms retain the full mass field.

The profile is not this finite-n certificate. For $m \in (0,1]$,

$$g_n(m) \leq \frac{2}{nm}, \qquad (1-m)^n \leq \frac{1}{nm},$$

which gives the scalar upper surrogate

$$\mathcal{C}_\theta(n,h) \leq \Lambda^2 h^{2\beta} + \frac{B^2 + 2\sigma^2}{n}\, \Psi_\theta(h),$$

in which $\Psi_\theta(h)$ carries all source dependence. Thus, at fixed h, minimizing $\Psi_\theta(h)$ minimizes this scalar upper surrogate; it does not minimize the exact certificate in general.

For source allocation, the finite-n objective is

$$\mathcal{J}_{n,h}(\alpha) = \int \left[B^2(1-m_\alpha)^n + \sigma^2 g_n(m_\alpha)\right]\, dQ.$$

This objective retains the finite-budget empty-neighborhood and noise terms rather than replacing them by the scalar profile. Under homoscedastic noise it is the exact zero-fallback NW risk for a

constant target $f \equiv c$ with $|c| = B$. For general $|f| \le B$, it is the allocation-dependent target-free part of the certificate at fixed h; the Hölder bias allowance $\Lambda^2 h^{2\beta}$ does not depend on α.

The two objectives coincide only asymptotically after scaling. If $\sigma > 0$ and $\Psi(\alpha) < \infty$ at the fixed scale—in particular under the strict finite-scale common-support condition used for allocation—then

$$n g_n(m) \to \frac{1}{m}, \qquad n(1-m)^n \to 0,$$

with dominating envelopes $2/m$ and $1/m$. Dominated convergence therefore gives

$$\frac{n\,\mathcal{J}_{n,h}(\alpha)}{\sigma^2} \longrightarrow \Psi(\alpha) \qquad \text{for fixed } h.$$

Thus Ψ is the scaled large-n fixed-scale limit of $\mathcal{J}_{n,h}$, not a finite-n identity.

Finite-budget transition and allocation diagnostics. Across the 23-design grid, with each objective evaluated on each scramble and then averaged, Ψ and $\mathcal{J}_{n,h}$ have Spearman correlation at least 0.93 at every experimental budget. Their five lowest-scoring designs overlap in at least four positions. The two objectives therefore give similar aggregate rankings in this family, but they are not interchangeable at finite budgets.

The finite-budget difference is largest when empty neighborhoods are common. At small budgets the median design-wise ratio

$$\frac{\mathcal{J}_{n,h}}{\sigma^2 \Psi / n}$$

exceeds one: at $h = 0.02$, the empty-neighborhood component of $\mathcal{J}_{n,h}$ accounts for as much as 97%, 89%, and 68% for the most affected design at $n = 50, 100, 200$, respectively. Consistent with the fixed-scale limit above, the median ratio approaches one as the budget grows: it first comes within about 1% of one at the diagnostic budget $n = 5000$ for $h = 0.02$ and at $n = 1000$ for $h = 0.05$, remaining within that range at the larger diagnostic budget. The value $n = 5000$ is used only for this large-budget diagnostic, not as an experimental allocation budget.

We also evaluate how the Ψ-selected mixture performs under the finite-n objective. Using the evaluation-objective lower bound from section C.3,

$$\mathrm{Reg}^{\mathrm{ub}}_{n,h}(\alpha^{\Psi}_{\mathrm{sel}})$$

is a certified relative suboptimality bound for the fixed *discretized finite-n evaluation objective*. Its maximum over the eight allocation cells is 3.7%, and it is at most 0.5% at $n = 1000$. This quantity concerns allocation-objective suboptimality, not downstream decision regret.

To check sensitivity to the chosen scale, define the FW lower bound on the minimum of the discretized evaluation profile,

$$L^{\mathrm{ev}}_{\Psi,h} := \widehat{\Psi}^{\mathrm{ev}}_h(\alpha^{\Psi}_{\mathrm{ev},h}) - \hat{g}^{\mathrm{ev}}_{\Psi,h}(\alpha^{\Psi}_{\mathrm{ev},h}),$$

with $\hat{g}^{\mathrm{ev}}_{\Psi,h}(\alpha)$ the discretized FW gap of $\widehat{\Psi}^{\mathrm{ev}}_h$ at α; the bound is positive at both allocation scales. Define

$$\mathrm{Reg}^{\mathrm{cross}}_{h_s \to h_t} := \frac{\widehat{\Psi}^{\mathrm{ev}}_{h_t}(\alpha^{\Psi}_{\mathrm{sel},h_s}) - L^{\mathrm{ev}}_{\Psi,h_t}}{L^{\mathrm{ev}}_{\Psi,h_t}}.$$

The cross-scale values are 0.058% for $0.02 \to 0.05$ and 0.057% for $0.05 \to 0.02$. Thus the Ψ allocation selected at either scale is nearly optimal for the other scale on the discretized evaluation profiles.

The two allocation rules can nevertheless produce meaningfully different local coverage. Let

$$\bar{m}^{\mathrm{ev}}_{\alpha}(z,h) := \frac{1}{4}\sum_{t=1}^{4}\sum_{i} \alpha_i m^{\mathrm{ev}}_{i,t}(z,h)$$

and define

$$d_{\infty,\mathrm{rel}}(n,h) := \max_{z \in \mathcal{G}_{3\mathrm{R}}} \frac{\left|\bar{m}^{\mathrm{ev}}_{\alpha^{\mathcal{J}}_{\mathrm{sel}}}(z,h) - \bar{m}^{\mathrm{ev}}_{\alpha^{\Psi}_{\mathrm{sel}}}(z,h)\right|}{\bar{m}^{\mathrm{ev}}_{\alpha^{\Psi}_{\mathrm{sel}}}(z,h)}.$$

This distance ranges from 3% to 20% across the eight cells, showing that similar aggregate objective rankings need not imply nearly identical local mass fields.

The two allocation rules show different performance across evaluation objectives, consistent with this finite-budget distinction. The heterogeneous-target comparison favored the Ψ-mixture over the finite-n allocation in all eight cells (section C.3). For the rank-1 pair at $h = 0.02$, the finite-n allocation achieves a 20.4% gain in its constant-target objective at $n = 100$, while the Ψ-mixture achieves a 3.6% gain in the profile objective (section C.4).

Finally, Ψ is convex on the simplex at every fixed scale, while $\mathcal{J}_{n,h}$ is convex when

$$B^2 \geq \frac{3}{2}\sigma^2$$

by proposition 8; the numerical values used here satisfy this condition.

Computation and reuse. Forward samples are generated once for each source and scramble. Those forward samples are reused when evaluating multiple scales, but the ball-mass field is recomputed for each new scale and then cached. Once the mass fields at a fixed scale are cached, with S scrambles, K sources, and G quadrature points, evaluating a fixed-scale Ψ value or gradient costs

$$O(SKG),$$

while the recurrence-based evaluation of $\mathcal{J}_{n,h}$ and its gradient costs

$$O(nSKG).$$

Constructing the mass fields from raw forward samples additionally depends on the number of forward samples and on the method used to compute neighborhood counts, so we report the measured construction time below rather than assigning that stage a cached-array complexity. Changing mixture weights or target functions does not require another generator run.

Representative microbenchmarks are medians of five timed runs after one warm-up on an Apple M1 Max with 64 GB memory (macOS 15.7, NumPy 2.4, SciPy 1.17). Computing mass fields for 23 designs, 8 scrambles, and 2 scales takes 13 s with 2^{20} forward samples per design–scramble. One Ψ evaluation takes 0.3 ms and its gradient 0.4 ms; the 23-source FW-certified Ψ solve takes 12 ms per scale, and a two-source pair solve takes 0.9 ms. The 23-source finite-n solve takes 0.2–8.1 s per cell.

For comparison, full end-to-end runs include additional mass-field construction, diagnostics, and output generation and are therefore a different timing category: the finite-n allocation run takes 1822 s, the 253-pair study 389 s, and the acquisition-law mass-field computation 381 s for five laws, eight seeds, four scales, and $16{,}384$ test points.

C.6 Experimental Objects and Numerical Settings

This subsection collects the experimental objects and numerical settings shared across sections C.1 to C.4 and D. Protocol-specific choices and outcomes are described in their respective subsections; the tables here provide the common design families, target banks, and seven-joint parameter variations.

Seven-joint numerical grids. Coverage is evaluated on the fixed 61×61 grid $\mathcal{G}_{\mathrm{Panda}}$ over the common deployment rectangle. Regression risks are estimated on a $16{,}384$-point Sobol sample from Q (seed 7070), and KRR calibration uses a separate 4096-point Sobol validation set (seed 6060). The decision study uses a fixed candidate set of 256 Sobol points in the common deployment region (seed 9090). For this candidate set, normalized decision regret is

$$\frac{f(\hat{z}) - f(z^\star)}{\max_{\mathcal{Z}_{\mathrm{cand}}} f - \min_{\mathcal{Z}_{\mathrm{cand}}} f},$$

as in lemma 11; it is reported only when the candidate range in the denominator exceeds the threshold 0.05.

The cosine-bump class membership follows analytically from the construction in table 6. For the random-smooth banks, the stated Lipschitz and sup-norm bounds are verified on the fixed 4001-point numerical grid used for evaluation.

Table 5: Planar three-revolute-joint (3R) design family used in the common-support experiments. The 23 link-length triples satisfy $\sum_j L_j = 1$ and are obtained from $L_1 \in \{0.36, 0.39, 0.42, 0.45, 0.48\}$ and $L_2 \in \{0.27, 0.30, 0.33, 0.36, 0.39, 0.42, 0.45\}$ by setting $L_3 = 1 - L_1 - L_2$ and retaining $L_1 \geq L_2 \geq L_3 \geq 0.05$. The displayed order is the design indexing used in the allocation and pair-complementarity analyses.

(0.36,0.33,0.31)	(0.36,0.36,0.28)	(0.39,0.33,0.28)	(0.39,0.36,0.25)
(0.39,0.39,0.22)	(0.42,0.30,0.28)	(0.42,0.33,0.25)	(0.42,0.36,0.22)
(0.42,0.39,0.19)	(0.42,0.42,0.16)	(0.45,0.30,0.25)	(0.45,0.33,0.22)
(0.45,0.36,0.19)	(0.45,0.39,0.16)	(0.45,0.42,0.13)	(0.45,0.45,0.10)
(0.48,0.27,0.25)	(0.48,0.30,0.22)	(0.48,0.33,0.19)	(0.48,0.36,0.16)
(0.48,0.39,0.13)	(0.48,0.42,0.10)	(0.48,0.45,0.07)	

Table 6: Target banks for the 3R experiments on the radial coordinate $r \in [0, 1]$, where Q has density $2r$. The cosine-bump construction uses $\beta = 1$, $\Lambda = 1$, and $B = 1$ and satisfies the stated Hölder constraint analytically. Random-smooth targets are rescaled using the 4001-point numerical verification grid described below.

Target bank	Construction
Cosine bump (u, s)	$f(r) = a_{\mathrm{bump}} \cos^2\!\big(\frac{\pi(r-c)}{2s}\big)$ for $\lvert r - c\rvert < s$ and 0 otherwise, with $c = \sqrt{u}$ (the radial location of deployment quantile u), width s, and $a_{\mathrm{bump}} = \min\{B, 2\Lambda s/\pi\}$. The construction is C^1 with Lipschitz constant $\pi a_{\mathrm{bump}}/(2s) \leq \Lambda$.
Target-reversal bump bank (section C.1)	$u \in \{0.1, 0.2, \ldots, 0.9\}$, $s \in \{0.05, 0.10\}$ (18 targets).
Allocation bump bank (section C.3)	$u \in \{0.05, 0.15, \ldots, 0.95\}$, $s \in \{0.05, 0.10\}$ (20 targets). The center locations alternate with those of the reversal bank; the $u = 0.95$ bump is truncated at $r = 1$.
Random smooth target (seed ς)	$\widetilde{f}(r) = \sum_{k=1}^{40} g_k\, k^{-(\beta+1)} \cos(k\pi r)$, $g_k \sim \mathcal{N}(0, 1)$, generated with NumPy `default_rng(ς)` and rescaled by the smaller factor required by the 4001-point numerical checks of the Lipschitz seminorm and maximum absolute value, with thresholds 0.9Λ and B, respectively.
Certificate-tracking smooth bank (section C.1)	Seeds $\varsigma \in \{1000, \ldots, 1029\}$ (30 targets).
Allocation smooth bank (section C.3)	Seeds $\varsigma \in \{2000, \ldots, 2029\}$ (30 targets), disjoint from the certificate-tracking seeds.

Seven-joint kinematic specification. The seven-joint family is a synthetic Panda-like arm family rather than an exact manufacturer model. Each variant is specified by a modified Denavit–Hartenberg chain in the Craig convention,

$$T_i(q_i) = R_x(\alpha_{i-1})\, T_x(a_{i-1})\, R_z(q_i + \theta_i)\, T_z(d_i), \qquad i = 1, \ldots, 7,$$

with all fixed joint offsets $\theta_i = 0$, followed by a fixed flange translation

$$d_{\mathrm{flange}} = 0.107\,\mathrm{m}$$

along the local z axis and then the variant-specific tool offset d_{tool}.

Table 7(a) lists the nominal chain and the joint limits shared by all variants. In the strict Craig indexing of part (a), the row-4 elbow offset is a_3; elsewhere in the paper and in the fixed variant specification we retain the row-wise label a_4 for this same parameter. Table 7(b) lists the four varied length parameters using that row-wise label, with each nonnominal variant changing at most two of them. The nonzero twist angles use the specified numerical values ± 1.5707963, which differ from $\pm\pi/2$ by approximately 3×10^{-8}. The source law is uniform on the joint-limit box.

All eight forward maps agree with an independently constructed kinematics implementation to a maximum tool-position discrepancy of $4.5 \times 10^{-16}\,\mathrm{m}$ over 10^4 in-limit configurations, all 128 joint-limit corners, and the home pose. This check verifies numerical agreement between the two implementations; the geometric regularity conditions used in the theory are established separately in section B.3.2.

Table 7: Synthetic Panda-like seven-joint kinematic specification. Lengths are in metres and angles in radians. (a) Nominal modified-DH parameters in strict Craig indexing (row i carries α_{i-1}, a_{i-1}, and d_i) and the shared joint limits $[q_i^-, q_i^+]$. (b) The four varied length parameters using the row-wise parameter labels retained elsewhere in the paper; every parameter not listed in (b) takes its nominal value from (a).

(a) Nominal chain and joint limits.

i	α_{i-1}	a_{i-1}	d_i	q_i^-	q_i^+
1	0	0	0.333	−2.8973	2.8973
2	−1.5707963	0	0	−1.7628	1.7628
3	1.5707963	0	0.316	−2.8973	2.8973
4	1.5707963	0.0825	0	−3.0718	−0.0698
5	−1.5707963	−0.0825	0.384	−2.8973	2.8973
6	1.5707963	0	0	−0.0175	3.7525
7	1.5707963	0.088	0	−2.8973	2.8973

(b) Controlled variant parameters.

variant	d_3	d_5	a_4	d_{tool}
V0 (nominal)	0.316	0.384	0.0825	0
V1 (upper arm +5%)	0.3318	0.384	0.0825	0
V2 (upper arm −5%)	0.3002	0.384	0.0825	0
V3 (forearm +5%)	0.316	0.4032	0.0825	0
V4 (forearm −5%)	0.316	0.3648	0.0825	0
V5 (tool +20 mm)	0.316	0.384	0.0825	0.020
V6 (elbow +15 mm)	0.316	0.384	0.0975	0
V7 (combined short)	0.3002	0.3648	0.0825	0

Bootstrap procedures. For the primary regression-risk study in section D, NW uses 100 training replicates and KRR uses 30. Reported 95% intervals are percentile intervals from 2000 bootstrap draws, using the 2.5th and 97.5th percentiles, generated from a single random-number stream seeded with 777 and shared across all estimator–field–budget cells in the order evaluated. Within each estimator–field–budget block, replicate indices are sampled with replacement and the same index set is used across variants or allocation arms, preserving the paired comparison.

For each bootstrap draw, the Spearman statistic is recomputed between $\Psi(h^\star)$ and the resampled replicate-mean risks of the eight variants. The V1–V2 and uniform–V2 relative-risk contrasts are recomputed from the same paired samples using the definitions in table 10. The V0–V2 comparison is reported as an additional contrast from the fixed risk arrays.

The expanded-grid KRR comparison in section D uses the same paired-resampling procedure with a separately initialized random-number stream seeded with 777.

The decision study uses 200 replicates and a separate bootstrap stream with seed 778. Paired replicate differences are resampled with replacement, and percentile intervals are reported for both the mean and median difference in normalized regret between uniform mixing and the Ψ-selected variant.

For the acquisition-law study in section D.1, the sixteen paired seeds are resampled with replacement using one common index set across arms and cells in each bootstrap draw. We use 2000 draws from a stream with seed 2026090207. Each criterion is evaluated using the one-sided 95% upper bootstrap quantile of the maximum over cells of the aggregate log-risk ratio.

D Synthetic Seven-Joint Sensitivity Studies

This appendix provides the sensitivity evidence underlying section 6, using synthetic observations generated by seven-joint forward maps. The first study varies the forward map at fixed acquisition, and the second varies the acquisition law at a fixed forward map. The kernel ridge regression (KRR) comparisons assess whether the pre-training coverage profile retains a useful regression-risk signal with a different learner. Figure 6 summarizes the experimental object: a seven-joint source

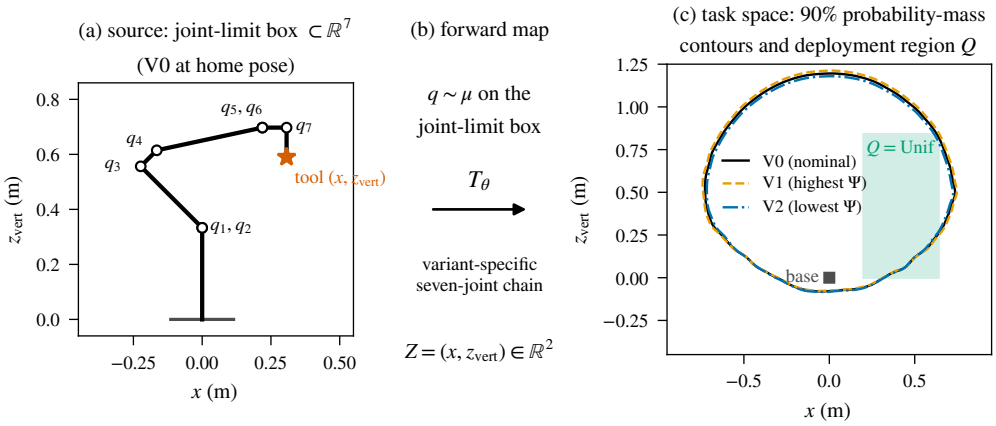

Figure 6: Seven-joint generator and task-space setup. (a) The nominal variant V0 at its home pose; configurations q are sampled uniformly from the shared joint-limit box. (b) Each variant defines a forward map T_θ from joint space to the sagittal tool-point coordinate $Z = (x, z_{\text{vert}}) \in \mathbb{R}^2$. (c) 90% task-space probability-mass contours for V0 (nominal), V1 (highest $\Psi(h^\star)$), and V2 (lowest $\Psi(h^\star)$) under the joint-uniform acquisition law, with the common deployment region Q. The contours nearly coincide; V1 and V2 differ from V0 only by a $\pm 5\%$ change in one link length (table 7), while the study focuses on finite-scale local mass over Q at $h^\star$.

space, a variant-specific forward map, and a two-dimensional sagittal task space with the common deployment region.

Seven-joint family and deployment region. We use the 8 synthetic arm-like variants in table 7 (illustrated in fig. 6), with the fixed modified Denavit–Hartenberg parameters listed there. The source law is uniform on the joint-limit box, and the forward map is the sagittal projection of the tool point; section B.3.2 verifies the coarea assumptions for all variants, and the implementation cross-check of all eight forward maps against an independent kinematic model is reported in section C.6.

Using coverage alone, we set the common deployment rectangle

$$x \in [x_{\min}, x_{\max}] = [0.20, 0.64]\,\text{m}, \qquad z_{\text{vert}} \in [z_{\text{vert,min}}, z_{\text{vert,max}}] = [0.00, 0.84]\,\text{m}.$$

Here $Z = (x, z_{\text{vert}})$ denotes the tool-point coordinate, with z_{vert} used to distinguish the vertical component from a generic task point z. The deployment law Q is uniform on this rectangle. On the numerical coverage grids, every variant has positive estimated mass throughout the rectangle at $h^\star = 0.10\,\text{m}$.

The same coverage calculation identifies V2 and V1 as the lower- and higher-$\Psi(h^\star)$ comparison variants. On the subsequent risk-run mass fields, $\Psi(h^\star)$ ranges from 64.3 for V2 to 67.2 for V1 across the eight variants, a relative spread of about 4.6%. All variants use the same target and conditional noise laws, so the statistical difference across variants is covariate shift only.

Coverage and pre-training allocation. For the numerical mass fields at $h^\star$, the Ψ-optimal mixture over the 8 variants is the V2 vertex. This does not follow from pointwise dominance: V7 has higher local mass than V2 on 50.7% of the deployment region. Nevertheless,

$$\Gamma_{\text{V2}\to\text{V7}} = +0.72 > 0,$$

so moving a small amount of allocation from V2 toward V7 would increase Ψ to first order. The least adverse inactive direction is V6, with

$$\Gamma_{\text{V2}\to\text{V6}} = +0.041 > 0.$$

Thus the mass fields cross, but none of the inactive directions is locally actionable at the V2 vertex, illustrating the distinction in proposition 2(b) between pointwise crossing and an improving allocation direction.

Table 8: Synthetic target fields used in the seven-joint studies. Task coordinates are (x, z_{vert}), with $u = (x - x_{\min})/(x_{\max} - x_{\min})$ and $v = (z_{\mathrm{vert}} - z_{\mathrm{vert,min}})/(z_{\mathrm{vert,max}} - z_{\mathrm{vert,min}})$. For each raw field $\widetilde{f}_j$, the evaluated field is $f_j = (\widetilde{f}_j - \widetilde{f}_j^{\min})/(\widetilde{f}_j^{\max} - \widetilde{f}_j^{\min})$, with the minimum and maximum taken on the fixed 61×61 numerical grid $\mathcal{G}_{\mathrm{Panda}}$. Gaussian label noise has $\sigma = 0.05$. F0 is used only for learner calibration.

field	raw construction $\widetilde{f}_j(u, v)$
F0 (calibration)	$(u-0.5)^2 + (v-0.5)^2 + 0.3\exp\big(-\frac{(u-0.3)^2+(v-0.7)^2}{2\cdot 0.15^2}\big)$
F1 (goal plus obstacles)	$\lambda_g\big((u-u_g)^2+(v-v_g)^2\big)+\sum_{r=1}^{3}\lambda_r\exp\big(-\frac{(u-c_{ru})^2+(v-c_{rv})^2}{2s_r^2}\big)$; $(u_g, v_g, \lambda_g) = (0.8, 0.35, 1.0)$; $(c_{ru}, c_{rv}, s_r, \lambda_r)$: $(0.4, 0.55, 0.10, 0.6)$, $(0.6, 0.20, 0.08, 0.5)$, $(0.25, 0.80, 0.12, 0.4)$
F2 (multimodal inspection)	$-\sum_{k=1}^{4}\omega_k\exp\big(-\frac{(u-u_k)^2+(v-v_k)^2}{2t_k^2}\big)$; $(u_k, v_k, t_k, \omega_k)$: $(0.2, 0.30, 0.10, 1.0)$, $(0.7, 0.75, 0.12, 0.9)$, $(0.5, 0.45, 0.07, 0.7)$, $(0.85, 0.15, 0.09, 0.8)$
F3 (random smooth)	$\sum_{\ell=1}^{12}\omega_\ell\exp\big(-\frac{(u-u_\ell)^2+(v-v_\ell)^2}{2s_\ell^2}\big)$; $(u_\ell, v_\ell) \sim \mathrm{Unif}[0,1]^2$, $s_\ell \sim \mathrm{Unif}[0.06, 0.18]$, $\omega_\ell \sim \mathcal{N}(0,1)$, drawn once with NumPy `default_rng(3000)` in the order centers, widths, weights

Target fields and learner calibration. F1–F3 are fixed synthetic target fields. The normalization in table 8 is a numerical normalization on $\mathcal{G}_{\mathrm{Panda}}$; no continuous-domain function-class statement is needed for this sensitivity study. For the primary comparisons, KRR hyperparameters are selected on the separate F0 calibration field and held fixed across F1–F3. NW on the same training distributions is retained as a reference estimator.

Each of the six field–budget settings,

$$F \in \{\mathrm{F1}, \mathrm{F2}, \mathrm{F3}\}, \qquad n \in \{200, 800\},$$

uses 30 KRR training replicates and 100 NW replicates. Risk is evaluated on the common deployment rectangle using the fixed numerical evaluation rule described in section C.6.

Regression-risk screening. Across the eight variants, the Spearman association between $\Psi(h^\star)$ and mean KRR risk is positive in all 6/6 settings, with 4/6 paired-bootstrap intervals excluding zero; all three settings at $n = 800$ exclude zero. The V2–V1 comparison also agrees with lower mean KRR risk for V2 in 5/6 settings. At $n = 800$, the positive rank association remains under every leave-one-variant-out recalculation.

The NW reference gives the same qualitative aggregate pattern: all six rank correlations are positive and four intervals exclude zero. For KRR, V2 has lower mean risk than uniform mixing in 5/6 settings, although only the F3, $n = 800$ paired interval excludes zero. Relative to the nominal variant V0, the V2 paired interval excludes zero in 3/6 settings (table 10).

For F2 at $n = 200$, the V1–V2 mean-risk ordering reverses the profile ordering, although the across-variant Spearman correlation remains positive and the paired interval for the V1–V2 contrast includes zero. The aggregate association supports $\Psi(h^\star)$ as a training-free screening signal for regression risk in this synthetic family. Figure 7 visualizes the six score–risk comparisons, and table 10 reports the corresponding contrasts.

Downstream argmin decision. For each field, the KRR regressor chooses an element of a fixed candidate set of 256 low-discrepancy task points by argmin. Normalized regret is defined using the candidate-set range as in lemma 11; we report normalized regret when the candidate-set range exceeds 0.05, and the observed ranges are 0.95–0.99.

At $n = 200$, over 200 replicates, the median normalized regret difference

$$\text{uniform mixing} - \Psi\text{-optimal V2}$$

is -0.007 for F1, 0.000 for F2, and $+0.019$ for F3. Thus the point estimates favor uniform mixing for F1, neither arm for F2, and V2 for F3; every paired-bootstrap interval crosses zero. Downstream decision effects are therefore field-dependent and unresolved; within this stress test, the scalar coverage score is more consistently associated with regression risk than with downstream argmin regret.

Table 9: KRR calibration sensitivity. Primary and expanded-grid configurations are $(b, \lambda) = (0.10\,\mathrm{m}, 10^{-4})$ and $(0.15\,\mathrm{m}, 10^{-5})$, respectively. ρ_s correlates $\Psi(h^\star)$ with replicate-mean risk across the eight variants. Brackets give 95% percentile intervals from 2000 paired bootstrap resamples of 30 training replicates. The final column reports the expanded-grid mean-risk difference $R_{\mathrm{unif}} - R_{\mathrm{V2}}$ in units of 10^{-3}; positive values favor V2. Both configurations use the same training samples and evaluation points.

Setting	Primary ρ_s	Expanded ρ_s	Expanded risk difference
F1, $n = 200$	$0.214\,[-0.262, 0.738]$	$0.762\,[-0.476, 0.858]$	$0.46\,[-2.3, 3.2]$
F1, $n = 800$	$0.714\,[0.262, 0.905]$	$0.952\,[0.214, 0.929]$	$0.040\,[-0.059, 0.15]$
F2, $n = 200$	$0.095\,[-0.167, 0.571]$	$0.048\,[-0.429, 0.429]$	$-5.0\,[-12, 0.59]$
F2, $n = 800$	$0.786\,[0.048, 0.952]$	$0.738\,[0.095, 0.929]$	$0.022\,[-0.21, 0.24]$
F3, $n = 200$	$0.714\,[0.262, 0.952]$	$0.357\,[-0.620, 0.929]$	$-0.46\,[-4.6, 2.8]$
F3, $n = 800$	$0.738\,[0.309, 0.834]$	$0.786\,[0.405, 0.952]$	$0.044\,[0.0093, 0.076]$

A descriptive analysis of the fixed candidate set gives decision margins of 0.0013, 0.0336, and 0.0213 for F1, F2, and F3, respectively. F1 has the smallest margin, with 9 candidates lying within 1% of the optimum. Small candidate-local prediction differences can therefore change the selected argmin even when aggregate regression risks are similar. The same analysis finds that V2 has 3–4% more local mass than the uniform mixture at the candidate-set optimum of each field, yet the argmin outcomes remain mixed. This is consistent with lemma 11: decision quality depends jointly on candidate-local error and the margin, not on aggregate coverage or regression risk alone.

KRR calibration and sensitivity. Gaussian KRR uses

$$k_b(z, z') = \exp\left(-\frac{\|z - z'\|^2}{2b^2}\right)$$

and solves

$$[\mathbf{K}_b + n\lambda I + 10^{-10} I]c = y$$

in float64 by Cholesky. Calibration minimizes validation mean-squared error on the separate F0 field at $n = 800$, averaged over ten training replicates and all eight variants, using the fixed 4096-point Sobol validation set with seed 6060.

The primary grid is

$$b \in \{0.03, 0.05, 0.08, 0.10\}\,\mathrm{m}, \qquad \lambda \in \{10^{-4}, 10^{-3}, 10^{-2}\}.$$

It selects $(b, \lambda) = (0.10\,\mathrm{m}, 10^{-4})$, used in the primary comparisons and in the acquisition-law study (table 12).

We assess calibration sensitivity on the expanded grid

$$b \in \{0.03, 0.05, 0.08, 0.10, 0.12, 0.15, 0.20\}\,\mathrm{m}, \qquad \lambda \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\},$$

using the same F0 data and selection criterion. The selected interior grid point, $(0.15\,\mathrm{m}, 10^{-5})$, reduces F0 validation MSE from 7.540×10^{-4} to 6.318×10^{-4}. With $h^\star = 0.10\,\mathrm{m}$ and all acquisition distributions unchanged, we evaluate this configuration on F1–F3 using the same 30 paired training replicates and deployment evaluation points.

Score–risk correlations are positive in all six settings under both configurations. V2 has lower mean risk than V1 in five of six settings under each configuration. At $n = 800$, all three paired-bootstrap 95% intervals for the score–risk correlations lie above zero under both configurations (table 9). The same coverage score thus retains its high-budget risk-screening value after recalibrating bandwidth and regularization.

D.1 Acquisition-Law Study on a Fixed Forward Map

Study design. This study holds the nominal seven-joint forward map, joint-space support, deployment law Q, target fields F1–F3, conditional noise law, and label budgets

$$n \in \{200, 800\}$$

Table 10: Primary KRR and argmin-decision results. ρ_s is the Spearman correlation between $\Psi(h^\star)$ and mean KRR risk across the eight variants, with paired-bootstrap 95% interval. The three contrast columns are relative risk differences against V2, in percent: V1 vs. V2 is $(R_{\mathrm{V1}} - R_{\mathrm{V2}})/R_{\mathrm{V1}}$, with the uniform and V0 columns defined analogously; positive values favor V2. The V1–V2 and uniform–V2 contrasts are the primary contrasts; V0–V2 is an additional contrast from the same fixed risk arrays. The decision column reports the median normalized-regret difference, uniform minus V2, at $n = 200$; positive values favor V2. In the aggregate row, the V1 and uniform entries count settings whose point estimates favor V2, whereas the V0 entry counts settings whose paired interval excludes zero.

Setting	ρ_s (95% CI)	V1 vs. V2	Unif. vs. V2	V0 vs. V2	Decision ($n = 200$)
F1, $n = 200$	$0.21\ [-0.26, 0.74]$	$+5.1\ [-1.9, +11.5]$	$+2.0\ [-4.8, +8.1]$	$+2.7\ [-1.3, +6.4]$	$-0.007\ [-0.011, 0.001]$
F1, $n = 800$	$0.71\ [0.26, 0.90]$	$+9.1\ [+5.0, +12.8]$	$+3.4\ [-1.1, +8.2]$	$+4.6\ [+1.9, +7.2]$	—
F2, $n = 200$	$0.10\ [-0.17, 0.57]$	$-3.2\ [-11.7, +3.7]$	$-5.0\ [-14.1, +3.9]$	$-1.0\ [-5.6, +3.0]$	$0.000\ [-0.034, 0.034]$
F2, $n = 800$	$0.79\ [0.05, 0.95]$	$+2.2\ [-2.7, +6.6]$	$+0.7\ [-3.0, +4.4]$	$+1.2\ [-0.9, +3.3]$	—
F3, $n = 200$	$0.71\ [0.26, 0.95]$	$+6.5\ [+1.7, +11.5]$	$+0.8\ [-8.4, +9.6]$	$+4.8\ [+2.0, +7.3]$	$+0.019\ [-0.009, 0.032]$
F3, $n = 800$	$0.74\ [0.31, 0.83]$	$+7.6\ [+4.9, +10.8]$	$+4.6\ [+2.3, +6.9]$	$+3.2\ [+1.6, +4.9]$	—
Aggregate	6/6 positive; 4/6 CI exclude 0	5/6 pt. est.	5/6 pt. est.	3/6 CI excl. 0	mixed

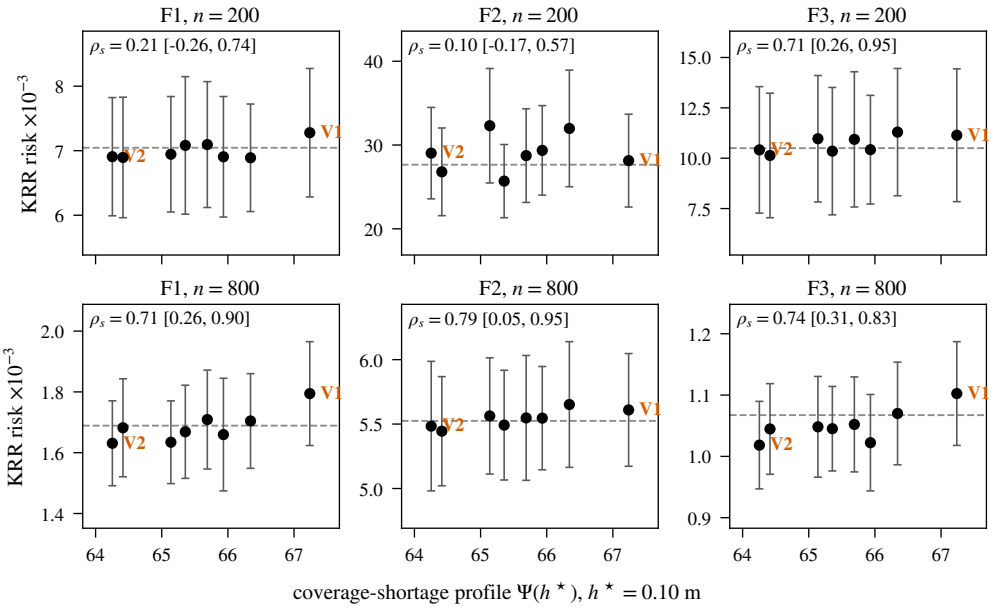

Figure 7: Coverage profile versus KRR risk in the synthetic seven-joint sensitivity study. Each panel contains the eight variants for one target field and label budget. Points show mean KRR risk with ± 2 standard errors over 30 replicates; the dashed line is the uniform-mixture risk, and V2 and V1 mark the lower- and higher-Ψ variants. Each panel reports Spearman's ρ_s with a paired-bootstrap 95% interval. The point estimate is positive in all six settings, and the interval excludes zero in four, including all three at $n = 800$.

fixed, and varies only the static i.i.d. acquisition law on the joint box. It therefore isolates acquisition-law effects without changing the forward map or label-generating process.

Each of the five source laws in table 11 has the form

$$\mu_i = 0.1\,\mu_{\mathrm{unif}} + 0.9\,\widetilde{\mu}_i.$$

Writing

$$P_{\text{joint-uniform}} := (T_\theta)_{\#}\mu_{\mathrm{unif}}, \qquad \widetilde{P}_i := (T_\theta)_{\#}\widetilde{\mu}_i,$$

linearity of the pushforward gives

$$P_i = 0.1\,P_{\text{joint-uniform}} + 0.9\,\widetilde{P}_i.$$

Hence every P_i has the same support as $P_{\text{joint-uniform}}$ and

$$m_i(z,h) \geq 0.1\, m_{\text{joint-uniform}}(z,h).$$

The comparison therefore concerns finite-scale source allocation rather than support loss.

Before training, we compute the following allocation arms from the mass fields at $h^\star$: equal mixing; the best single source by Ψ; Ψ-ranking, with

$$\Psi_i(h^\star) = \int \frac{dQ(z)}{m_i(z,h^\star)}, \qquad \alpha_i \propto \frac{1}{\Psi_i(h^\star)};$$

the Ψ-optimal mixture of proposition 2; and two label-free global-matching baselines. An ideal control samples $Z \sim Q$ directly.

Numerical allocation weights. In the source order of table 11 (joint-uniform, goal-biased, joint-limit-biased, trajectory-local, outer-reach), the Ψ-optimal weights computed before training are approximately

$$(0.012558,\ 0.228105,\ 0.746060,\ 0.013277,\ 0),$$

and the Ψ-ranking weights are

$$(0.328107,\ 0.143334,\ 0.327700,\ 0.073165,\ 0.127694).$$

These weights apply to the complete acquisition laws of table 11, each of which contains its own uniform component.

The moment-matching baseline minimizes

$$\left\| \mathbf{D}_{\text{rect}}^{-1} \left(\sum_i \alpha_i \bar{z}_i - \bar{z}_Q \right) \right\|_2^2 + \left\| \mathbf{D}_{\text{rect}}^{-1} \left(\sum_i \alpha_i \Xi_i - \Xi_Q \right) \mathbf{D}_{\text{rect}}^{-1} \right\|_F^2$$

over the simplex, where

$$\bar{z}_i = \mathbb{E}_{P_i}[Z], \qquad \Xi_i = \mathbb{E}_{P_i}[ZZ^\top],$$

and $\bar{z}_Q, \Xi_Q$ are the corresponding moments of the uniform deployment law. Here

$$\mathbf{D}_{\text{rect}} = \text{diag}(x_{\max} - x_{\min}, z_{\text{vert,max}} - z_{\text{vert,min}}).$$

The histogram baseline minimizes

$$\left\| \sum_i \alpha_i \nu_i - \nu_Q \right\|_2^2,$$

where ν_i is the source probability-mass vector on a fixed 30×60 equal-area grid over the deployment rectangle plus one outside cell, and ν_Q is uniform on the rectangle cells and zero outside. Both baselines are label-free, use only the selection-seed geometry samples, and are solved by SLSQP on the simplex.

For the profile-equated pair study, define

$$P_{\text{unif}} := P_{\text{joint-uniform}}, \qquad P_{\text{peq}} := \lambda^\star_{\text{pair}} P_{\text{joint-limit}} + (1 - \lambda^\star_{\text{pair}}) P_{\text{goal}},$$

and

$$P_{\text{mid}} := \tfrac{1}{2}(P_{\text{unif}} + P_{\text{peq}}).$$

The two component laws in P_{peq} are the full induced laws from the corresponding rows of table 11, including their $0.1 P_{\text{joint-uniform}}$ components. Thus P_{peq} is itself a realizable source mixture.

Numerical profile construction. For $X \in \{\text{unif}, \text{peq}, \text{mid}\}$, let $\widehat{m}^{\text{sel}}_{X,s}(z,h^\star)$ and $\widehat{m}^{\text{ev}}_{X,t}(z,h^\star)$ denote the mass-field estimates of P_X on four selection scrambles $s \in \{1,\dots,4\}$ and four disjoint evaluation scrambles $t \in \{1,\dots,4\}$. Define

$$\widehat{\Psi}^{\text{sel}}_{X,s}(h^\star) := \int \frac{dQ(z)}{\widehat{m}^{\text{sel}}_{X,s}(z,h^\star)}, \qquad \widehat{\Psi}^{\text{ev}}_{X,t}(h^\star) := \int \frac{dQ(z)}{\widehat{m}^{\text{ev}}_{X,t}(z,h^\star)},$$

Table 11: Acquisition laws used in the fixed-forward-map study. Each law is $\mu_i = 0.1\,\mu_{\mathrm{unif}} + 0.9\,\widetilde{\mu}_i$ on the joint box. The value $\rho_{\mathrm{ref}} = 0.8577\,\mathrm{m}$ used by the outer-reach law is a numerical reference obtained from the Sobol construction.

law	$\widetilde{\mu}_i$	sampler
joint-uniform	uniform on the joint box	direct
goal-biased	$\frac{d\tilde{\mu}}{d\mu_{\mathrm{unif}}}(q) \propto \exp\Big(-\frac{\|T_\theta(q)-z_g\|^2}{2(0.11\,\mathrm{m})^2}\Big)$, $z_g = (0.42, 0.42)\,\mathrm{m}$	exact rejection
joint-limit-biased	each joint independently $\mathrm{Beta}(0.5, 0.5)$ rescaled to its range	direct
trajectory-local	uniform four-component mixture over centers drawn once i.i.d. from μ_{unif} with NumPy `default_rng(8001)`; conditional on a center, each joint is an independent truncated normal with s.d. 5% of its joint range	direct (inverse CDF)
outer-reach	$\frac{d\tilde{\mu}}{d\mu_{\mathrm{unif}}}(q) \propto \exp\Big(-\frac{\max(\rho_{\mathrm{ref}}-\rho(T_\theta(q)),0)}{0.11\,\mathrm{m}}\Big)$, $\rho = \sqrt{x^2 + (z_{\mathrm{vert}} - 0.333\,\mathrm{m})^2}$ ($0.333\,\mathrm{m}$ is the base-link height d_1 of table 7)	exact rejection

using the fixed numerical quadrature, and

$$\widehat{\Psi}_{X,\mathrm{sel}}(h^\star) := \frac{1}{4}\sum_{s=1}^{4} \widehat{\Psi}^{\mathrm{sel}}_{X,s}(h^\star).$$

The root rule selects the smallest solution

$$\lambda^\star_{\mathrm{pair}} = 0.4017$$

for which

$$\widehat{\Psi}_{\mathrm{unif,sel}}(h^\star) = \widehat{\Psi}_{\mathrm{peq,sel}}(h^\star).$$

The evaluation scrambles are then used only to assess numerical stability. Their maximum relative endpoint discrepancy,

$$\max_t \frac{\big|\widehat{\Psi}^{\mathrm{ev}}_{\mathrm{unif},t}(h^\star) - \widehat{\Psi}^{\mathrm{ev}}_{\mathrm{peq},t}(h^\star)\big|}{\frac{1}{2}\big[\widehat{\Psi}^{\mathrm{ev}}_{\mathrm{unif},t}(h^\star) + \widehat{\Psi}^{\mathrm{ev}}_{\mathrm{peq},t}(h^\star)\big]},$$

equals 0.3%, and for every evaluation scramble,

$$\widehat{\Psi}^{\mathrm{ev}}_{\mathrm{mid},t}(h^\star) < \min\{\widehat{\Psi}^{\mathrm{ev}}_{\mathrm{unif},t}(h^\star), \widehat{\Psi}^{\mathrm{ev}}_{\mathrm{peq},t}(h^\star)\}.$$

The 0.3% endpoint discrepancy makes this a numerical near-equal-profile analogue of corollary 7.

Learners. The four learners are specified in table 12. KRR, RF, and MLP hyperparameters are calibrated on the disjoint F0 field before the primary F1–F3 risk study. The primary experiment uses sixteen paired training seeds, and its evaluation criteria and comparison statistics were specified beforehand.

Table 12: Learner configurations used in the acquisition-law study. $\tilde{z}$ denotes task coordinates linearly scaled to $[-1, 1]^2$ using the common deployment rectangle.

learner	configuration
NW	zero-fallback uniform-ball NW at $h^\star = 0.10\,\mathrm{m}$
KRR	Gaussian kernel $k_b(z, z') = \exp[-\|z - z'\|^2/(2b^2)]$; $(b, \lambda) = (0.10\,\mathrm{m}, 10^{-4})$ selected once on F0; solve $[\mathbf{K}_b + n\lambda I + 10^{-10} I]c = y$ in float64 by Cholesky
RF (random-feature ridge)	512 fixed features $\phi_k(\tilde{z}) = \tanh\big(\sum_{j=1}^{2} \tilde{z}_j W_{jk} + b_k\big)$, with $W_{jk} \sim \mathcal{N}(0, 2^2)$, $b_k \sim \mathrm{Unif}[-\pi, \pi]$, seed 90210; solve $(\Phi^\top\Phi + n\lambda I)c_{\mathrm{RF}} = \Phi^\top y$ with $\lambda = 10^{-3}$ selected on F0
MLP	two tanh hidden layers of width 64 on $\tilde{z}$; scikit-learn `MLPRegressor` with `solver="lbfgs"`, `alpha`$= n\lambda$, and $\lambda = 10^{-5}$ selected on F0; caps `max_iter`$=$ 20000 and `max_fun`$=$ 100000, with `tol`$= 10^{-9}$

Local stress targets for the profile-equated pair. Let

$$\bar{m}^{\mathrm{sel}}_X(z,h^\star) := \frac{1}{4}\sum_{s=1}^{4} \widehat{m}^{\mathrm{sel}}_{X,s}(z,h^\star), \qquad X \in \{\mathrm{unif},\mathrm{peq},\mathrm{mid}\}.$$

Locations at which P_{unif} has smaller local mass than P_{peq}, $\bar{m}^{\mathrm{sel}}_{\mathrm{unif}}(z,h^\star) < \bar{m}^{\mathrm{sel}}_{\mathrm{peq}}(z,h^\star)$, are ranked by

$$\bar{\gamma}(z) := \frac{1}{4}\sum_{s=1}^{4} \frac{[\widehat{m}^{\mathrm{sel}}_{\mathrm{unif},s}(z,h^\star) - \widehat{m}^{\mathrm{sel}}_{\mathrm{peq},s}(z,h^\star)]^2}{\widehat{m}^{\mathrm{sel}}_{\mathrm{unif},s}(z,h^\star)\widehat{m}^{\mathrm{sel}}_{\mathrm{peq},s}(z,h^\star)\widehat{m}^{\mathrm{sel}}_{\mathrm{mid},s}(z,h^\star)},$$

which is four times the selection-scramble-averaged midpoint mixture-gap integrand in eq. (11). Locations where P_{peq} has smaller local mass are defined with the roles reversed.

Three centers from each side are selected greedily from the selection mass fields in decreasing order of $\bar{\gamma}$ on the 61×61 coverage grid $\mathcal{G}_{\mathrm{Panda}}$, subject to a margin of at least $h^\star$ from the rectangle boundary and separation of at least $2h^\star$ from all previously selected centers on either side. Rounded to three decimals in metres, they are

$$c^{\mathrm{unif}}_{1,2,3} = (0.530, 0.364), (0.383, 0.504), (0.332, 0.308),$$

and

$$c^{\mathrm{peq}}_{1,2,3} = (0.537, 0.728), (0.303, 0.728), (0.376, 0.112).$$

For each center c, define the symmetric local stress targets

$$f^{\pm}_c(Z) = \pm\left(1 - \frac{\|Z-c\|_2}{h^\star}\right)_+.$$

They have absolute height one and support radius $h^\star$ and are used only for this empirical local-coverage stress test; they are not range-normalized.

For $X \in \{\mathrm{unif},\mathrm{peq},\mathrm{mid}\}$, learner L, budget n, paired training seed k, and $j \in \{1,2,3\}$, let $R^{(k,L,n)}_X(f)$ denote the deployment mean-squared test risk of L trained on n examples from P_X with target f. Define

$$R^{(k,L,n)}_{X,\mathrm{unif},j} := \frac{1}{2}\left[R^{(k,L,n)}_X(f^{+}_{c^{\mathrm{unif}}_j}) + R^{(k,L,n)}_X(f^{-}_{c^{\mathrm{unif}}_j})\right],$$

$$R^{(k,L,n)}_{X,\mathrm{peq},j} := \frac{1}{2}\left[R^{(k,L,n)}_X(f^{+}_{c^{\mathrm{peq}}_j}) + R^{(k,L,n)}_X(f^{-}_{c^{\mathrm{peq}}_j})\right],$$

and the pooled stress risk

$$R^{(k,L,n)}_{X,j} := \frac{1}{2}\left(R^{(k,L,n)}_{X,\mathrm{unif},j} + R^{(k,L,n)}_{X,\mathrm{peq},j}\right).$$

The local-stress criterion uses the pooled risks $R^{(k,L,n)}_{X,j}$. A stress cell is one center pair and one label budget, giving six cells in total.

MLP evaluation criteria. The confirmatory criteria in this study, specified beforehand, are MLP-based; the low-coverage mechanism check additionally contains a source-coverage condition. For an arm X, field $F \in \{\mathrm{F1},\mathrm{F2},\mathrm{F3}\}$, budget $n \in \{200, 800\}$, and paired training seed $k \in \{1,\dots,16\}$, let $R^{(k)}_X(F,n)$ denote the deployment mean-squared test risk of the MLP trained under arm X, and

$$\ell_k(X{:}Y;F,n) := \log R^{(k)}_X(F,n) - \log R^{(k)}_Y(F,n).$$

For each of the six natural target–budget cells define

$$a(X{:}Y;F,n) := \exp \mathrm{median}_k\, \ell_k(X{:}Y;F,n),$$

and the across-cell aggregate

$$\bar{a}(X{:}Y) := \mathrm{median}_{(F,n)}\, a(X{:}Y;F,n).$$

For each of 2000 paired bootstrap draws, the same resampled multiset $\mathcal{I}_b$ of the sixteen seed indices is used across all arms and cells. Define

$$a^{\mathrm{boot}}_b(X{:}Y;F,n) := \exp \mathrm{median}_{k\in\mathcal{I}_b}\, \ell_k(X{:}Y;F,n), \qquad T_b(X{:}Y) := \max_{(F,n)} \log a^{\mathrm{boot}}_b(X{:}Y;F,n).$$

A two-arm comparison with point threshold κ is satisfied only when

(i) $\bar{a}(X{:}Y) \le \kappa$;

(ii) $a(X{:}Y; F, n) < 1$ in at least five of the six cells; and

(iii) the empirical 95th percentile of $T_b(X{:}Y)$ is negative.

For the profile-equated pair, define

$$\ell_k^{\max}(F, n) := \max\{\ell_k(\text{mid:unif}; F, n), \ell_k(\text{mid:peq}; F, n)\},$$

with $a^{\max}(F, n)$, $\bar{a}^{\max}$, and $T_b^{\max}$ defined analogously. The same construction is applied to the six local-stress cells using $\log R_{X,j}^{(k,\text{MLP},n)}$ in place of $\log R_X^{(k)}(F, n)$.

For the low-coverage mechanism check, let $\mathcal{L}$ be the bottom 20% of the $16{,}384$ deployment test points ranked by evaluation-scramble-averaged equal-mixing mass at $h^\star$. Let $R_{X,\mathcal{L}}^{(k)}(F, n)$ denote the test risk restricted to this subset, $\bar{a}_{\mathcal{L}}$ the corresponding overall aggregate, and $\widehat{E}_X(n)$ the realized fraction of test points with no training example within $h^\star$ under arm X at budget n, reported as the median over seeds.

The confirmatory evaluation uses five MLP criteria:

1. **Natural-target mixture improvement.** The Ψ-optimal mixture must satisfy the two-arm criterion with $\kappa = 0.95$ against both equal mixing and the best single source.
2. **Scalar-ranking comparison.** Conditional on the first criterion, the Ψ-optimal mixture must satisfy $\bar{a}(\text{full:}\Psi\text{-ranking}) < 1$ together with conditions (ii)–(iii) above.
3. **Profile-equated pair on F1–F3.** The midpoint must satisfy $\bar{a}^{\max} < 1$ with conditions (ii)–(iii) applied to $a^{\max}$ and $T_b^{\max}$.
4. **Profile-equated local-stress targets.** The same better-endpoint criterion must hold over the six local-stress cells.
5. **Low-coverage mechanism check.** Conditional on the first criterion,

   $$\widehat{E}_{\text{full}}(n) < \widehat{E}_{\text{equal}}(n)$$

   must hold at both budgets, together with

   $$\bar{a}_{\mathcal{L}}(\text{full:equal}) < 1$$

   and conditions (ii)–(iii) on the restricted risks.

Here "full" denotes the Ψ-optimal mixture, "equal" equal mixing, and "best" the best single source by $\Psi(h^\star)$. The MLP fit diagnostics track failures and iteration-cap hits, with thresholds of zero failures, 5% cap hits overall, and 10% per arm.

MLP outcome. The MLP response is target-dependent. The Ψ-optimal mixture has aggregate MLP risk ratios of 1.00 relative to equal mixing, 1.00 relative to the best single source, and 1.05 relative to Ψ-ranking. Relative to equal mixing, the two F1 cells favor the mixture, whereas F3 at $n = 200$ has higher risk. All 1920 MLP fits converged with zero failures. These target-dependent outcomes do not satisfy the five confirmatory criteria. The direct comparison ratios correspond to the criteria above; table 13 reports every arm relative to equal mixing.

Global matching and the ideal control. Direct Q-matching yields lower aggregate point estimates than equal mixing for every learner, with ratios between 0.21 and 0.51. Both global-matching baselines select the same pure goal-biased source. That source places 89% of its probability inside the deployment rectangle, but its $\Psi(h^\star)$ is $2.3\times$ that of the best single source, and its aggregate risks are approximately 2.1–3.3 times those under equal mixing across the four learners. Thus, in this experiment, matching global moments or coarse occupancy selects a source with substantial finite-scale undercoverage despite strong global overlap with the deployment region, whereas $\Psi(h^\star)$ identified this undercoverage before training.

Profile-based allocation. Relative to equal mixing, Ψ-ranking has aggregate risk ratios 0.87 for NW, 0.94 for KRR, and 0.92 for RF, with one-sided 95% paired-bootstrap upper limits below one. Its MLP point estimate is 0.96.

The Ψ-optimal mixture has aggregate risk ratios of 0.87 for KRR and 0.92 for NW relative to equal mixing; for each of these two learners, the cellwise point estimate favors the mixture in five of the six target–budget settings. Its MLP aggregate ratio is 1.00 (table 13). The MLP response is strongly target-dependent; for example, the two F1 ratios are 0.89 and 0.91, whereas F3 at $n = 200$ gives 1.78. The Ψ-ranking rule has lower aggregate risk point estimates than the Ψ-optimal mixture for NW, RF, and MLP, while KRR favors the Ψ-optimal mixture. The convex program optimizes the fixed-scale profile, whereas these comparisons evaluate finite-budget prediction risk on heterogeneous targets.

Profile-equated pair. For the local stress targets, define the side-specific seed-median ratios

$$\rho^{(L,n)}_{\mathrm{unif},j} := \mathrm{median}_k \left[\frac{R^{(k,L,n)}_{\mathrm{mid,unif},j}}{R^{(k,L,n)}_{\mathrm{unif,unif},j}} \right], \qquad \rho^{(L,n)}_{\mathrm{peq},j} := \mathrm{median}_k \left[\frac{R^{(k,L,n)}_{\mathrm{mid,peq},j}}{R^{(k,L,n)}_{\mathrm{peq,peq},j}} \right].$$

Across $j \in \{1, 2, 3\}$ and $n \in \{200, 800\}$, the MLP ratios satisfy

$$\rho^{(\mathrm{MLP},n)}_{\mathrm{unif},j} \in [0.51, 1.02], \qquad \rho^{(\mathrm{MLP},n)}_{\mathrm{peq},j} \in [0.28, 0.87].$$

Thus the midpoint often improves prediction on the side where one endpoint is locally deficient, but the improvement is not uniform across every unif-side cell.

On the natural targets F1–F3, the midpoint has lower aggregate risk than equal mixing for all four learners, while the cellwise comparison with the better profile-equated endpoint is mixed. The realized local counts support the intended coverage mechanism: at the unif-deficient centers, the counts under P_{mid} are 2.7–$4.3\times$ those under the locally deficient endpoint P_{unif}.

The finite-scale profile provides a target-agnostic coverage criterion, while the downstream value of reallocating that coverage depends on the learner and target.

Table 13: Aggregate risk ratios relative to equal mixing; values below one indicate lower risk. Entries are medians over the six F1–F3 target–budget cells. All rows use 16 paired training seeds except the global-matching row, which is reported from the 8-seed robustness run. The MLP has two hidden layers of width 64.

arm	NW	KRR	RF	MLP
Ψ-optimal mixture	0.92	0.87	0.99	1.00
Ψ-ranking	0.87	0.94	0.92	0.96
best single ($= P_{\mathrm{unif}}$)	0.86	1.06	1.01	1.05
pair P_{mid}	0.90	0.91	0.86	0.94
pair P_{peq}	1.18	1.17	1.10	1.11
Q-matched (ideal control)	0.51	0.33	0.25	0.21
moment and histogram matching (both select pure goal-biased)	2.64	3.26	2.11	2.68