Approximating an Ideal Experiment with Induced Order Statistics*

Federico A. Bugni
Department of Economics
Northwestern University
federico.bugni@northwestern.edu

Ivan A. Canay
Department of Economics
Northwestern University
iacanay@northwestern.edu

Deborah Kim
Department of Economics
University of Warwick
deborah.kim@warwick.ac.uk

November 12, 2026

Abstract

Many economic questions concern the conditional distribution of an outcome at a fixed covariate value. We refer to observing i.i.d. draws from this distribution as an *ideal experiment*, under which many statistical procedures admit exact, finite-sample guarantees. When the covariate is continuously distributed, however, this ideal experiment is unavailable. We provide a general method for approximating this ideal experiment using *induced order statistics*, the outcomes attached to the nearest observed covariates. We derive explicit bounds on the approximation error in Hellinger and total variation distances. These bounds quantify how local approximation errors accumulate as the nearest-neighbor sample grows and provide a unified way to transfer rejection and coverage guarantees across statistical procedures. Under quadratic mean differentiability and local regularity of the covariate distribution, the bounds apply to both interior and boundary conditioning points and give explicit conditions on the growth of the nearest-neighbor sample. We illustrate the approach in four distinct applications: covariate balance testing in regression discontinuity designs, conditional quantile inference, conformal prediction at a fixed covariate value, and policy-irrelevance testing under network interference.

KEYWORDS: total variation distance, Hellinger distance, induced order statistics, nearest neighbors, boundary points, regression discontinuity designs.

JEL classification codes: C12, C14.

*We thank Matias Cattaneo for comments and discussion. All errors are our own.

1 Introduction

Many economic questions concern the conditional distribution of an outcome at a fixed covariate value. Fix such a value x_0, and let P denote the conditional distribution of an outcome Y given $X = x_0$. In this paper, we refer to observing an i.i.d. sample from P as the *ideal experiment*. If this ideal experiment were available, many of these economic questions could be addressed using statistical procedures with exact, finite-sample guarantees. In a regression discontinuity design, for example, covariate balance testing would be a standard two-sample problem if we could draw independent samples of treated and untreated units at the cutoff. An i.i.d. sample from P would similarly permit finite-sample rank-based inference on a conditional quantile. Related ideal experiments arise when constructing conformal prediction sets at a fixed covariate value and when testing policy irrelevance at a target network configuration.

If X is continuously distributed, the observed sample does not provide an ideal experiment. In this case, the event $X = x_0$ has probability zero, and so the observed sample contains no observations drawn exactly from P. A natural approximation uses the outcomes attached to the k observations whose covariates lie closest to x_0. These outcomes are *induced order statistics* (IOS), or, equivalently for our purposes, the responses in a k-nearest-neighbor sample. The quality of this approximation depends on two opposing forces. Keeping k fixed ensures that the selected covariates become arbitrarily close to x_0, but limits the information available for inference. Allowing k to grow provides more observations but can also increase the discrepancy between the feasible sample and the ideal experiment.

Classical work on IOS develops their probabilistic properties, limit distributions, and representations in terms of mixtures and conditional laws; see, among others, David and Galambos (1974); Bhattacharya (1974); Reiss (1989); Kaufmann and Reiss (1992) and the book-length treatment by Falk et al. (2010). Empirical practice often relies on asymptotic analyses that keep the dimension of the IOS vector fixed, reflecting the lack of readily available theoretical results that allow this dimension to grow. On the theoretical front, the main general result allowing k to grow, Falk et al. (2010, Theorem 3.5.2), relies on a multiplicative expansion of the joint density in a full neighborhood of x_0. This condition yields a fast approximation rate but restricts how the conditional support of Y may vary with X and rules out the possibility that x_0 is a boundary point. Boundary points are particularly important in regression discontinuity designs, where each side of the cutoff is analyzed separately, and arise naturally in other applications considered below.

This paper derives approximation rates for the entire IOS vector under primitive conditions that accommodate both interior and boundary conditioning points. Let d denote the dimension of X. Let S_n collect the outcomes attached to the k covariate observations nearest to x_0, and let S denote the ideal experiment, consisting of k i.i.d. draws from P. We study the Hellinger and total variation distances between the two laws, $\mathcal{L}(S_n)$ and $\mathcal{L}(S)$. Hellinger distance tensorizes over product measures and describes how local approximation errors accumulate across the k observations. Total variation distance controls the maximal discrepancy in event probabilities, translating

the joint approximation into bounds on rejection and coverage probabilities.

Our analysis proceeds in two steps. For $r > 0$, let P_r denote the outcome distribution among observations whose covariates X lie within distance r of x_0. The local approximation concerns the distance between P_r and P as r shrinks. Our high-level result shows how this discrepancy accumulates across the k selected outcomes and thereby determines the joint approximation between $\mathcal{L}(S_n)$ and $\mathcal{L}(S)$. We next provide primitive conditions for the local approximation. Our main condition requires the conditional densities to be differentiable in quadratic mean at x_0. Combined with local regularity of the covariate distribution and its support, this condition implies that P_r approaches P at rate $O(r)$ in both Hellinger and total variation distances, whether x_0 is an interior point or a boundary point. Consequently,

$$\begin{aligned} \mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) &= O\big(k^{1/2}(k/n)^{1/d}\big) \ , \\ \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) &= O\big(k^{1/2}(k/n)^{1/d}\big) \ . \end{aligned} \tag{1.1}$$

The feasible experiment therefore approximates the ideal one whenever $k = o(n^{2/(d+2)})$. For a scalar covariate, this condition becomes $k = o(n^{2/3})$. We show that the local approximation rate underlying (1.1) is sharp if x_0 lies on the boundary of the support of X. At interior points, both local distances are $o(r)$ for every law satisfying our conditions, but there is such a model for which neither distance is $O(r^{1+\varepsilon})$ for any $\varepsilon > 0$. For comparison, Falk et al. (2010) obtain the less restrictive sufficient condition $k = o(n^{4/(d+4)})$ under stronger smoothness conditions at an interior conditioning point. Section 3.2 discusses the source of this difference, while the supplemental appendix provides complementary results under a Taylor/Hölder condition.

The joint total variation bound provides a general blueprint for inference with IOS. It has three steps: establish a finite-sample guarantee for a procedure in the ideal experiment, apply the same procedure to S_n, and use the total variation bound to quantify the resulting approximation error. The feasible guarantee and an explicit error rate then follow without a separate approximation argument for each application.

We apply this argument in four settings. First, we revisit the covariate balance test for regression discontinuity designs proposed by Canay and Kamat (2018). In the ideal experiment, the observations selected on the two sides of the cutoff are exchangeable under the null, and the permutation test has finite-sample size control. Writing k for the total number of observations selected across the two sides, our approximation yields a rejection probability bounded by $\alpha + O(k^{3/2}/n)$. This bound quantifies the approximation error as the local sample grows and gives the sufficient restriction $k = o(n^{2/3})$ for asymptotic size control. The original Canay–Kamat rule of thumb, whose deterministic rate component is $n^{0.9}/\log n$, lies outside this guarantee. Section 4.1 describes a modification satisfying the restriction.

For our second application, we construct a confidence interval for a conditional quantile using two order statistics of the k nearest outcomes. In the ideal experiment, their indices are cali-

brated using exact binomial probabilities. The finite-sample rank argument remains valid when the conditional distribution has an atom at the target quantile, in which case the interval may be conservative. Our total variation bound transfers this coverage guarantee directly to the feasible interval. The construction is closely related to the direct approach of Fan and Liu (2016). Their quantile-specific analysis permits a faster growth rate under smooth-interior conditions, whereas our general approximation result yields a sufficient growth condition under assumptions that accommodate boundary points.

In our third application, we construct a split-conformal prediction set for the outcome of a new observation. We study its coverage conditional on the observed covariate value, a pointwise target studied by Guan (2023) and Min et al. (2026). Conditional on an independent training sample, the usual rank argument gives finite-sample coverage in the ideal experiment. Calibrating with the nearest outcomes carries this guarantee to the feasible prediction set up to the joint total variation error. Under our main conditions, this gives coverage loss $O(k^{1/2}(k/n)^{1/d})$, including at boundary points. A separate argument exploiting conditional exchangeability sharpens the bound for this prediction rule. The argument requires no continuity of the nonconformity-score distribution; ties can only make the ideal procedure conservative.

As our final application, we study the policy-irrelevance test under network interference considered by Auerbach et al. (2026). For each target, the problem can be represented as an IOS experiment after each community is reduced to the outcome attached to its closest rooted network and the corresponding minimum distance. The ideal experiment yields a finite-sample permutation argument, while the feasible procedure uses the outcomes associated with the closest configurations. Whereas the original paper establishes asymptotic validity of the test for a fixed local sample size k, our results establish validity with k growing at an explicit rate with the number of independent communities, under a prespecified sample split and suitable local regularity conditions.

The remainder of the paper is organized as follows. Section 2 defines the ideal and feasible experiments, recalls the mixture representation of the feasible law, and presents the high-level approximation result. Section 3 gives primitive conditions based on quadratic mean differentiability, establishes the boundary and interior results, and compares them with the classical IOS benchmark. Section 4 develops the four applications, and Section 5 concludes. The main appendix contains the proofs and the conditional-quantile rank result; the supplemental appendix contains alternative smoothness results and their proofs.

2 The Ideal and Feasible Experiments

What we call an *ideal experiment* consists of observing i.i.d. draws from the conditional distribution of Y given $X = x_0$. We approximate it using *induced order statistics*, the outcomes attached to the k observed covariate values closest to x_0, and refer to observing these outcomes as the *feasible experiment*. We first formalize these two experiments, then recall a mixture representation for the

feasible law and state the high-level result that converts local approximation rates into rates for the entire IOS vector.

2.1 The two experiments

Let $X \in \mathcal{X} \subseteq \mathbf{R}^d$ and $Y \in \mathcal{Y} \subseteq \mathbf{R}^m$, and fix $x_0 \in \mathcal{X}$. Let $\mathbf{Q}$ be a model class of laws of (X, Y), each dominated by the product measure $d\mu := dx \otimes \nu(dy)$, where dx is Lebesgue measure on $\mathbf{R}^d$ and ν is an arbitrary fixed σ-finite measure on $\mathbf{R}^m$. The supports $\mathcal{X}$ and $\mathcal{Y}$ may depend on Q; we suppress this dependence in the notation. The conditioning point x_0 remains fixed. Fix $Q \in \mathbf{Q}$ and denote its joint density with respect to $d\mu$ by $f(x, y)$ and the marginal density of X by

$$g(x) := \int f(x, y)\, \nu(dy) \ .$$

We denote the conditional distribution of Y given $X = x_0$ by P. For any $x \in \mathbf{R}^d$ with $g(x) > 0$, we denote the density of the conditional distribution of Y given $X = x$ by

$$p_x(y) := \frac{f(x, y)}{g(x)} \ .$$

This setup allows Y to be continuous, discrete, or mixed, while requiring that X admits a Lebesgue density.

Let $r > 0$ be given. We denote the Euclidean distance of X from x_0 by

$$R := \|X - x_0\| \tag{2.1}$$

and the open ball of radius r with center x_0 by

$$B_r := \{x \in \mathbf{R}^d : \|x - x_0\| < r\} \ .$$

We denote the conditional distribution of Y given $X \in B_r$ by P_r and its density by

$$h_r(y) := \frac{\int_{B_r} f(x, y)\, dx}{\int_{B_r} g(x)\, dx} \ .$$

Under the smoothness conditions introduced in Section 3, it follows that $h_r(y)$ approximates $p_{x_0}(y)$ as $r \to 0$. For notational simplicity, the dependence of B_r and $h_r(y)$ on x_0 is kept implicit, since the center x_0 remains fixed throughout the paper.

We observe an i.i.d. sample of size n from Q and denote it by

$$\{(X_i, Y_i) : 1 \le i \le n\} \ . \tag{2.2}$$

We also denote the order statistics of $\{R_i : 1 \le i \le n\}$ in (2.1) by $R_{(1)} \le \cdots \le R_{(n)}$. Since X

admits a Lebesgue density, ties occur with probability zero. Define the (random) permutation σ_n of $\{1,\dots,n\}$ by

$$\sigma_n(i) = j \quad \Longleftrightarrow \quad R_i = R_{(j)} \ , \qquad i,j = 1,\dots,n \ ,$$

so that $\sigma_n(i)$ is the rank of observation i. The index set of the k observations closest to x_0 is

$$K_n := \{\, i \in \{1,\dots,n\} : \ \sigma_n(i) \le k \,\} \ .$$

Let ι_n be the increasing rearrangement of K_n (i.e., $\iota_n(1) < \cdots < \iota_n(k)$ and $\{\iota_n(1),\dots,\iota_n(k)\} = K_n$). Finally, for $j = 1,\dots,k$, let $Y_{\iota_n(j)}$ be the response attached to the observation whose distance from x_0 ranks among the k smallest, with the indices $\iota_n(1) < \cdots < \iota_n(k)$ listed in the original sample order. These outcomes are the *induced order statistics*, listed in original sample order (equivalently, a permutation of the distance-ordered concomitants; see, e.g., David and Galambos (1974); Bhattacharya (1974); Canay and Kamat (2018)).

Collect these observations into

$$S_n := (S_{n,1},\dots,S_{n,k}) := \big(Y_{\iota_n(1)},\dots,Y_{\iota_n(k)}\big) \ . \tag{2.3}$$

By definition, S_n collects the outcomes attached to the k covariates closest to x_0. These observations can be used as an approximation to the following counterpart random vector

$$S := (S_1,\dots,S_k) := (\tilde{Y}_1,\dots,\tilde{Y}_k), \quad \text{with} \quad \tilde{Y}_1,\dots,\tilde{Y}_k \ \text{ i.i.d. with } P \ . \tag{2.4}$$

We refer to S_n as the *feasible experiment* and to S as the corresponding *ideal experiment*. This paper studies the approximation problem of comparing the laws of these two random vectors. Section 4 gives detailed examples.

Example 1. *Let $m = d = 1$, $n = 10$, $k = 4$, and suppose $x_0 = 0$. Consider the hypothetical sample*

$$\{(-3,2),\ (4,-1),\ (-1,0),\ (2,5),\ (0.5,1),\ (-0.2,3),\ (7,4),\ (-5,-2),\ (1.4,6),\ (-0.7,8)\} \ .$$

In this case $R_i = |X_i|$ and so the rank map σ_n is given by

i	1	2	3	4	5	6	7	8	9	10
$\sigma_n(i)$	7	8	4	6	2	1	10	9	5	3

.

Since $k = 4$, the index set of the k nearest observations is

$$K_n = \{i : \sigma_n(i) \le 4\} = \{6,5,10,3\} \ .$$

Applying the rearrangement determined by ι_n we get $\iota_n = (3,5,6,10)$. Hence the induced order

statistics are the Y-values attached to these indices, listed in the original sample order:

$$S_n = (Y_{\iota_n(1)}, Y_{\iota_n(2)}, Y_{\iota_n(3)}, Y_{\iota_n(4)}) = (Y_3, Y_5, Y_6, Y_{10}) = (0, 1, 3, 8) \ .$$

■

2.2 Mixture representation and transfer

We use $\mathcal{L}(\cdot)$ to denote a probability law. The mixture representation of S_n makes the approximation problem transparent. Our assumptions below imply that $R = \|X - x_0\|$ has a continuous CDF and that $P(R < r) > 0$ for every $r > 0$. Under these conditions, Kaufmann and Reiss (1992, Theorem 1) implies that, for any $k < n$ and $r > 0$, the conditional law of S_n given $R_{(k+1)} = r$ is that of k i.i.d. draws from P_r:

$$\mathcal{L}\big(S_n \mid R_{(k+1)} = r\big) = P_r^k \ .$$

If P_r approximates the target conditional law P for small $r > 0$ and $R_{(k+1)}$ concentrates near zero, the preceding representation suggests

$$\mathcal{L}(S_n) = \int P_r^k \, \mathcal{L}\big(R_{(k+1)}\big)(dr) \approx \int P^k \, \mathcal{L}\big(R_{(k+1)}\big)(dr) = P^k = \mathcal{L}(S) \, . \tag{2.5}$$

In words, the Y-values attached to the k nearest X-neighbors of x_0 behave approximately like k i.i.d. draws from P. Our results below give conditions under which this approximation is valid and quantify its accuracy as k grows.

This approximation has a direct implication for hypothesis testing. Let $\phi : (\mathbf{R}^m)^k \to [0, 1]$ be any measurable function. This allows ϕ to represent the rejection indicator of a nonrandomized test or the rejection probability of a randomized test given the sample. Since $\phi \leq 1$, the variational characterization of total variation implies

$$\big|E_{Q^n}[\phi(S_n)] - E_{P^k}[\phi(S)]\big| \ \leq \ \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ , \tag{2.6}$$

where TV denotes the total variation distance. Moreover, the standard comparison between total variation and the Hellinger distance H yields

$$\mathrm{H}^2\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ \leq \ \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ \leq \ \sqrt{2}\,\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ . \tag{2.7}$$

Together, (2.6) and (2.7) imply that a bound in either distance lets us carry probability guarantees from the ideal experiment to the feasible experiment, with an explicit bound on the error. This is the transfer step used in the applications below.

For estimation, let $\psi_k : (\mathbf{R}^m)^k \to \mathbf{R}^\ell$ estimate a parameter $\theta(P)$, and let L be a loss satisfying

$0 \leq L(\psi_k(s), \theta(P)) \leq \bar{L} < \infty$ uniformly in s and k. Applying (2.6) to $L(\psi_k(\cdot), \theta(P))/\bar{L}$ gives

$$E_{Q^n}\big[L\big(\psi_k(S_n), \theta(P)\big)\big] \leq E_{P^k}\big[L\big(\psi_k(S), \theta(P)\big)\big] + \bar{L} \times \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ .$$

The first term is the risk in the ideal k-sample experiment and typically decreases with k. For example, under squared-error loss it is of order $O(k^{-1})$ for the sample mean with bounded outcomes and for a pointwise empirical CDF. The second term is the approximation error from using the k nearest observations, and our bound on this term increases with k for fixed n. Thus, k trades off statistical precision in the ideal experiment against the accuracy with which that experiment is approximated by the IOS vector.

2.3 From local to joint approximation

In this section, we derive rates for $\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big)$ and $\mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big)$ in terms of the local approximation errors $\mathrm{H}(P_r, P)$ and $\mathrm{TV}(P_r, P)$. We allow x_0 to be either an interior point or a boundary point of $\mathcal{X}$. The point $x_0 \in \mathcal{X}$ is an interior point if $B_r \subset \mathcal{X}$ for some $r > 0$, and a boundary point otherwise. This distinction is required to accommodate applications in regression discontinuity designs.

We begin with a mild regularity condition on the marginal density g at x_0.

Assumption 1 (Local regularity of g at x_0)**.** *The law Q satisfies the following conditions:*

(i) The marginal density g of X satisfies $g(x_0) > 0$ and

$$|g(x) - g(x_0)| \leq C_g \|x - x_0\|$$

for all $x \in \mathcal{X}$ in a neighborhood of x_0 and some constant $C_g < \infty$.

(ii) There exist constants $C_{x_0} > 0$ and $r_0 > 0$ such that

$$\mathrm{Vol}\big(B_r \cap \mathcal{X}\big) \geq C_{x_0} r^d \qquad \text{for all } 0 < r < r_0 \ ,$$

where Vol *denotes volume.*

Continuity of the distribution of X together with positivity $g(x_0) > 0$ are standard primitives in the induced order statistics literature; see, for example, Reiss (1989), Kaufmann and Reiss (1992), and Falk et al. (2010). We strengthen these conditions slightly by imposing local Lipschitz regularity of g at x_0 and a local thickness condition on $B_r \cap \mathcal{X}$, a formulation that remains compatible with x_0 being a boundary point.

Although Assumption 1 may appear unusual at first glance, it is quite mild once one distinguishes between interior and boundary points. If x_0 is an interior point of $\mathcal{X}$, then for all sufficiently

small r we have $B_r \subset \mathcal{X}$, hence $\mathrm{Vol}(B_r \cap \mathcal{X}) = \mathrm{Vol}(B_r)$ is of order r^d, so condition (ii) holds automatically. In this case, Assumption 1 reduces to condition (i), which imposes local Lipschitz continuity of g at x_0. If x_0 is a boundary point of $\mathcal{X}$, condition (ii) simply requires that $\mathcal{X}$ occupy a nonvanishing fraction of each small ball centered at x_0, a property that holds, for example, when $\mathcal{X}$ has a Lipschitz boundary near x_0. Condition (i) remains a one-sided Lipschitz condition on g at x_0.

Theorem 1. *Let Assumption 1 hold. Suppose that* $\mathrm{H}(P_r, P) = O(r^{a_h})$ *for some* $a_h \geq 0$ *and* $\mathrm{TV}(P_r, P) = O(r^{a_{tv}})$ *for some* $a_{tv} \geq 0$. *Then the following bounds hold uniformly for* $n \in \mathbf{N}$ *and* $1 \leq k < n$*:*

(a) $\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{a_h/d}\big)$.

(b) $\mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\left(\min\left\{k(k/n)^{a_{tv}/d}, k^{1/2}(k/n)^{a_h/d}\right\}\right)$.

Theorem 1 combines the local approximation rates with the nearest-neighbor radius, whose relevant scale under Assumption 1 is $(k/n)^{1/d}$. Tensorization of the Hellinger affinity over the conditional product laws yields the $k^{1/2}$ factor in part (a). For total variation, part (b) retains the smaller of the direct product bound and the bound obtained from Hellinger distance. The latter yields the faster rate under the smoothness conditions considered in later sections. The direct bound yields a sharper rate under the alternative smoothness conditions in Appendix S.2, where the local total variation distance converges faster than the Hellinger distance.

By (2.6), the resulting total variation rate bounds the difference between probabilities under the ideal and feasible experiments. The proof of Theorem 1 adapts the argument of Falk et al. (2010, Theorem 3.5.2) to accommodate general local exponents and to obtain the direct total variation bound.

3 Primitive Conditions and Boundary-Valid Rates

The high-level result in Theorem 1 reduces the joint approximation problem to the local behavior of P_r around P. In this section, we establish the required local rates under quadratic mean differentiability, study their sharpness at boundary and interior points, and compare them with the smooth-interior benchmark that is currently available in the literature.

3.1 Quadratic Mean Differentiability

We develop local approximation rates for the conditional densities p_x at x_0 under quadratic mean differentiability (QMD). We show that QMD yields linear local approximation rate of P_r to P in both Hellinger and total variation distances. We also study sharpness and the limits of possible

improvements over the linear rate, distinguishing what can and cannot be strengthened at boundary points versus interior points.

Assumption 2 (QMD at x_0)**.** *The conditional densities p_x of Q, defined in Section 2.1, are such that there exists a measurable function $\dot{\ell}_{x_0} : \mathbf{R}^m \to \mathbf{R}^d$ satisfying*

$$\int \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y)\nu(dy) < \infty \ ,$$

and, as $t \to 0$ through values satisfying $x_0 + t \in \mathcal{X}$,

$$\int \left[\sqrt{p_{x_0+t}(y)} - \sqrt{p_{x_0}(y)} - \tfrac{1}{2}\big(t^\top \dot{\ell}_{x_0}(y)\big)\sqrt{p_{x_0}(y)}\right]^2 \nu(dy) = o(\|t\|^2) \ .$$

QMD is a central smoothness condition in asymptotic statistics. It underlies the local asymptotic normality (LAN) framework and plays a key role in the asymptotic theory of likelihood-based procedures (see, e.g., Le Cam (1986), Van der Vaart (2000)), with $\dot{\ell}_{x_0}$ serving as the score function. Beyond likelihood theory, QMD also supports diffusion/LAN approximations in statistical decision problems, such as optimal design and inference for adaptive experiments and bandits, by justifying Gaussian limit experiments and the associated risk analyses; see Adusumilli (2025).

QMD imposes weaker restrictions on the local behavior of the conditional law than the smooth-interior conditions proposed by Falk et al. (2010). We make this comparison precise in Section 3.2, where we introduce the latter conditions. The next theorem establishes the local approximation rates under QMD, gives a boundary example attaining the linear rate, and gives an interior example admitting no positive polynomial improvement. For the existence claims below, the dominating measure ν may be chosen as part of the construction.

Theorem 2. *Suppose Assumptions 1–2 hold. Then,*

$$\mathrm{H}(P_r, P) = O(r) \quad \textit{and} \quad \mathrm{TV}(P_r, P) = O(r) \ . \tag{3.1}$$

Moreover:

(i) Boundary sharpness. *There exists a law $Q \in \mathbf{Q}$ satisfying Assumptions 1–2 such that x_0 lies on the boundary of $\mathcal{X}$, and constants $C > 0$ and $\bar{r} > 0$ such that*

$$\mathrm{H}(P_r, P) \geq Cr \quad \textit{and} \quad \mathrm{TV}(P_r, P) \geq Cr \quad \textit{for all } r \in (0, \bar{r}) \ .$$

(ii) Interior refinement and its limitation. *For every $Q \in \mathbf{Q}$ satisfying Assumptions 1–2 such that x_0 lies in the interior of $\mathcal{X}$,*

$$\mathrm{H}(P_r, P) = o(r) \quad \textit{and} \quad \mathrm{TV}(P_r, P) = o(r) \ .$$

However, for every $\tilde{r} > 0$, there exists a law $Q \in \mathbf{Q}$ satisfying these assumptions, with $B_{\tilde{r}} \subseteq \mathcal{X}$, such that, for every $\varepsilon > 0$,

$$\frac{\mathrm{H}(P_r, P)}{r^{1+\varepsilon}} \longrightarrow \infty \qquad \textit{and} \qquad \frac{\mathrm{TV}(P_r, P)}{r^{1+\varepsilon}} \longrightarrow \infty \qquad \textit{as } r \to 0 \ . \tag{3.2}$$

Theorem 2 gives the local upper rate $O(r)$ in both distances at interior and boundary points. Part (i) shows that this rate is sharp at boundary points by providing a boundary example that attains it. Part (ii) shows that, at interior points, both distances are $o(r)$ for every fixed law satisfying the assumptions. Nevertheless, the $O(r)$ rates are *nearly sharp* in the sense that there exists a single interior QMD model for which neither distance is $O(r^{1+\varepsilon})$ for any $\varepsilon > 0$.

Theorem 2 implies that Theorem 1 applies with $a_h = a_{tv} = 1$, yielding

$$\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{1/d}\big) \ , \qquad \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{1/d}\big) \ .$$

Since $a_h = a_{tv} = 1$, the relevant bound for $\mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big)$ is provided by the inequality $\mathrm{TV} \leq \sqrt{2}\mathrm{H}$, and both distances share the same joint rate. This rate directly determines how fast k may grow relative to n while still ensuring convergence of $\mathcal{L}(S_n)$ to $\mathcal{L}(S)$:

$$k^{1/2}(k/n)^{1/d} \to 0 \qquad \Longleftrightarrow \qquad k = o\big(n^{2/(2+d)}\big) \ .$$

In particular, for $d = 1$ this becomes $k = o(n^{2/3})$. This threshold coincides with the growth conditions derived, through different arguments, in Bugni and Canay (2021) and Bugni et al. (2025) for the specific uses of induced order statistics in those papers.

Remark 1 (Joint approximation at interior points)**.** *Suppose Assumptions 1–2 hold and x_0 lies in the interior of $\mathcal{X}$. For each fixed law Q and every sequence $1 \leq k < n$ with $k/n \to 0$,*

$$\begin{aligned} \mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) &= o\big(k^{1/2}(k/n)^{1/d}\big) \ , \\ \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) &= o\big(k^{1/2}(k/n)^{1/d}\big) \ . \end{aligned}$$

Consequently, $k = O(n^{2/(d+2)})$ suffices for joint convergence at an interior point. This refinement is pointwise in Q. ■

Theorem 2 gives our default local approximation rate. Appendix S.2 provides complementary results under the Taylor/Hölder remainder condition in Assumption S.4. This condition and the QMD condition in Assumption 2 are nonnested. Assumption S.4 indexes smoothness by the Taylor order $\kappa_{\mathrm{s}} \in \mathbf{N}_0$ and the Hölder exponent $\kappa_{\mathrm{r}} \in (0, 1]$. Theorem S.1 gives exponents (a_h, a_{tv}) that depend on their sum and on whether x_0 is an interior point or not. At interior points, first-order terms cancel when integrated over a symmetric neighborhood. At boundary points, the support truncates this neighborhood, so cancellation may fail and rates can be slower. For example, in the Hölder-continuity case $\kappa_{\mathrm{s}} = 0$, $\mathrm{H}(P_r, P) = O(r^{\kappa_{\mathrm{r}}/2})$ and $\mathrm{TV}(P_r, P) = O(r^{\kappa_{\mathrm{r}}})$. As $\kappa_{\mathrm{r}} \downarrow 0$,

the admissible growth rate for k becomes correspondingly more restrictive; no common positive polynomial growth rate is available uniformly over $\kappa_{\mathrm{r}} \in (0, 1]$.

3.2 Comparison with the Standard Smooth-Interior Benchmark

The monograph Falk et al. (2010) provides an influential benchmark for the convergence of induced order statistics in Hellinger distance. Their result allows k to grow with n under the following condition.

Assumption 3 (FHR assumption)**.** *There exists $t_0 > 0$ such that, uniformly for $t \in \mathbf{R}^d$ with $\|t\| \le t_0$ and all $y \in \mathbf{R}^m$,*

$$f(x_0 + t, y) = f(x_0, y)\Big\{1 + t^\top \zeta_1(y) + O\big(\|t\|^2 \zeta_2(y)\big)\Big\} \ , \tag{3.3}$$

where $\zeta_1 : \mathbf{R}^m \to \mathbf{R}^d$, $\zeta_2 : \mathbf{R}^m \to [0, \infty)$, and

$$\int \big(\|\zeta_1(y)\|^2 + \zeta_2(y)^2\big) f(x_0, y) \nu(dy) < \infty \ . \tag{3.4}$$

The following result uses the preceding assumption to bound the joint approximation error between the feasible and ideal experiments.

Theorem 3 (Falk–Hüsler–Reiss, Theorem 3.5.2)**.** *Suppose $g(x_0) > 0$ and Assumption 3 holds. Then, uniformly over $n \in \mathbf{N}$ and $k \in \{1, \dots, n\}$,*

$$\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{2/d}\big) \ .$$

We now briefly explain the idea behind the proof of Theorem 3. Under Assumption 3 and $g(x_0) > 0$, Lemma S.5 establishes $\mathrm{H}(P_r, P) = O(r^2)$ and $\mathrm{TV}(P_r, P) = O(r^2)$. Part (a) of Theorem 1 therefore reproduces the Hellinger rate in Theorem 3. Part (b) also yields

$$\mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{2/d}\big) \ ,$$

so the same rate transfers finite-sample probability guarantees from the ideal experiment to the feasible one under the FHR conditions. The corresponding sufficient growth condition is $k = o(n^{4/(d+4)})$, compared with $k = o(n^{2/(d+2)})$ under Assumptions 1–2.

The previous paragraph shows that the standard smooth-interior conditions, given by Assumption 3 and $g(x_0) > 0$, permit faster growth of k than Assumptions 1–2. This improvement relies on stronger assumptions. Formally, we show in Lemma S.6 in the supplemental appendix that Assumption 3, together with $g(x_0) > 0$, implies Assumptions 1 and 2, with $x_0 \in \mathrm{int}(\mathcal{X})$, but the converse does not hold. The next paragraphs explain the key features of Assumption 3 that contribute to the faster FHR rate.

First, Assumption 3 requires the expansion in (3.3) to hold for all perturbations in a full neighborhood of x_0. For any $(x_0, y) \in \mathcal{X} \times \mathbf{R}^m$ with $f(x_0, y) > 0$, this expansion implies $f(x_0+t, y) > 0$ for all sufficiently small perturbations $t \in \mathbf{R}^d$. Thus, x_0 must be an interior point of $\mathcal{X}$. Otherwise, arbitrarily small perturbations could lie outside the support, where $f(x_0 + t, y) = 0$. At an interior point, integration over the symmetric ball B_r cancels the leading first-order contribution. At a boundary point, the support truncates the ball, so this cancellation need not occur.

Second, (3.3) imposes a quadratic bound on the relative remainder, $O(\|t\|^2\zeta_2(y))$, where ζ_2 is square-integrable under P. Together with the cancellation of the leading first-order contribution, this bound yields an approximation error of order r^2. By contrast, QMD in Assumption 2 requires only that the $L^2(\nu)$ remainder in the first-order expansion of $\sqrt{p_{x_0+t}}$ be $o(\|t\|)$, without specifying a polynomial rate. This is also the mechanism behind Theorem 2(ii): at an interior point, symmetry reduces the averaged first-order contribution to $O(r^2)$, but QMD supplies no polynomial rate for the remaining term. The interior example in Theorem 2(ii) shows that no positive polynomial improvement over $O(r)$ may be attainable, even for a single model.

The multiplicative form in (3.3) has a further implication for the zero-density region

$$\mathcal{N} := \{\, y \in \mathcal{Y} : f(x_0, y) = 0 \,\} = \{\, y \in \mathcal{Y} : p_{x_0}(y) = 0 \,\} \,.$$

Because the remainder in (3.3) is multiplicative in $f(x_0, y)$, the assumption implies that $f(x_0 + t, y) = 0$ for every $y \in \mathcal{N}$ and all sufficiently small t. It therefore excludes models in which y-values having zero density at x_0 acquire positive density as x moves locally away from x_0. QMD in Assumption 2 permits such local changes, requiring only that

$$\int_{\mathcal{N}} p_{x_0+t}(y)\, \nu(dy) = o(\|t\|^2) \,.$$

Thus, QMD controls the probability entering $\mathcal{N}$ in an integrated sense, whereas Assumption 3 rules it out pointwise. We examine the role of $\mathcal{N}$ under alternative smoothness conditions further in Appendix S.2.1.

In summary, the faster FHR rate reflects both symmetric interior geometry and a quadratic bound on the relative remainder. The FHR conditions also restrict local changes in the support of the conditional law. Our QMD framework does not impose these restrictions and accommodates boundary points, as arise in regression discontinuity designs.

4 Applications

Fix $\alpha \in (0, 1)$. A finite-sample guarantee in the ideal experiment continues to hold when the same procedure is applied to the nearest outcomes, up to the total variation distance between the two experiments. The four applications below use this transfer for rejection and coverage probabilities.

Section 4.3 also illustrates how additional structure can sharpen the generic coverage bound.

4.1 Covariate Balance Testing in Regression Discontinuity Designs

In a sharp regression discontinuity design, treatment is assigned when a real-valued running variable X crosses a known cutoff, normalized to zero. A standard specification check asks whether the distribution of a baseline covariate Y is continuous at the cutoff. Write

$$P_- := \mathcal{L}(Y \mid X = 0^-), \qquad P_+ := \mathcal{L}(Y \mid X = 0^+) \ .$$

Here P_- and P_+ denote the respective one-sided limits of the conditional law of Y at the cutoff. The null hypothesis studied by Canay and Kamat (2018) is

$$H_0 : P_- = P_+ \ . \tag{4.1}$$

This is the distributional analog of the continuity-of-means checks commonly reported in empirical applications of regression discontinuity designs.

In the ideal experiment, covariate balance is a standard two-sample problem. As before, S and S_n denote the pooled ideal and feasible samples, respectively. Let $S_- = (Y_1^-, \ldots, Y_q^-)$ and $S_+ = (Y_1^+, \ldots, Y_q^+)$ be independent, with laws P_-^q and P_+^q, respectively, and set $S := (S_-, S_+)$. Under H_0, the components of S are $2q$ i.i.d. draws from a common distribution, and hence S is exchangeable. Consider the Cramér–von Mises permutation test proposed by Canay and Kamat (2018), which compares the empirical distributions of the two blocks. Its critical values are computed by permuting the pooled sample. Randomization at the critical value gives rejection probability α in the ideal experiment; without this randomization, the rejection probability is at most α. Let ϕ denote either version of the test, so that $E[\phi(S)] \leq \alpha$.

When X is continuously distributed, the two ideal samples are unavailable. Canay and Kamat (2018) approximate them using the outcomes attached to the q observations closest to the cutoff on each side. Let $\iota_n^-(1), \ldots, \iota_n^-(q)$ denote the indices of the q observations with $X_i < 0$ closest to zero, listed in their original sample order. Define $\iota_n^+(1), \ldots, \iota_n^+(q)$ analogously using the observations with $X_i > 0$. The resulting feasible experiment is

$$S_n := \left(Y_{\iota_n^-(1)}, \ldots, Y_{\iota_n^-(q)}, Y_{\iota_n^+(1)}, \ldots, Y_{\iota_n^+(q)}\right) \ .$$

Applying the same permutation rule to S_n gives its feasible test, $\phi(S_n)$. Thus, the remaining question is how closely the law of S_n approximates the law of S.

To answer this question, let Q_- be the law of $(-X, Y)$ conditional on $X < 0$ and let Q_+ be the law of (X, Y) conditional on $X > 0$. For $u > 0$, let $P_{-,u}$ and $P_{+,u}$ denote versions of the conditional laws of Y under Q_- and Q_+, respectively, at covariate value u. Each block is a one-dimensional IOS experiment with conditioning point zero, a boundary point of its covariate support.

Corollary 1. *Suppose $Q\{X<0\}>0$ and $Q\{X>0\}>0$. Suppose also that $P_{-,u}$ and $P_{+,u}$ extend to zero with $P_{-,0}=P_-$ and $P_{+,0}=P_+$, and that Assumptions 1–2 hold at $x_0=0$ for Q_- and Q_+ with these extensions. If $q=o(n)$, Theorems 1–2 imply*

$$\mathrm{TV}(\mathcal{L}(S_n),\mathcal{L}(S))=O\big(q^{3/2}/n\big)\ .$$

Under H_0, it follows that

$$E_{Q^n}[\phi(S_n)]\le\alpha+O\big(q^{3/2}/n\big)\ . \tag{4.2}$$

Consequently, asymptotic size control holds if

$$q=o\big(n^{2/3}\big)\ . \tag{4.3}$$

The proof, given in Appendix A.3, applies the IOS approximation separately to the two one-sided samples and combines the bounds using the product inequality for total variation. The boundary-valid rate in Section 3 is needed because zero is a boundary point for each sample.

Remark 2 (Choice of q)**.** *The rule of thumb proposed by Canay and Kamat (2018) takes the form $q_{\mathrm{rot}}=C_{\mathrm{ck}}n^{0.9}/\log n$, where C_{ck} is the adjustment derived there under a bivariate normal model. The factor $n^{0.9}/\log n$ does not satisfy (4.3), so Corollary 1 does not cover the original rule. Since (4.3) restricts the growth rate rather than the multiplicative constant, we suggest retaining this adjustment and modifying the deterministic rate component as follows:*

$$q_{\mathrm{rot}}=C_{\mathrm{ck}}n^{\gamma},\qquad 0<\gamma<2/3\ .$$

■

4.2 Conditional Quantile Inference

Suppose that Y is scalar and fix $\tau\in(0,1)$. The parameter of interest is the τ-quantile of the conditional distribution of Y given $X=x_0$,

$$\theta_\tau(P):=\inf\{y:P\{Y\le y\}\ge\tau\}\ , \tag{4.4}$$

where P denotes the conditional law of Y given $X=x_0$.

Consider first the ideal experiment $S=(S_1,\dots,S_k)\sim P^k$. Let $S_{(1)}\le\cdots\le S_{(k)}$ denote its order statistics, and set $S_{(0)}=-\infty$ and $S_{(k+1)}=+\infty$. Choose integers $0\le\ell\le u\le k+1$ such that

$$b_{k,\tau}(\ell,u):=\sum_{j=\ell}^{u-1}\binom{k}{j}\tau^j(1-\tau)^{k-j}\ge1-\alpha\ . \tag{4.5}$$

Lemma 1 in Appendix A.3.2 gives

$$P^k\{\theta_\tau(P) \in [S_{(\ell)}, S_{(u)}]\} \geq b_{k,\tau}(\ell, u) \ .$$

This finite-sample guarantee does not require P to have a continuous CDF. Equality holds if it is continuous at $\theta_\tau(P)$; an atom at the target quantile may make the interval conservative.

The feasible interval replaces the ideal sample with the k nearest outcomes. Let $S_{n,(1)} \leq \cdots \leq S_{n,(k)}$ denote the order statistics of the components of S_n, with the same endpoint conventions, and define

$$C_{n,k} := [S_{n,(\ell)}, S_{n,(u)}] \ . \tag{4.6}$$

Corollary 2. *For every $Q \in \mathbf{Q}$, $n \in \mathbf{N}$, and $1 \leq k < n$,*

$$\begin{aligned} Q^n\{\theta_\tau(P) \in C_{n,k}\} &\geq b_{k,\tau}(\ell, u) - \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \\ &\geq 1 - \alpha - \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ . \end{aligned} \tag{4.7}$$

If, in addition, Assumptions 1–2 hold, Theorems 1–2 imply

$$Q^n\{\theta_\tau(P) \in C_{n,k}\} \geq 1 - \alpha - O\big(k^{1/2}(k/n)^{1/d}\big) \ . \tag{4.8}$$

Consequently, if $k = o(n^{2/(d+2)})$, then

$$\liminf_{n\to\infty} Q^n\{\theta_\tau(P) \in C_{n,k}\} \geq 1 - \alpha \ . \tag{4.9}$$

The bound in (4.7) is a finite-sample transfer statement and imposes no smoothness on the conditional law. Smoothness enters only when bounding the total variation distance. No quantile-process approximation is needed.

A closely related construction appears in Fan and Liu (2016, Example 2.1). For the asymmetric k-nearest-neighbor estimator, their direct approach also produces an interval with order-statistic endpoints. Fan and Liu choose those endpoints using an asymptotic Gaussian shift in the quantile index and establish first-order coverage through a quantile-process limit. Under the smooth-interior conditions used to verify that limit, they impose $k = o(n^{4/(d+4)})$, but do not report a rate for the coverage error.

Our interval instead uses finite-sample binomial calibration in the ideal experiment and the explicit transfer bound in (4.7). Under QMD, this bound accommodates boundary points and gives the sufficient condition $k = o(n^{2/(d+2)})$. Under Assumption 3, our general result also gives $k = o(n^{4/(d+4)})$, although through approximation of the joint IOS law rather than a quantile-process limit. Fan and Liu's conditions and QMD are not nested: theirs are tailored to the conditional-quantile process, whereas QMD controls the local conditional law.

4.3 Conditional Conformal Prediction

Conformal prediction constructs a prediction set by comparing a future outcome's nonconformity score with scores from a calibration sample. Let $(X_0, Y_0) \sim Q$ denote the covariates and outcome of the unit for which we seek a prediction. Standard conformal prediction provides a set $\widehat{C}(X_0)$ that covers Y_0 with probability at least $1-\alpha$. This marginal guarantee allows for systematic undercoverage at particular values of X_0. A growing literature on conditional conformal prediction instead targets coverage of at least $1-\alpha$ conditional on X_0. See Remark 3 for empirical settings that motivate this conditional coverage objective. Recent contributions to this literature include Guan (2023); Hore and Barber (2025); Min et al. (2026). Nontrivial, distribution-free conditional coverage cannot generally be guaranteed in finite samples without restrictions on the data-generating process (Barber et al., 2021); our approximation results provide a conditional coverage bound under local regularity conditions.

Let the calibration sample consist of n i.i.d. observations from Q. A separate training sample $\mathcal{D}_{\text{tr}}$ may be used to construct a measurable nonconformity score conditional on $X_0 = x_0$, $v_{x_0} : \mathbf{R}^m \to \mathbf{R}$ (for example, $v_{x_0}(y) = |y - \widehat{\mu}(x_0)|$ when Y is scalar). The two samples and the new observation (X_0, Y_0) are mutually independent, so that, conditional on $X_0 = x_0$, the calibration sample has law Q^n and $Y_0 \sim P$ is independent of both samples. We condition throughout on $\mathcal{D}_{\text{tr}}$, so the score is fixed. For a prediction set $\widehat{C}(x_0)$ constructed from the calibration sample, the target is

$$(Q^n \otimes P)\{Y_0 \in \widehat{C}(x_0)\} \geq 1 - \alpha \ , \tag{4.10}$$

where $\alpha \in (0,1)$. This probability averages over the calibration sample and the future outcome, given $X_0 = x_0$ and $\mathcal{D}_{\text{tr}}$. The rates below are pointwise in x_0.

Remark 3 (Prediction at an observed covariate value)**.** *Suppose a hospital wants to predict a patient's length of stay Y_0 using the patient's age and initial clinical measurements X_0. At admission, the hospital observes $X_0 = x_0$ and constructs a prediction interval for Y_0 at that value. An interval that achieves $1-\alpha$ coverage across all admissions may have lower coverage among patients with age and clinical measurements x_0. The target in (4.10) requires coverage of at least $1-\alpha$ for these patients. Other examples include predicting loan repayment given a borrower's characteristics at application and predicting subsequent earnings given a participant's characteristics at entry into a job-training program.* ■

Consider first the ideal experiment $S = (S_1, \ldots, S_k) \sim P^k$, independent of Y_0. Let $T_j = v_{x_0}(S_j)$, write $T_{(1)} \leq \cdots \leq T_{(k)}$ for their order statistics, and set $T_{(k+1)} = +\infty$. Define

$$j_\alpha := \lceil (k+1)(1-\alpha) \rceil \ , \qquad \widehat{C}_k(x_0) := \{y : v_{x_0}(y) \leq T_{(j_\alpha)}\} \ .$$

The same fixed score is applied to $k+1$ i.i.d. outcomes, so the scores are exchangeable. The

standard conformal rank argument therefore gives

$$(P^k \otimes P)\{Y_0 \in \widehat{C}_k(x_0)\} \geq 1 - \alpha \ .$$

The feasible procedure uses S_n, the outcomes attached to the k calibration covariates closest to x_0. Let $T_{n,j} = v_{x_0}(S_{n,j})$, denote their order statistics by $T_{n,(1)} \leq \cdots \leq T_{n,(k)}$, and set $T_{n,(k+1)} = +\infty$. The prediction set is

$$\widehat{C}_{n,k}(x_0) := \{y : v_{x_0}(y) \leq T_{n,(j_\alpha)}\} \ . \tag{4.11}$$

The cutoff $T_{n,(j_\alpha)}$ is finite only if $k \geq (1-\alpha)/\alpha$; otherwise the set is all of $\mathbf{R}^m$. At $\alpha = 0.05$, this requires at least 19 neighbors. For the absolute-residual score, this is an interval centered at $\widehat{\mu}(x_0)$ with radius $T_{n,(j_\alpha)}$. Every selected calibration outcome is scored at the same target value x_0. Using $|Y_i - \widehat{\mu}(X_i)|$ instead requires additional regularity of the covariate-dependent score law.

Corollary 3. *For the fixed measurable score v_{x_0}, every $Q \in \mathbf{Q}$, $n \in \mathbf{N}$, and $1 \leq k < n$,*

$$(Q^n \otimes P)\{Y_0 \in \widehat{C}_{n,k}(x_0)\} \geq 1 - \alpha - \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ . \tag{4.12}$$

If, in addition, Assumptions 1–2 hold, Theorems 1–2 imply

$$(Q^n \otimes P)\{Y_0 \in \widehat{C}_{n,k}(x_0)\} \geq 1 - \alpha - O\big(k^{1/2}(k/n)^{1/d}\big) \ . \tag{4.13}$$

Consequently, if $k = o(n^{2/(d+2)})$, the limit inferior of the coverage probability is at least $1 - \alpha$.

The first bound requires no smoothness beyond the setup. Theorems 1–2 then give (4.13) directly, without approximating the score quantile. The proof is in Appendix A.3.3. The guarantee holds for every fixed measurable score, permits ties, and requires no accuracy condition on the fitted prediction rule; the score still affects the size and shape of the set. If Assumptions 1–2 hold at Q_X-almost every x_0, where Q_X denotes the marginal law of X under Q, the preceding bounds give coverage conditional on X_0 almost surely, with constants that may depend on its value.

The set in (4.11) is a nearest-neighbor version of localized split conformal prediction, with equal weights and a common target-point score; see Guan (2023) and Hore and Barber (2025, Section 2.1) for the local weighting construction. Our result gives a one-sided conditional coverage guarantee at x_0 under local conditions on the outcome law, including at boundary points, for any fixed measurable target-point score. It follows from the joint transfer bound without a separate approximation of the score quantile.

Remark 4 (A sharper bound for prediction)**.** *Our blueprint delivers rates automatically once the ideal guarantee and approximation conditions are established. Additional structure may yield faster rates, at the cost of further analysis. For prediction, such an argument removes the factor $k^{1/2}$ from the corollary's coverage bound: instead of approximating the joint law of all k calibration outcomes, it suffices to compare the law of a single future outcome.*

To see this, condition on $R_{(k+1)} = r$. Up to ordering, which does not affect the prediction set, the selected calibration outcomes are then i.i.d. from P_r. If the future outcome also had law P_r and were independent of these outcomes, the conformal rank argument would give coverage at least $1 - \alpha$. Its actual law is P, so the coverage loss is at most $\mathrm{TV}(P_r, P)$. Under Assumptions 1–2, averaging over the radius gives

$$(Q^n \otimes P)\{Y_0 \in \widehat{C}_{n,k}(x_0)\} \geq 1 - \alpha - O\big((k/n)^{1/d}\big) \ . \tag{4.14}$$

Thus, $k/n \to 0$ suffices for the lower coverage guarantee. This single-observation comparison does not apply to our other applications: changing the local law changes the quantile target, while the two testing problems need not have a common local law under the null. The proof is in Appendix A.3.3; a related comparison appears in Hore and Barber (2025, online supplementary Appendix A.3). ■

4.4 Policy-Irrelevance Testing under Network Interference

Auerbach et al. (2026) propose a nonparametric framework for causal inference with network data. Their model allows an agent's outcome to depend on the agent's treatment status, the agent's connections with other individuals in the community, and the treatment statuses of other agents in the network. The network and treatment information from the agent's perspective is called a "rooted network," which we denote by G. Under their assumptions, a rooted network represents a policy of interest. Fix two target rooted networks, $\mathtt{g}_0$ and $\mathtt{g}_1$, and for $a \in \{0, 1\}$ let P_a denote the distribution of the outcome under $\mathtt{g}_a$. The null hypothesis of policy irrelevance is

$$H_0 : P_0 = P_1 \ . \tag{4.15}$$

In the ideal experiment, policy irrelevance is again a two-sample problem. Throughout this subsection, S and S_n denote the pooled ideal and feasible samples, respectively. Let $S(0)$ and $S(1)$ be independent, with laws P_0^k and P_1^k, respectively, and set $S := (S(0), S(1))$. Under H_0, the components of S are $2k$ i.i.d. draws from a common distribution, and hence S is exchangeable. Let ϕ denote the permutation test of Auerbach et al. (2026), whose critical values are computed by permuting the pooled sample, so that $E[\phi(S)] \leq \alpha$.

The data consist of an i.i.d. sample of n communities, each composed of a network of individuals. We implement the procedure in Auerbach et al. (2026) using a prespecified split of the communities into two halves of size $n/2$, assuming for simplicity that n is even; the first half is used for $\mathtt{g}_0$ and the second for $\mathtt{g}_1$. In each community of the half used for $\mathtt{g}_a$, select the individual whose rooted network is closest to $\mathtt{g}_a$ according to the distance ρ in Auerbach et al. (2026, Eq. (1)), breaking ties within the community by independent uniform randomization, as in their Algorithm 1. Let X^a denote the distance between the selected individual's rooted network and $\mathtt{g}_a$, and let Y^a denote the selected individual's outcome. This produces $n/2$ "community representatives." Let $S_n(a)$ collect the outcomes of the k representatives with the smallest values of X^a, listed in their original

sample order. The resulting feasible experiment is $S_n := (S_n(0), S_n(1))$, and applying the same permutation rule to S_n gives the feasible test $\phi(S_n)$. Here Q denotes the law of a community together with its tie-breaking draws.

To connect this construction with our framework, let Q_a be the law of the representative pair (X^a, Y^a) for $a \in \{0,1\}$. Within each half, the representative pairs are i.i.d. draws from Q_a, and the two halves are independent. Hence $S_n(a)$ is the IOS vector attached to the k smallest values of X^a in a sample of size $n/2$ from Q_a. For $r > 0$ with positive marginal density under Q_a, let $P_{a,r}$ denote the conditional law of Y^a given $X^a = r$, defined through the densities as in Section 2.1. Since distances are nonnegative, each half is a one-dimensional IOS experiment with conditioning point zero, a boundary point of its covariate support.

Corollary 4. *Suppose that, for each $a \in \{0,1\}$, Q_a satisfies Assumptions 1–2 at zero with P_a as the conditional law of Y^a at $X^a = 0$. Then, uniformly over even $n \in \mathbf{N}$ and $1 \le k < n/2$, Theorems 1–2 imply*

$$\mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{3/2}/n\big) \ . \tag{4.16}$$

Under H_0, it follows that

$$E_{Q^n}[\phi(S_n)] \le \alpha + O\big(k^{3/2}/n\big) \ . \tag{4.17}$$

Consequently, asymptotic size control holds if $k = o(n^{2/3})$.

The proof, given in Appendix A.3, applies the IOS approximation separately to the two halves and combines the bounds using the product inequality for total variation. As in Corollary 1, requiring P_a to be the conditional law at zero identifies the limiting law of the selected outcome with the target law, and the common 2/3 exponent reflects that both applications reduce to a one-dimensional boundary experiment under QMD.

Auerbach et al. (2026, Theorem 4) establish asymptotic validity for fixed k under their Assumptions 1, 5, and 7. Their simulation evidence suggests the diverging choice $k = 2\lfloor \log n \rfloor$ in our notation (p. 192). This choice satisfies our growth condition and yields a size bound of $\alpha + O((\log n)^{3/2}/n)$ under our assumptions. Our result does not cover the data-dependent alternating partition they recommend in practice because that partition need not produce the independent subsamples used in our proof.

5 Concluding Remarks

The paper provides a three-step blueprint for inference with nearest outcomes. First, establish a finite-sample guarantee in an ideal experiment consisting of independent draws from the conditional law at the target point. Second, apply the same procedure to the outcomes attached to the nearest observed covariates. Third, bound the total variation distance between the two experiments. The ideal guarantee then carries over to the feasible procedure, with an error no larger than this distance.

Our approximation results give sufficient restrictions on how fast the number of selected neighbors may grow with the sample size. Under quadratic mean differentiability and local regularity of the covariate distribution, the local Hellinger and total variation errors are at most linear in the neighborhood radius, including at boundary points. Theorem 2 also establishes the limits of these rates. A boundary example attains the linear rate in both distances. At interior points, both the Hellinger and total variation errors are smaller than linear for every law satisfying the assumptions. Nevertheless, a single such interior model rules out every positive polynomial improvement on the linear rate in either distance. Interior location alone therefore does not deliver a faster polynomial rate under these conditions.

The four applications combine the same transfer argument with different finite-sample guarantees: permutation validity, binomial calibration, and conformal ranks. Controlling the joint law of the nearest outcomes allows the approximation to be used across procedures, without a separate analysis of each statistic. Additional structure can yield sharper bounds, as the prediction application illustrates, at the cost of further analysis.

Appendix A Proof of the main results

For probability measures M and N with densities p_M and p_N relative to a common dominating measure λ, we use the normalization

$$\mathrm{H}^2(M,N) := \frac{1}{2}\int (\sqrt{p_M} - \sqrt{p_N})^2 \, d\lambda, \qquad \mathrm{TV}(M,N) := \frac{1}{2}\int |p_M - p_N| \, d\lambda \ .$$

A.1 Proof of Theorem 1

Part (a). Construct $S \sim P^k$ independently of the observed sample. Let r_1 be as in Lemma 2, reduced if necessary so that the assumed local Hellinger and total variation bounds hold for $0 < r \leq r_1$. All constants below may depend on the fixed law and the exponents, but not on n or k. Lemma 2 implies that $F(r) \geq F(r/2) > 0$ for all $r \in (0, r_1]$. Then, Kaufmann and Reiss (1992, Theorem 1) implies that, for $\mathcal{L}(R_{(k+1)})$-almost every $r \in (0, r_1)$,

$$\mathcal{L}(S_n | R_{(k+1)} = r) = P_r^k \ , \tag{A-1}$$

where P_r^k denotes the k-fold product measure of P_r.

Consider the following derivation for any $n \in \mathbf{N}$ and $1 \leq k < n$,

$$\begin{aligned}
\mathrm{H}^2(\mathcal{L}(S_n), \mathcal{L}(S)) &\overset{(1)}{\leq} \int_0^\infty \mathrm{H}^2(\mathcal{L}(S_n|R_{(k+1)} = r), \mathcal{L}(S|R_{(k+1)} = r)) dP_{R_{(k+1)}}(r) \\
&\overset{(2)}{=} \int_0^{r_1} \mathrm{H}^2(P_r^k, P^k) dP_{R_{(k+1)}}(r) + \int_{r_1}^\infty \mathrm{H}^2(\mathcal{L}(S_n|R_{(k+1)} = r), \mathcal{L}(S|R_{(k+1)} = r)) dP_{R_{(k+1)}}(r)
\end{aligned}$$

$$\overset{(3)}{\leq} k \int_0^{r_1} \mathrm{H}^2(P_r, P) dP_{R_{(k+1)}}(r) + P(R_{(k+1)} \geq r_1) \ , \tag{A-2}$$

where (1) holds by joint convexity of the squared Hellinger distance (e.g., Reiss (1993, Lemma 3.1.3)), applied to $\mathcal{L}(S_n) = \int \mathcal{L}(S_n | R_{(k+1)} = r)\, dP_{R_{(k+1)}}(r)$ and $\mathcal{L}(S) = \int \mathcal{L}(S | R_{(k+1)} = r)\, dP_{R_{(k+1)}}(r)$, (2) by (A-1), $S \perp R_{(k+1)}$, and $S \sim P^k$, and (3) by $\mathrm{H}^2(P', P'') \leq 1$ and Bernoulli's inequality, which states that $1 - (1-x)^k \leq kx$ for any $x \in [0,1]$ and $k \in \mathbf{N}$.

Given (A-2), we complete the proof by finding constants $C_1, C_2 \in (0, \infty)$ such that, for all $n \in \mathbf{N}$ and $1 \leq k < n$,

$$\int_0^{r_1} \mathrm{H}^2(P_r, P) dP_{R_{(k+1)}}(r) \leq C_1 (k/n)^{2a_h/d} \tag{A-3}$$

$$P(R_{(k+1)} \geq r_1) \leq C_2 k (k/n)^{2a_h/d} \ . \tag{A-4}$$

We begin by proving (A-3). By $\mathrm{H}(P_r, P) = O(r^{a_h})$, there is $C_3 < \infty$ such that

$$\mathrm{H}^2(P_r, P) \leq C_3 r^{2a_h} \qquad \text{for all } 0 < r \leq r_1 \ . \tag{A-5}$$

Set $\beta := 2a_h/d$. If $\beta = 0$, (A-3) and (A-4) follow immediately from $\mathrm{H}^2(P_r, P) \leq 1$ and $P\{R_{(k+1)} \geq r_1\} \leq 1 \leq k$. Suppose, therefore, that $\beta > 0$. For every $u \in (0,1)$,

$$\big(F^{-1}(u)\big)^{2a_h} I\{F^{-1}(u) \leq r_1\} \leq C_L^{-\beta} u^{\beta} \ . \tag{A-6}$$

To see this, if $u \leq C_L r_1^d$, Lemma 2 gives $F^{-1}(u) \leq (u/C_L)^{1/d}$ and so $(F^{-1}(u))^{2a_h} \leq C_L^{-\beta} u^{\beta}$. If $u > C_L r_1^d$ and the indicator in (A-6) equals one, then $F^{-1}(u) \leq r_1 < (u/C_L)^{1/d}$; if the indicator equals zero, the inequality is immediate. It follows that

$$\begin{aligned}
\int_0^{r_1} \mathrm{H}^2(P_r, P)\, dP_{R_{(k+1)}}(r) &\leq C_3 E\left[R_{(k+1)}^{2a_h} I\{R_{(k+1)} \leq r_1\}\right] \\
&= C_3 E\left[\big(F^{-1}(U_{(k+1:n)})\big)^{2a_h} I\{F^{-1}(U_{(k+1:n)}) \leq r_1\}\right] \\
&\leq C_3 C_L^{-\beta} E\left[U_{(k+1:n)}^{\beta}\right] \\
&\leq C_4 (k/n)^{\beta} \ .
\end{aligned}$$

The last inequality follows from Lemma S.4, applied to the $(k+1)$st order statistic, and $k+1 \leq 2k$. This proves (A-3).

We now prove (A-4). Let $r_\star := r_1/2$ and $p_\star := F(r_\star) > 0$, where positivity follows from Lemma 2. Define

$$N_\star := \sum_{i=1}^{n} I\{R_i < r_\star\} \ .$$

Then $N_\star \sim \text{Binomial}(n, p_\star)$ and

$$\{R_{(k+1)} \geq r_1\} \subseteq \{R_{(k+1)} \geq r_\star\} = \{N_\star \leq k\} \ .$$

If $k/n \leq p_\star/2$, the Chernoff bound for the lower tail of a binomial random variable gives

$$P\{R_{(k+1)} \geq r_1\} \leq P\{N_\star \leq k\} \leq P\{N_\star \leq np_\star/2\} \leq \exp(-np_\star/8) \leq C_5 n^{-\beta} \leq C_5 k(k/n)^\beta \ , \quad \text{(A-7)}$$

for some $C_5 < \infty$ which does not depend on n or k, as exponential decay dominates polynomial decay. The last inequality uses $k \geq 1$. If $k/n > p_\star/2$, then

$$P\{R_{(k+1)} \geq r_1\} \leq 1 \leq \left(\frac{2}{p_\star}\right)^\beta (k/n)^\beta \leq \left(\frac{2}{p_\star}\right)^\beta k(k/n)^\beta \ . \quad \text{(A-8)}$$

Combining the two cases proves (A-4) with $C_2 := \max\{C_5, (2/p_\star)^\beta\}$. Combining (A-2)–(A-4), including the case $\beta = 0$, gives

$$\text{H}^2(\mathcal{L}(S_n), \mathcal{L}(S)) \leq (C_1 + C_2)k(k/n)^{2a_h/d} \ .$$

Taking square roots proves part (a).

Part (b). By part (a) and (2.7), we obtain that, for any $n \in \mathbf{N}$ and $1 \leq k < n$, $\text{TV}(\mathcal{L}(S_n), \mathcal{L}(S)) = O(k^{1/2}(k/n)^{a_h/d})$. To complete the proof, it then suffices to show that, for any $n \in \mathbf{N}$ and $1 \leq k < n$, we also have $\text{TV}(\mathcal{L}(S_n), \mathcal{L}(S)) = O(k(k/n)^{a_{tv}/d})$.

The argument used to derive (A-1) still applies. Thus, for any $n \in \mathbf{N}$ and $1 \leq k < n$, we have

$$\begin{aligned}
\text{TV}(\mathcal{L}(S_n), \mathcal{L}(S)) &\overset{(1)}{=} \sup_{B \in \mathcal{B}} \left| \int_0^\infty \left(P\left(S_n \in B | R_{(k+1)} = r\right) - P\left(S \in B | R_{(k+1)} = r\right) \right) dP_{R_{(k+1)}}(r) \right| \\
&\leq \int_0^\infty \sup_{B \in \mathcal{B}} \left| P\left(S_n \in B | R_{(k+1)} = r\right) - P\left(S \in B | R_{(k+1)} = r\right) \right| dP_{R_{(k+1)}}(r) \\
&= \int_0^\infty \text{TV}(\mathcal{L}(S_n | R_{(k+1)} = r), \mathcal{L}(S | R_{(k+1)} = r)) dP_{R_{(k+1)}}(r) \\
&\overset{(2)}{=} \int_0^{r_1} \text{TV}(P_r^k, P^k) dP_{R_{(k+1)}}(r) + \int_{r_1}^\infty \text{TV}(\mathcal{L}(S_n | R_{(k+1)} = r), \mathcal{L}(S | R_{(k+1)} = r)) dP_{R_{(k+1)}}(r) \\
&\overset{(3)}{\leq} k \int_0^{r_1} \text{TV}(P_r, P) dP_{R_{(k+1)}}(r) + P(R_{(k+1)} \geq r_1) \ ,
\end{aligned} \quad \text{(A-9)}$$

where (1) holds by the law of total probability and using $\mathcal{B}$ to denote the common σ-algebra to $\mathcal{L}(S_n)$ and $\mathcal{L}(S)$, and (2) by (A-1), $S \perp R_{(k+1)}$, and $S \sim P^k$, and (3) by Hoeffding and Wolfowitz (1958, Eq. (4.5)) and $\text{TV} \leq 1$. Given (A-9), we complete the proof by finding constants $C_1', C_2' \in (0, \infty)$ such that, for all $n \in \mathbf{N}$ and $1 \leq k < n$,

$$\int_0^{r_1} \text{TV}(P_r, P) dP_{R_{(k+1)}}(r) \leq C_1'(k/n)^{a_{tv}/d} \quad \text{(A-10)}$$

$$P(R_{(k+1)} \geq r_1) \leq C_2' k(k/n)^{a_{tv}/d} \ . \tag{A-11}$$

The bounds in (A-10) and (A-11) follow by repeating the preceding argument with $\mathrm{H}(P_r, P) = O(r^{a_h})$ replaced by $\mathrm{TV}(P_r, P) = O(r^{a_{tv}})$ and $\beta := a_{tv}/d$. Combining this total variation bound with the bound from part (a), and taking the larger of their constants, proves the stated minimum.

A.2 Proof of Theorem 2

Let r_1 be as in Lemma 2. For $x \in \mathcal{X}$ sufficiently close to x_0, the density $g(x)$ is positive, so p_x and $D(x)$ are defined. Lemma 3 with Cauchy-Schwarz inequality give, as $x \to x_0$ through these points,

$$\begin{aligned} D(x) &= \tfrac{1}{4} \int ((x - x_0)^\top \dot{\ell}_{x_0}(y))^2 p_{x_0}(y) \nu(dy) + o(\|x - x_0\|^2) \\ &\leq \tfrac{1}{4} \|x - x_0\|^2 \int \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y) \nu(dy) + o(\|x - x_0\|^2) \ . \end{aligned} \tag{A-12}$$

Consequently, there are $C < \infty$ and $\bar{r} > 0$ such that, for every $x \in \mathcal{X}$ with $\|x - x_0\| < \bar{r}$,

$$D(x) \leq C \|x - x_0\|^2 \ . \tag{A-13}$$

By Lemma 2, for $0 < r \leq r_1$,

$$F(r) = \int_{B_r \cap \mathcal{X}} g(x)\, dx \geq C_L r^d > 0 \ . \tag{A-14}$$

Thus the probability measure

$$\mu_r(dx) := \frac{g(x) I\{x \in B_r \cap \mathcal{X}\}}{F(r)}\, dx \ , \tag{A-15}$$

satisfies $P_r = \int P_x \, \mu_r(dx)$. Let $r_2 := \min\{r_1, \bar{r}\}$. For any $r \in (0, r_2)$, consider the following derivation.

$$\begin{aligned} \mathrm{H}^2(P_r, P) &\overset{(1)}{\leq} \int \mathrm{H}^2(P_x, P) \mu_r(dx) \\ &\overset{(2)}{=} \frac{1}{2F(r)} \int_{B_r \cap \mathcal{X}} D(x) g(x) dx \\ &\overset{(3)}{\leq} \frac{C}{2F(r)} \int_{B_r \cap \mathcal{X}} \|x - x_0\|^2 g(x) dx \\ &\overset{(4)}{\leq} \frac{C}{2C_L r^d} (g(x_0) + C_g r) \int_{B_r \cap \mathcal{X}} \|x - x_0\|^2 dx \\ &\overset{(5)}{=} O(r^2) \ , \end{aligned} \tag{A-16}$$

where (1) holds by the convexity lemma (e.g., Reiss (1993, Lemma 3.1.3)), with measure $\mu_r(dx)$ and $P_r = \int P_x \mu_r(dx)$, (2) by $\mathrm{H}^2(P_x, P) = D(x)/2$ and (A-15), (3) by (A-13), (4) by (A-14) and Assumption 1, and (5) by $\int_{B_r \cap \mathcal{X}} \|x - x_0\|^2 dx = O(r^{d+2})$. From (A-16), we obtain

$$\mathrm{H}(P_r, P) = O(r) \ .$$

The convergence rate of the total variation distance follows from its relationship with the Hellinger distance, $\mathrm{TV}(P_r, P) \leq \sqrt{2}\mathrm{H}(P_r, P)$. This completes the proof of (3.1).

Part (i) of the Theorem follows from Lemma S.1. The first statement in Part (ii) follows from Lemma S.3, and the second follows from Lemma S.2. This completes the proof. ■

Proof of Remark 1. Fix Q and a sequence $k = k_n$ as in the remark. Let r_1 be as in Lemma 2. Given $\epsilon > 0$, Theorem 2(ii) gives $\delta \in (0, r_1)$ such that $\mathrm{H}^2(P_r, P) \leq \epsilon r^2$ for $0 < r < \delta$. Applying the argument in (A-2) with δ in place of r_1 yields

$$\begin{aligned} \mathrm{H}^2\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) &\leq \epsilon k E_{Q^n}\Big[R_{(k+1)}^2 I\{R_{(k+1)} < \delta\}\Big] + Q^n\{R_{(k+1)} \geq \delta\} \\ &\leq \epsilon C k (k/n)^{2/d} + Q^n\{R_{(k+1)} \geq \delta\} \ . \end{aligned}$$

The second inequality follows from (A-6) with $a_h = 1$ and Lemma S.4, since $\delta < r_1$. The constant C does not depend on ϵ, δ, n, or k.

Let $p_\delta := F(\delta/2) > 0$. Since $k/n \to 0$, eventually $k/n \leq p_\delta/2$. The binomial argument in (A-7) then gives

$$Q^n\{R_{(k+1)} \geq \delta\} \leq \exp(-np_\delta/8) \ .$$

Since $k \geq 1$, we have $k(k/n)^{2/d} \geq n^{-2/d}$, and hence

$$\frac{Q^n\{R_{(k+1)} \geq \delta\}}{k(k/n)^{2/d}} \leq n^{2/d} \exp(-np_\delta/8) \to 0 \ .$$

It follows that

$$\limsup_{n\to\infty} \frac{\mathrm{H}^2\big(\mathcal{L}(S_n), \mathcal{L}(S)\big)}{k(k/n)^{2/d}} \leq \epsilon C \ .$$

Letting $\epsilon \downarrow 0$ proves the Hellinger claim. The total variation claim follows from (2.7), and $k = O(n^{2/(d+2)})$ makes $k^{1/2}(k/n)^{1/d}$ bounded. ■

A.3 Proofs for the Applications

A.3.1 Covariate balance testing in regression discontinuity designs

Proof of Corollary 1. Let

$$N_{n,-} := \sum_{i=1}^{n} 1\{X_i < 0\}, \qquad N_{n,+} := \sum_{i=1}^{n} 1\{X_i > 0\},$$

and write $\pi_- := Q\{X < 0\}$ and $\pi_+ := Q\{X > 0\}$. Let $\mathcal{G}_n$ be the σ-field generated by the signs of $X_1, \ldots, X_n$. Conditional on $\mathcal{G}_n$, the observations on the two sides are independent samples from Q_- and Q_+. On the event

$$\mathcal{E}_n := \{N_{n,-} \geq \pi_- n/2,\ N_{n,+} \geq \pi_+ n/2\},$$

we have $q < N_{n,-} \wedge N_{n,+}$ for all sufficiently large n. Use the selected vector whenever both sides contain at least q observations, and define S_n to be a fixed vector otherwise. Since $q = o(n)$, the two conditional applications of Theorem 1 below are valid on $\mathcal{E}_n$ for all sufficiently large n. Theorems 1–2, applied conditionally and separately to the two sides, and the product inequality for total variation give

$$\mathrm{TV}\big(\mathcal{L}(S_n \mid \mathcal{G}_n), P_-^q \otimes P_+^q\big) \leq C_- \frac{q^{3/2}}{N_{n,-}} + C_+ \frac{q^{3/2}}{N_{n,+}} \leq \left(\frac{2C_-}{\pi_-} + \frac{2C_+}{\pi_+}\right) \frac{q^{3/2}}{n} = O(q^{3/2}/n)\ ,$$

uniformly on $\mathcal{E}_n$. Binomial concentration implies that $Q^n\{\mathcal{E}_n^c\} \leq 2\exp(-cn)$ for some $c > 0$. Convexity of total variation over the conditional laws therefore yields

$$\mathrm{TV}\big(\mathcal{L}(S_n), P_-^q \otimes P_+^q\big) \leq O(q^{3/2}/n) + 2\exp(-cn) = O(q^{3/2}/n)\ .$$

Under H_0, the $2q$ pooled ideal outcomes are i.i.d. from a common law and hence exchangeable. The ideal permutation rule therefore has rejection probability at most α. The variational characterization of total variation gives

$$E_{Q^n}[\phi(S_n)] \leq \alpha + O(q^{3/2}/n)\ ,$$

which proves (4.2). Finally, $q = o(n^{2/3})$ makes the approximation error vanish. ■

A.3.2 Conditional quantile inference

The next lemma gives the ideal-experiment coverage statement without imposing continuity of P. Let $S_1, \ldots, S_k$ be i.i.d. with common law P and CDF F, write $S_{(1)} \leq \cdots \leq S_{(k)}$ for their order statistics, and set $S_{(0)} = -\infty$ and $S_{(k+1)} = +\infty$.

Lemma 1. *Fix $\tau \in (0,1)$ and let $\theta_\tau := \inf\{y : F(y) \geq \tau\}$. For integers $0 \leq \ell \leq u \leq k+1$,*

$$P^k\{\theta_\tau \in [S_{(\ell)}, S_{(u)}]\} \geq b_{k,\tau}(\ell, u) := P\{\ell \leq B_{k,\tau} \leq u-1\}, \tag{A-17}$$

where $B_{k,\tau} \sim \text{Binomial}(k,\tau)$. If F is continuous at θ_τ, equality holds in (A-17).

Proof. Define

$$\tau_- := F(\theta_\tau -) = P\{S_i < \theta_\tau\}, \qquad \tau_+ := F(\theta_\tau) = P\{S_i \leq \theta_\tau\} \ .$$

The definition of θ_τ implies $\tau_- \leq \tau \leq \tau_+$. Let

$$N_< := \sum_{i=1}^k 1\{S_i < \theta_\tau\}, \qquad N_\leq := \sum_{i=1}^k 1\{S_i \leq \theta_\tau\} \ .$$

The interval covers θ_τ if and only if

$$\{\theta_\tau \in [S_{(\ell)}, S_{(u)}]\} = \{N_\leq \geq \ell, \ N_< \leq u-1\} \ . \tag{A-18}$$

Since $N_\leq \sim \text{Binomial}(k, \tau_+)$ and $N_< \sim \text{Binomial}(k, \tau_-)$, stochastic ordering and the union bound give

$$\begin{aligned} P^k\{\theta_\tau \notin [S_{(\ell)}, S_{(u)}]\} &\leq P^k\{N_\leq \leq \ell - 1\} + P^k\{N_< \geq u\} \\ &\leq \sum_{j=0}^{\ell-1} \binom{k}{j} \tau^j (1-\tau)^{k-j} + \sum_{j=u}^{k} \binom{k}{j} \tau^j (1-\tau)^{k-j} \\ &= 1 - b_{k,\tau}(\ell, u) \ . \end{aligned} \tag{A-19}$$

Taking complements proves (A-17). If F is continuous at θ_τ, then $\tau_- = \tau_+ = \tau$. In this case $N_< = N_\leq$ almost surely, this common count has the Binomial(k,τ) distribution, and (A-18) gives equality. ■

Proof of Corollary 2. Let

$$\mathcal{A}_\tau := \{s \in \mathbf{R}^k : s_{(\ell)} \leq \theta_\tau(P) \leq s_{(u)}\} \ ,$$

where the conventions for the zeroth and $(k+1)$st order statistics are as above. The variational characterization of total variation implies

$$Q^n\{S_n \in \mathcal{A}_\tau\} \geq P^k\{S \in \mathcal{A}_\tau\} - \text{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) \ . \tag{A-20}$$

Lemma 1 and (4.5) bound the first term on the right-hand side from below by $b_{k,\tau}(\ell, u) \geq 1 - \alpha$, which proves (4.7). Under Assumptions 1–2, Theorems 1–2 give

$$\text{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{1/d}\big) \ .$$

Substituting this bound into (4.7) proves (4.8). Finally, $k^{1/2}(k/n)^{1/d} = o(1)$ if and only if $k = o(n^{2/(d+2)})$, which gives (4.9). ■

A.3.3 Conditional conformal prediction

Throughout this subsection, we condition on $X_0 = x_0$ and the training sample $\mathcal{D}_{\mathrm{tr}}$, so that the calibration sample has law Q^n, the independent future outcome has law P, and the score v_{x_0} is fixed.

Proof of Corollary 3. Fix the score v_{x_0}. For $s = (s_1, \ldots, s_k)$, let $t_{(j)}(s)$ denote the jth order statistic of $v_{x_0}(s_1), \ldots, v_{x_0}(s_k)$, with $t_{(k+1)}(s) = +\infty$, and define

$$A := \Big\{(s, y) \in (\mathbf{R}^m)^k \times \mathbf{R}^m : v_{x_0}(y) \le t_{(j_\alpha)}(s)\Big\} \ .$$

For any probability law G on $\mathbf{R}^m$, consider $k+1$ i.i.d. draws from G and apply the same score v_{x_0} to each. Break score ties using independent $\mathrm{Unif}(0,1)$ variables. The randomized rank of the $(k+1)$st score is then uniform on $\{1, \ldots, k+1\}$, and rank at most j_α implies membership in A. Hence,

$$(G^k \otimes G)(A) \ge \frac{j_\alpha}{k+1} \ge 1 - \alpha \ . \tag{A-21}$$

This argument allows ties. If $j_\alpha = k+1$, then A is the entire sample space, which gives the small-k qualification following Corollary 3.

Since $Y_0 \sim P$ is independent of both S_n and the ideal sample $S \sim P^k$, the feasible and ideal coverage probabilities are $(\mathcal{L}(S_n) \otimes P)(A)$ and $(P^k \otimes P)(A)$, respectively. By the variational characterization of total variation and the fact that tensoring with a common probability measure preserves total variation,

$$\begin{aligned}(\mathcal{L}(S_n) \otimes P)(A) &\ge (P^k \otimes P)(A) - \mathrm{TV}\big(\mathcal{L}(S_n) \otimes P, P^k \otimes P\big) \\ &= (P^k \otimes P)(A) - \mathrm{TV}\big(\mathcal{L}(S_n), P^k\big) \\ &\ge 1 - \alpha - \mathrm{TV}\big(\mathcal{L}(S_n), P^k\big) \ ,\end{aligned}$$

where the last inequality follows from (A-21) with $G = P$. This proves (4.12). Under Assumptions 1–2, Theorems 1–2 give

$$\mathrm{TV}\big(\mathcal{L}(S_n), P^k\big) = O\big(k^{1/2}(k/n)^{1/d}\big) \ .$$

Substituting this bound into (4.12) proves (4.13). Finally, $k = o\big(n^{2/(d+2)}\big)$ implies $k^{1/2}(k/n)^{1/d} \to 0$. ■

Proof of Remark 4. Use the set A from the preceding proof, and let r_1 be as in Lemma 2, reduced if necessary so that $\mathrm{TV}(P, P_r) \le Cr$ for all $0 < r \le r_1$, as guaranteed by Theorem 2. By Lemma 2

and monotonicity of F, $F(r) > 0$ for every $r > 0$, so Kaufmann and Reiss (1992, Theorem 1) gives the representation in (A-1) for $\mathcal{L}(R_{(k+1)})$-almost every $r > 0$, not only for $r < r_1$. By (A-1), conditional on $R_{(k+1)} = r$, the calibration sample has law P_r^k, while the independent future outcome Y_0 has law P. Hence, for every such r, the conditional coverage probability is $(P_r^k \otimes P)(A)$. Under the auxiliary experiment P_r^{k+1}, the $k+1$ observations are i.i.d. from P_r. Therefore, (A-21), applied with $G = P_r$, gives

$$(P_r^{k+1})(A) \geq 1 - \alpha \ .$$

The actual and auxiliary experiments differ only in the law of the future outcome. Thus,

$$\begin{aligned}(P_r^k \otimes P)(A) &\geq (P_r^k \otimes P_r)(A) - \mathrm{TV}(P_r^k \otimes P, P_r^k \otimes P_r) \\ &\geq 1 - \alpha - \mathrm{TV}(P, P_r) \ ,\end{aligned}$$

where the equality $\mathrm{TV}(P_r^k \otimes P, P_r^k \otimes P_r) = \mathrm{TV}(P, P_r)$ again follows from invariance of total variation under tensoring with a common probability measure. Integrating over $R_{(k+1)}$ gives

$$(Q^n \otimes P)\{Y_0 \in \widehat{C}_{n,k}(x_0)\} \geq 1 - \alpha - E_{Q^n}\left[\mathrm{TV}(P, P_{R_{(k+1)}})\right] \ . \tag{A-22}$$

It remains to bound the final expectation. The arguments used in the proof of Theorem 1, with $a_{tv} = 1$, give

$$E_{Q^n}\left[\mathrm{TV}(P, P_{R_{(k+1)}})\right] \leq \int_0^{r_1} \mathrm{TV}(P, P_r)\, dP_{R_{(k+1)}}(r) + P\{R_{(k+1)} \geq r_1\} = O((k/n)^{1/d}) \ .$$

The first term is $O((k/n)^{1/d})$ by (A-10) with $a_{tv} = 1$. For the second term, let $p_\star := F(r_1/2) > 0$ and argue as in (A-7)–(A-8) with $\beta = 1/d$, but without the factor k. If $k/n \leq p_\star/2$, then $P\{R_{(k+1)} \geq r_1\} \leq \exp(-np_\star/8) \leq Cn^{-1/d} \leq C(k/n)^{1/d}$ for a constant $C < \infty$ that does not depend on n or k. If $k/n > p_\star/2$, then $P\{R_{(k+1)} \geq r_1\} \leq 1 \leq (2/p_\star)^{1/d}(k/n)^{1/d}$. Hence $P\{R_{(k+1)} \geq r_1\} = O((k/n)^{1/d})$ uniformly over $n \in \mathbf{N}$ and $1 \leq k < n$. Substituting this bound into (A-22) proves (4.14). Finally, $k/n \to 0$ makes the coverage error vanish. ■

A.3.4 Policy-irrelevance testing under network interference

Proof of Corollary 4. Fix $a \in \{0, 1\}$. Within the half used for $\mathbf{g}_a$, the representative pairs (X^a, Y^a) are $n/2$ i.i.d. draws from Q_a, and $S_n(a)$ is the corresponding IOS vector with conditioning point zero. Since Assumptions 1 and 2 hold at zero for Q_a with P_a as the conditional law of Y^a at $X^a = 0$, Theorems 1–2, applied with $d = 1$ and $a_h = a_{tv} = 1$, give

$$\mathrm{TV}\big(\mathcal{L}(S_n(a)), P_a^k\big) = O\big(k^{3/2}/(n/2)\big) = O\big(k^{3/2}/n\big) \tag{A-23}$$

uniformly over even $n \in \mathbf{N}$ and $1 \leq k < n/2$. The two halves consist of disjoint sets of independent communities, together with their tie-breaking draws, so $S_n(0)$ and $S_n(1)$ are independent. The

product inequality for total variation and (A-23) yield (4.16). Under H_0, the $2k$ pooled ideal outcomes are i.i.d. from a common law, so $E[\phi(S)] \le \alpha$. Combining (2.6) and (4.16) proves (4.17). Finally, $k = o(n^{2/3})$ makes the approximation error vanish. ■

Appendix B Auxiliary results

Lemma 2. *Under Assumption 1, there exists $r_1 > 0$ and $0 < C_L < C_H < \infty$ such that, for all $r \in [0, r_1]$,*

$$F(r) := \int_{B_r} g(x)dx \quad \textit{satisfies} \quad C_L r^d \le F(r) \le C_H r^d \ . \tag{B-1}$$

Here $F(0) := 0$, and $F(r) = Q\{R < r\} = Q\{R \le r\}$ is the continuous CDF of R. In addition, for any $u \in (0, C_L r_1^d]$, the (left-continuous) quantile function of F,

$$F^{-1}(u) := \inf\{r : F(r) \ge u\}, \quad \textit{satisfies} \quad F^{-1}(u) \le (u/C_L)^{1/d} \ .$$

Proof. Let $r_g > 0$ be such that the Lipschitz bound in Assumption 1(i) holds for $x \in \mathcal{X}$ with $\|x - x_0\| < r_g$, and set

$$r_1 := \min\left\{\frac{r_g}{2}, \frac{r_0}{2}, \frac{g(x_0)}{2\max\{C_g, 1\}}\right\} > 0 \ .$$

Both parts of Assumption 1 therefore apply for $0 < r \le r_1$. Since X has a Lebesgue density, every sphere has probability zero, which proves the asserted continuity and identifies F as the CDF of R. For any $x \in \mathcal{X}$ with $\|x - x_0\| \le r$ and $r \in [0, r_1]$, we have

$$g(x) \overset{(1)}{\ge} g(x_0) - C_g r \overset{(2)}{\ge} (g(x_0)/2)(2 - C_g/(\max\{C_g, 1\})) \ge g(x_0)/2 \overset{(3)}{>} 0 \ , \tag{B-2}$$

where (1) and (3) hold by Assumption 1 and (2) by $r \le r_1$. Then, for all $r \in [0, r_1]$,

$$F(r) \overset{(1)}{\ge} \frac{g(x_0)}{2} \int_{B_r \cap \mathcal{X}} dx \overset{(2)}{\ge} \frac{g(x_0)}{2} C_{x_0} r^d \overset{(3)}{=} C_L r^d \ , \tag{B-3}$$

where (1) holds by (B-2) and $\|x - x_0\| \le r$, (2) by Assumption 1, and (3) by setting $C_L := \frac{g(x_0)}{2} C_{x_0}$. This proves the first inequality in (B-1).

For any $x \in \mathcal{X}$ with $\|x - x_0\| \le r$ and $r \in [0, r_1]$, we also have

$$g(x) \overset{(1)}{\le} g(x_0) + C_g r \overset{(2)}{\le} g(x_0) + C_g r_1 \ , \tag{B-4}$$

where (1) holds by Assumption 1 and (2) by $r \le r_1$. Then, for all $r \in [0, r_1]$,

$$F(r) \overset{(1)}{\le} (g(x_0) + C_g r_1)\mathrm{Vol}(B_r) \overset{(2)}{=} C_H r^d \ ,$$

where (1) holds by (B-4) and $\|x - x_0\| \le r$, and (2) by using that $\mathrm{Vol}(B_r) = \pi^{d/2}/\Gamma(d/2+1) r^d$ and setting $C_H := (g(x_0) + C_g r_1)\pi^{d/2}/\Gamma(d/2+1)$. This proves the second inequality in (B-1).

For the last claim, take any $r \in (0, r_1]$, and note that

$$F^{-1}(C_L r^d) \overset{(1)}{\le} F^{-1}(F(r)) \overset{(2)}{\le} r \ , \tag{B-5}$$

where (1) holds by the weak monotonicity of the quantile function and (B-3), and (2) by elementary properties of the quantile function (e.g., Pollard (2002, Eq. 36)). For any $u \in (0, C_L r_1^d]$, applying (B-5) to $r = (u/C_L)^{1/d} \in (0, r_1]$ proves the desired result. ■

Lemma 3. *Let Assumptions 1 and 2 hold and define, for $x \in \mathcal{X}$ near x_0 for which p_x is defined,*

$$D(x) := \int \left(\sqrt{p_x(y)} - \sqrt{p_{x_0}(y)}\right)^2 \nu(dy) \ .$$

Let

$$I(x_0) := \int \dot{\ell}_{x_0}(y)\, \dot{\ell}_{x_0}(y)^\top \, p_{x_0}(y)\, \nu(dy)$$

denote the $d \times d$ matrix induced by the score $\dot{\ell}_{x_0}$. Then, as $x \to x_0$,

$$D(x) = \frac{1}{4} (x - x_0)^\top I(x_0) (x - x_0) + o\big(\|x - x_0\|^2\big) \ .$$

Proof. For $t = x - x_0$ with $x \in \mathcal{X}$ near x_0 and p_x defined, set

$$a_t(y) := \tfrac{1}{2} t^\top \dot{\ell}_{x_0}(y) \sqrt{p_{x_0}(y)}, \qquad e_t(y) := \sqrt{p_{x_0+t}(y)} - \sqrt{p_{x_0}(y)} - a_t(y) \ .$$

Writing $\|\cdot\|_2$ for the norm in $L^2(\nu)$, QMD gives $\|e_t\|_2^2 = o(\|t\|^2)$ and score integrability gives

$$\begin{aligned} \|a_t\|_2^2 = \tfrac{1}{4} t^\top I(x_0) t = \tfrac{1}{4} \int (t^\top \dot{\ell}_{x_0}(y))^2 p_{x_0}(y) \ \nu(dy) \\ \le \tfrac{1}{4} \|t\|^2 \int \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y) \ \nu(dy) = \tfrac{1}{4} \|t\|^2 \operatorname{tr} I(x_0) \ . \end{aligned} \tag{B-6}$$

where the inequality follows from Cauchy-Schwarz. Expanding the square yields

$$D(x) = \|a_t + e_t\|_2^2 = \|a_t\|_2^2 + \|e_t\|_2^2 + 2 \int a_t(y) e_t(y) \nu(dy) \ . \tag{B-7}$$

Since $\|a_t\|_2^2 = \frac{1}{4} t^\top I(x_0) t$, the triangle and Cauchy–Schwarz inequalities,

$$\left| D(x) - \tfrac{1}{4} t^\top I(x_0) t \right| \le \|e_t\|_2^2 + 2 \|a_t\|_2 \|e_t\|_2 = o(\|t\|^2) \ . \tag{B-8}$$

All expressions are zero at $x = x_0$, so the decomposition holds there as well. This proves the claimed expansion. ■

References

Abramowitz, M. and I. A. Stegun (1964): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards. Applied Mathematics Series.

Adusumilli, K. (2025): "Risk and Optimal Policies in Bandit Experiments," *Econometrica*, forthcoming.

Auerbach, E., H. Guo, and M. Tabord-Meehan (2026): "The local approach to causal inference under network interference," *Quantitative Economics*, 17, 173–199.

Barber, R. F., E. J. Candès, A. Ramdas, and R. J. Tibshirani (2021): "The limits of distribution-free conditional predictive inference," *Information and Inference*, 10, 455–482.

Bhattacharya, P. (1974): "Convergence of sample paths of normalized sums of induced order statistics," *The Annals of Statistics*, 1034–1039.

Bugni, F. A. and I. A. Canay (2021): "Testing Continuity of a Density via g-order statistics in the Regression Discontinuity Design," *Journal of Econometrics*, 221, 138–159.

Bugni, F. A., I. A. Canay, and D. Kim (2025): "Testing Conditional Stochastic Dominance at Target Points," .

Canay, I. A. and V. Kamat (2018): "Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design," *The Review of Economic Studies*, 85, 1577–1608.

David, H. and J. Galambos (1974): "The asymptotic theory of concomitants of order statistics," *Journal of Applied Probability*, 762–770.

David, H. A. and H. N. Nagaraja (2003): *Order Statistics*, Wiley Series in Probability and Statistics, third edition ed.

Falk, M., J. Hüsler, and R.-D. Reiss (2010): *Laws of small numbers: extremes and rare events*, Springer Science & Business Media.

Fan, Y. and R. Liu (2016): "A direct approach to inference in nonparametric and semiparametric quantile models," *Journal of Econometrics*, 191, 196–216.

Guan, L. (2023): "Localized conformal prediction: A generalized inference framework for conformal prediction," *Biometrika*, 110, 33–50.

Hoeffding, W. and J. Wolfowitz (1958): "Distinguishability of Sets of Distributions," *The Annals of Mathematical Statistics*, 29, 700–718.

Hore, R. and R. F. Barber (2025): "Conformal prediction with local weights: Randomization enables robust guarantees," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87, 549–578.

Kaufmann, E. and R.-D. Reiss (1992): "On conditional distributions of nearest neighbors," *Journal of Multivariate Analysis*, 42, 67–76.

Le Cam, L. (1986): *Asymptotic Methods in Statistical Decision Theory*, Springer.

Min, Y., L. Peng, and C. Zou (2026): "A Unified Theory of Conditional Coverage in Conformal Prediction with Applications," Tech. Rep. 2605.11602, arXiv.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*, Cambrigde University Press, New York.

Reiss, R.-D. (1989): *Approximate distributions of order statistics: with applications to nonparametric statistics*, Springer-Verlag, New York.

——— (1993): *A Course on Point Processes*, Springer Series in Statistics, New York: Springer.

Van der Vaart, A. W. (2000): *Asymptotic statistics*, vol. 3, Cambridge university press.

Supplementary Appendix for

"Approximating an Ideal Experiment with Induced Order Statistics"

Federico A. Bugni
Department of Economics
Northwestern University
federico.bugni@northwestern.edu

Ivan A. Canay
Department of Economics
Northwestern University
iacanay@northwestern.edu

Deborah Kim
Department of Economics
University of Warwick
deborah.kim@warwick.ac.uk

November 12, 2026

Abstract

This document provides supplementary material for the paper "Approximating an Ideal Experiment with Induced Order Statistics." It collects auxiliary lemmas and technical proofs omitted from the main text, develops additional local approximation rate results under alternative smoothness assumptions, and provides further examples and discussion.

KEYWORDS: Hellinger distance, total variation distance, induced order statistics, rank tests, permutation tests.

JEL codes: C12, C14.

Supplementary Contents

Appendix S.1 Supplementary Lemmas **1**

Appendix S.2 Results under Alternative Assumptions **13**

S.2.1 Auxiliary Results under Alternative Assumptions 16

S.2.2 Proof of Theorem S.1 16

Appendix S.1 Supplementary Lemmas

Lemma S.1. *For every $d, m \geq 1$, there exists an admissible joint law Q, a covariate support $\mathcal{X}$, and a boundary point $x_0 \in \mathcal{X}$ satisfying Assumptions 1–2, together with constants $C > 0$ and $\bar{r} > 0$, such that*

$$\mathrm{H}(P_r, P) \geq Cr, \qquad \mathrm{TV}(P_r, P) \geq Cr \quad \text{for all } r \in (0, \bar{r}) \ . \tag{S.1.1}$$

Proof. Take $x_0 = 0_d$ and $X \sim \mathrm{Unif}([0,1]^d)$. Then $\mathcal{X} = [0,1]^d$, $g(x_0) = 1$, and Assumption 1 holds with $C_g = 0$, $r_0 = 1$, and $C_{x_0} = 2^{-d}\pi^{d/2}/\Gamma(d/2+1)$, since $\mathrm{Vol}(B_r \cap \mathcal{X}) = 2^{-d}\mathrm{Vol}(B_r) = C_{x_0} r^d$ for $0 < r < 1$. Let $Y \mid X = x \sim N((x_1, 0, \ldots, 0)^\top, I_m)$, with ν equal to Lebesgue measure on $\mathbf{R}^m$. The score with respect to the d-dimensional covariate is

$$\nabla_x \log p_x(y) = (y_1 - x_1, 0, \ldots, 0)^\top \in \mathbf{R}^d \ .$$

At x_0, take $\dot{\ell}_{x_0}(y) = (y_1, 0, \ldots, 0)^\top \in \mathbf{R}^d$, so that

$$\int \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y)\, dy = 1 \ .$$

For an admissible perturbation t, let $s = t_1$, $q_s = \sqrt{p_{x_0+t}}$, and $q_0 = \sqrt{p_{x_0}}$. Completing the square in the Gaussian integrals gives

$$\int q_s q_0 \, dy = e^{-s^2/8}, \qquad \int y_1 q_s q_0 \, dy = \tfrac{s}{2} e^{-s^2/8} \ .$$

Therefore,

$$\int \left(q_s - q_0 - \tfrac{s}{2} y_1 q_0\right)^2 dy = 2 + \tfrac{s^2}{4} - 2\left(1 + \tfrac{s^2}{4}\right) e^{-s^2/8} = O(s^4) = o(\|t\|^2) \ .$$

The remainder is exactly zero when $t_1 = 0$. This proves Assumption 2.

For the lower bound, let $A = \{y : y_1 > 0\}$ so that $P(A) = 1/2$. Let $Z = (Z_1, \ldots, Z_d)^\top$ be uniform on $B_1 \cap [0, \infty)^d$. Writing Φ and ϕ for the standard normal CDF and density, respectively,

gives the following:

$$\begin{aligned} P_r(A) - P(A) &= E[P(Y_1 > 0 \mid X) \mid X \in B_r] - \tfrac{1}{2} \\ &= E[\Phi(X_1) \mid X \in B_r] - \tfrac{1}{2} \\ &= E\big[\Phi(rZ_1) - \tfrac{1}{2}\big] \\ &\geq \phi(1) E[Z_1] r \ . \end{aligned}$$

where the third equality holds because $X \mid X \in B_r \stackrel{d}{=} rZ$ for $r \in (0,1)$, and the final inequality uses $E[\Phi(rZ_1) - \frac{1}{2}] = E[\int_0^{rZ_1} \phi(u)du] \geq E[\int_0^{rZ_1} \phi(1)du]$ for $r \in (0,1)$. Integrating over the truncated unit ball yields

$$E[Z_1] = \frac{\Gamma(d/2+1)}{\sqrt{\pi}\,\Gamma((d+3)/2)} > 0 \ .$$

The event bound and $\mathrm{TV} \leq \sqrt{2}\mathrm{H}$ give (S.1.1) with $\bar{r} = 1$ and $C = \phi(1)E[Z_1]/\sqrt{2}$. ■

Lemma S.2. *For every $d, m \geq 1$ and $\tilde{r} > 0$, there exists an admissible joint law Q satisfying Assumptions 1–2, with $B_r \subset \mathcal{X}$ for all $0 < r < \tilde{r}$, such that, for every $\varepsilon > 0$,*

$$\frac{\mathrm{H}(P_r, P)}{r^{1+\varepsilon}} \longrightarrow \infty \qquad \text{as } r \downarrow 0 \ , \tag{S.1.2}$$

$$\frac{\mathrm{TV}(P_r, P)}{r^{1+\varepsilon}} \longrightarrow \infty \qquad \text{as } r \downarrow 0 \ . \tag{S.1.3}$$

In particular, neither distance admits a common $O(r^{1+\varepsilon})$ bound over a class containing this model.

Proof. Fix $d, m \geq 1$ and $\tilde{r} > 0$. We construct one law that satisfies both conclusions for every $\varepsilon > 0$. We define the distribution Q as follows. Let $d \geq 1$, $m \geq 1$, and $x_0 = 0_d$. Let $X \sim \mathrm{Unif}(B_{\tilde{r}})$, so that $g(x) = 1[x \in B_{\tilde{r}}]/\mathrm{Vol}(B_{\tilde{r}})$ and $\mathcal{X} = B_{\tilde{r}}$. Next, let $e_1 := (1, 0, \ldots, 0)^\top \in \mathbf{R}^m$ and let $\delta : \mathbf{R} \to \mathbf{R}$ be defined as follows: $\delta(u) := u/(1 - \ln u)$ for $u \in (0, \exp(-2))$, and $\delta(u) := 0$ otherwise. Let $\{Y|X = x\}$ be supported on $\{0_m, e_1\} \subset \mathbf{R}^m$ with the following distribution:

$$P(Y = e_1|X = x) = \pi(x) := \frac{1}{2} + \delta(\|x\|) \text{ and } P(Y = 0_m|X = x) = 1 - \pi(x) \ .$$

Since $X \sim \mathrm{Unif}(B_{\tilde{r}})$ and $x_0 = 0_d$, Assumption 1 holds with $C_g = 0$, $r_0 = \tilde{r}$, and $C_{x_0} = \pi^{d/2}/\Gamma(d/2+1)$. Also, $B_r \cap \mathcal{X} = B_r$ for all $r \in (0, \tilde{r})$. Next, we show that Assumption 2 holds with $\dot{\ell}_{x_0}(y) := 0_d$ and ν equal to the counting measure on $\{0_m, e_1\}$. First, note that $\dot{\ell}_{x_0}(y) = 0_d$ implies $\int \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y)\nu(dy) = 0 < \infty$. Second, for any $\|t\| \in (0, \exp(-2))$, we have

$$\begin{aligned} &\int (\sqrt{p_{x_0+t}(y)} - \sqrt{p_{x_0}(y)} - \frac{1}{2} t^\top \dot{\ell}_{x_0}(y) \sqrt{p_{x_0}(y)})^2 \nu(dy) \\ &\stackrel{(1)}{=} (\sqrt{\tfrac{1}{2} + \delta(\|t\|)} - \sqrt{\tfrac{1}{2}})^2 + (\sqrt{\tfrac{1}{2} - \delta(\|t\|)} - \sqrt{\tfrac{1}{2}})^2 \\ &\stackrel{(2)}{=} 2 - \sqrt{1 + \tfrac{2\|t\|}{(1 - \ln \|t\|)}} - \sqrt{1 - \tfrac{2\|t\|}{(1 - \ln \|t\|)}} \ , \end{aligned} \tag{S.1.4}$$

where (1) holds by $\dot{\ell}_{x_0}(y) = 0_d$, $p_{x_0+t}(e_1) = 1/2 + \delta(\|x_0 + t\|)$, $p_{x_0+t}(0_m) = 1/2 - \delta(\|x_0 + t\|)$, and $p_{x_0}(y) = 1/2$ for $y \in \{0_m, e_1\}$, (2) by $\delta(u) := u/(1 - \ln u)$ for $u \in (0, \exp(-2))$. By a second-order Taylor expansion of $h(u) = 2 - \sqrt{1+2u} - \sqrt{1-2u}$ centered at $u = 0$, and that $h(0) = h'(0) = 0$,

$$h(u) = ((1+2\tilde{u})^{-3/2} + (1-2\tilde{u})^{-3/2})\frac{u^2}{2}$$

with $\tilde{u} \in [0, u]$. Next, note that $\|t\|/(1 - \ln\|t\|) \in (0, \exp(-2)/3)$ for $\|t\| \in (0, \exp(-2))$, and $h''(u) \in (2, 2.1)$ for $u \in (0, \exp(-2)/3)$. By plugging the results on (S.1.4), we conclude that

$$\int (\sqrt{p_{x_0+t}(y)} - \sqrt{p_{x_0}(y)} - \frac{1}{2}t^\top \dot{\ell}_{x_0}(y)\sqrt{p_{x_0}(y)})^2 \nu(dy) = O\Big(\tfrac{\|t\|^2}{(1-\ln\|t\|)^2}\Big) .$$

The right-hand side expression of the previous equation is $o(\|t\|^2)$ as $\|t\| \downarrow 0$, which verifies Assumption 2. Thus Q is an admissible interior model with the required neighborhood.

Let $A := \{e_1\}$. For $0 < r < \min\{\tilde{r}, e^{-2}\}$,

$$\begin{aligned}
\sqrt{2}\mathrm{H}(P_r, P) \geq \mathrm{TV}(P_r, P) &\geq |P_r(A) - P(A)| \\
&\overset{(1)}{=} \int_{\|x\| \leq r} \delta(\|x\|)dx/\mathrm{Vol}(B_r) \\
&\overset{(2)}{=} \tfrac{d}{r^d}\int_0^r v^d/(1 - \ln v)dv \\
&\overset{(3)}{\geq} \tfrac{d}{r^d}\int_{r/2}^r v^d/(1 - \ln v)dv \\
&\geq \tfrac{d}{(1-\ln(r/2))r^d}\int_{r/2}^r v^d dv \\
&= \tfrac{d(1-2^{-d-1})}{d+1}\tfrac{r}{1-\ln(r/2)} ,
\end{aligned} \tag{S.1.5}$$

where (1) holds because $P(Y = e_1 \mid X = x) = 1/2 + \delta(\|x\|) \geq 1/2$ for all x, $P(Y = e_1 \mid X = x_0) = 1/2$ since $\delta(0) = 0$, $X \sim \mathrm{Unif}(B_{\tilde{r}})$, and $r < \tilde{r}$, (2) by change of variables to polar coordinates, and (3) by $v^d/(1 - \ln v) \geq 0$ for $v < r < \exp(-2) < 1$. For every $\varepsilon > 0$, dividing the positive lower bound in (S.1.5) by $r^{1+\varepsilon}$ gives a quantity proportional to $r^{-\varepsilon}/(1-\log(r/2))$, which tends to infinity. This proves (S.1.3). Since $\mathrm{H}(P_r, P) \geq \mathrm{TV}(P_r, P)/\sqrt{2}$, it also proves (S.1.2). ■

Lemma S.3. *Suppose Assumptions 1–2 hold and x_0 lies in the interior of $\mathcal{X}$. Then,*

$$\mathrm{H}(P_r, P) = o(r) \qquad \textit{and} \qquad \mathrm{TV}(P_r, P) = o(r) .$$

Proof. Let r_1 be as in Lemma 2. Since x_0 is an interior point of $\mathcal{X}$, there exists $r_2 > 0$ such that $B_{r_2} \subset \mathcal{X}$. Set $\tilde{r} = \min\{r_1, r_2\} > 0$. For $r \in (0, \tilde{r})$, Lemma 2 gives $F(r) \geq C_L r^d > 0$, and Assumption 1 gives $g(x) > 0$ for $x \in B_r$, so p_x is well defined there.

For $r \in (0, \tilde{r})$, let T represent the shift from x_0 of the variable X drawn conditional on lying

within distance r of x_0. Its density is

$$\mu_r(t) := \frac{g(x_0+t)I\{\|t\| < r\}}{F(r)} ,$$

so that we can write $h_r(y) = E[p_{x_0+T}(y)]$. Define the QMD remainder

$$u_t(y) := \sqrt{p_{x_0+t}(y)} - \sqrt{p_{x_0}(y)} - \tfrac{1}{2}t^\top \dot{\ell}_{x_0}(y)\sqrt{p_{x_0}(y)} .$$

Since $h_r(y)$ is the expectation of $(\sqrt{p_{x_0+T}(y)})^2$, we obtain a useful auxiliary function by replacing the square-root density with its QMD approximation,

$$\widetilde{h}_r(y) := E\left[\left(\sqrt{p_{x_0}(y)} + \tfrac{1}{2}T^\top \dot{\ell}_{x_0}(y)\sqrt{p_{x_0}(y)}\right)^2\right] . \tag{S.1.6}$$

For ν-almost every fixed y, the definitions of h_r and $\widetilde{h}_r$ give

$$\sqrt{h_r(y)} = \left(\int \left|\sqrt{p_{x_0+t}(y)}\right|^2 \mu_r(t)\,dt\right)^{1/2}$$

and

$$\sqrt{\widetilde{h}_r(y)} = \left(\int \left|\sqrt{p_{x_0}(y)} + \tfrac{1}{2}t^\top \dot{\ell}_{x_0}(y)\sqrt{p_{x_0}(y)}\right|^2 \mu_r(t)\,dt\right)^{1/2} .$$

That is, these are the $L^2(\mu_r)$ norms of the two functions inside the absolute values, viewed as functions of t with y held fixed. Their difference is $u_t(y)$, so the reverse triangle inequality gives

$$\left|\sqrt{h_r(y)} - \sqrt{\widetilde{h}_r(y)}\right| \le \left(\int |u_t(y)|^2 \mu_r(t)\,dt\right)^{1/2} = \left(E[u_T(y)^2]\right)^{1/2} .$$

Combining this inequality with $(a+b)^2 \le 2a^2 + 2b^2$ and Tonelli's theorem yields

$$\mathrm{H}^2(P_r, P) \le E\left[\int u_T(y)^2\,\nu(dy)\right] + \int \left(\sqrt{\widetilde{h}_r(y)} - \sqrt{p_{x_0}(y)}\right)^2 \nu(dy) . \tag{S.1.7}$$

Next, we show that both terms on the right-hand side of (S.1.7) are $o(r^2)$. For the first, note that Assumption 2 implies that

$$\epsilon(r) := \sup_{0<\|t\|<r} \frac{\int u_t(y)^2\,\nu(dy)}{\|t\|^2} \to 0 .$$

Since $\|T\| < r$ almost surely and $u_0 = 0$, this gives

$$E\left[\int u_T(y)^2\,\nu(dy)\right] \le \epsilon(r)E[\|T\|^2] \le \epsilon(r)r^2 = o(r^2) .$$

For the second term, first note that

$$\|E[T]\| \overset{(1)}{=} \frac{1}{F(r)} \left\| \int_{\{\|t\|<r\}} t\{g(x_0+t) - g(x_0)\}\, dt \right\| \overset{(2)}{\leq} \frac{C_g}{C_L r^d} \int_{\{\|t\|<r\}} \|t\|^2\, dt = O(r^2) \ , \tag{S.1.8}$$

where (1) uses $\int_{\{\|t\|<r\}} t\, dt = 0$, and (2) follows from Assumption 1 and Lemma 2. Next, expanding the square in (S.1.6) gives

$$\widetilde{h}_r(y) = p_{x_0}(y) \left\{ 1 + E[T]^\top \dot{\ell}_{x_0}(y) + \tfrac{1}{4} E\left[(T^\top \dot{\ell}_{x_0}(y))^2 \right] \right\} = p_{x_0}(y)\{1 + O(r^2)\} \ ,$$

where the second equality follows from (S.1.8) and $E[(T^\top \dot{\ell}_{x_0}(y))^2] \leq r^2 \|\dot{\ell}_{x_0}(y)\|^2$. That is, the term in braces is $1 + O(r^2)$ for each fixed y. Thus, if $p_{x_0}(y) > 0$,

$$\frac{\sqrt{\widetilde{h}_r(y)} - \sqrt{p_{x_0}(y)}}{r} = \frac{\widetilde{h}_r(y) - p_{x_0}(y)}{r\{\sqrt{\widetilde{h}_r(y)} + \sqrt{p_{x_0}(y)}\}} \to 0 \ .$$

If $p_{x_0}(y) = 0$, then $\widetilde{h}_r(y) = 0$ and the same conclusion holds. Next, note that $\sqrt{\widetilde{h}_r(y)}$ and $\sqrt{p_{x_0}(y)}$ are the $L^2(\mu_r)$ norms of $\sqrt{p_{x_0}(y)} + \frac{1}{2} t^\top \dot{\ell}_{x_0}(y) \sqrt{p_{x_0}(y)}$ and $\sqrt{p_{x_0}(y)}$, respectively, viewed as functions of t with y held fixed. Thus, application of the reverse triangle inequality in $L^2(\mu_r)$ to these functions yield

$$\left| \sqrt{\widetilde{h}_r(y)} - \sqrt{p_{x_0}(y)} \right| \leq \tfrac{1}{2} \sqrt{p_{x_0}(y)} \left(E[(T^\top \dot{\ell}_{x_0}(y))^2] \right)^{1/2} \leq \tfrac{1}{2} r \|\dot{\ell}_{x_0}(y)\| \sqrt{p_{x_0}(y)} \ .$$

Consequently,

$$\frac{\left| \sqrt{\widetilde{h}_r(y)} - \sqrt{p_{x_0}(y)} \right|^2}{r^2} \leq \tfrac{1}{4} \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y) \ .$$

The right-hand side is integrable by Assumption 2. Dominated convergence therefore implies that the second term in (S.1.7) is $o(r^2)$. This proves $\mathrm{H}(P_r, P) = o(r)$, and the total variation conclusion follows from $\mathrm{TV}(P_r, P) \leq \sqrt{2}\, \mathrm{H}(P_r, P)$. ■

Lemma S.4. *Let $U_{(k:n)}$ be the k-th order statistic from an i.i.d. sample $U_1, \ldots, U_n \sim \mathrm{U}(0,1)$. For any real $m > 0$,*

$$E[U_{(k:n)}^m] = O((k/n)^m) \ .$$

Proof. For any $m > 0$, $k \in \{1, \ldots, n\}$, and $n \in \mathbf{N}$, consider the following derivation:

$$E[U_{(k:n)}^m] \overset{(1)}{=} \frac{B(m+k, n-k+1)}{B(k, n-k+1)} \overset{(2)}{=} \frac{\Gamma(m+k)/\Gamma(k)}{\Gamma(m+n+1)/\Gamma(n+1)} \ , \tag{S.1.9}$$

where (1) holds by David and Nagaraja (2003, p. 10-11), Abramowitz and Stegun (1964, 6.2.1, p. 257), where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt : (0,\infty)^2 \to (0,\infty)$ is the Beta function, and (2) by

Abramowitz and Stegun (1964, 6.2.2, p. 257), which gives $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, where $\Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)dt : (0,\infty) \to (0,\infty)$ is the Gamma function.

Next, we show that, for all $m > 0$, we can find constants $C_1(m), C_2(m)$ with $0 < C_1(m) < C_2(m) < \infty$ such that for all $t \in \mathbf{N}$,

$$0 < C_1(m) \le \frac{\Gamma(m+t)}{t^m\Gamma(t)} \le C_2(m) < \infty \ . \tag{S.1.10}$$

By definition, $\Gamma : (0,\infty) \to \mathbf{R}$ is continuous and positive. Therefore, for all $t \in \mathbf{N}$, $t^{-m}\Gamma(m+t)/\Gamma(t)$ is positive. Also, Abramowitz and Stegun (1964, 6.1.46, p. 257) implies that $\lim_{t\to\infty} t^{-m}\Gamma(m+t)/\Gamma(t) = 1$ for any $m \in \mathbf{R}$. Therefore, $\exists t_1 \in \mathbf{N}$ such that for all $t > t_1$, $t^{-m}\Gamma(m+t)/\Gamma(t) \in (1/2, 2)$. Then, for all $t \in \mathbf{N}$,

$$\frac{\Gamma(m+t)}{t^m\Gamma(t)} \le \max\left\{\max_{t\le t_1}\frac{\Gamma(m+t)}{t^m\Gamma(t)}, \max_{t>t_1}\frac{\Gamma(m+t)}{t^m\Gamma(t)}\right\} \le \max\left\{\max_{t\le t_1}\frac{\Gamma(m+t)}{t^m\Gamma(t)}, 2\right\} := C_2(m) < \infty$$

and

$$\frac{\Gamma(m+t)}{t^m\Gamma(t)} \ge \min\left\{\min_{t\le t_1}\frac{\Gamma(m+t)}{t^m\Gamma(t)}, \min_{t>t_1}\frac{\Gamma(m+t)}{t^m\Gamma(t)}\right\} \ge \min\left\{\min_{t\le t_1}\frac{\Gamma(m+t)}{t^m\Gamma(t)}, \frac{1}{2}\right\} := C_1(m) > 0 \ ,$$

as desired in (S.1.10).

To conclude the proof, for any $k \in \{1,\dots,n\}$ and $n \in \mathbf{N}$, by (S.1.9),

$$E[U_{(k:n)}^m] = \frac{\Gamma(m+k)/\Gamma(k)}{\Gamma(m+n+1)/\Gamma(n+1)} \ .$$

Applying (S.1.10) with $t = k$ in the numerator and with $t = n+1$ in the denominator gives

$$E[U_{(k:n)}^m] \le \frac{C_2(m)k^m}{C_1(m)(n+1)^m} = \frac{C_2(m)}{C_1(m)}\Big(\frac{k}{n+1}\Big)^m \le \frac{C_2(m)}{C_1(m)}\Big(\frac{k}{n}\Big)^m \ .$$

Hence $E[U_{(k:n)}^m] = O((k/n)^m)$. ■

Lemma S.5. *Suppose $g(x_0) > 0$ and Assumption 3 holds with additional integrability condition: $\zeta_2 \in L^p(p_{x_0}\cdot\nu)$. Then, for $p \ge 1$,*

$$\mathrm{H}(P_r, P) = O(r^{\min\{p,2\}}) \quad \textit{and} \quad \mathrm{TV}(P_r, P) = O(r^2) \ .$$

Proof. By Assumption 3, there exist $\tilde{\zeta}_{2,v}$ and $\varepsilon_0 > 0, C_1 > 0$ uniformly for all $v \in \mathbf{R}^d$ with $\|v\| \le \varepsilon_0$ and all $y \in \mathbf{R}^m$,

$$f(x_0+v, y) = f(x_0, y)\big[1 + \zeta_1(y)^\top v + \|v\|^2\tilde{\zeta}_{2,v}(y)\big], \qquad |\tilde{\zeta}_{2,v}(y)| \le C_1\zeta_2(y) \ .$$

Integrating with respect to y gives

$$g(x_0+v) = g(x_0)\left[1 + \bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2\right] ,$$

where

$$\bar{\zeta}_1 := \int \zeta_1(y) p_{x_0}(y)\nu(dy), \qquad \bar{\zeta}_{2,v} := \int \tilde{\zeta}_{2,v}(y) p_{x_0}(y)\nu(dy) .$$

Since $\zeta_2(y) \in L^p(p_{x_0}\cdot\nu)$, $\bar{\zeta}_{2,v}$ is bounded uniformly for $\|v\| \le \varepsilon_0$:

$$|\bar{\zeta}_{2,v}| \le \int |\tilde{\zeta}_{2,v}(y)| p_{x_0}(y)\nu(dy) \ \le C_1 \int \zeta_2(y) p_{x_0}(y)\nu(dy) =: C_1\bar{\zeta}_2 < \infty .$$

Choose $r \in (0, \varepsilon_0)$. To obtain $h_r(y)$, we integrate the joint density over B_r. Let $B_r(0)$ be a ball of radius r around the origin. Since $x_0 \in \mathrm{int}(\mathcal{X})$, by symmetry, the numerator is

$$\int_{B_r(0)} f(x_0+v, y)\, dv = \mathrm{Vol}(B_r) f(x_0, y)[1 + A_r(y)], \quad \text{where } A_r(y) := \frac{1}{\mathrm{Vol}(B_r)} \int_{B_r(0)} \|v\|^2 \tilde{\zeta}_{2,v}(y)\, dv .$$

On $B_r(0)$, $\|v\|^2 \le r^2$, so $A_r(y)$ is bounded from above:

$$|A_r(y)| \le \frac{1}{\mathrm{Vol}(B_r)} \int_{B_r(0)} \|v\|^2 |\tilde{\zeta}_{2,v}(y)|\, dv \le \frac{1}{\mathrm{Vol}(B_r)} \int_{B_r(0)} r^2 C_1 \zeta_2(y)\, dv = C_1 r^2\, \zeta_2(y) . \quad \text{(S.1.11)}$$

To get the denominator, integrate $g(x_0+v)$ over $B_r(0)$. Then, by Tonelli's theorem, we have

$$\int_{B_r(0)} g(x_0+v)\, dv = \int_{B_r(0)} \int f(x_0+v, y)\, \nu(dy)\, dv = \mathrm{Vol}(B_r) \int f(x_0, y)[1 + A_r(y)]\, \nu(dy) .$$

Therefore, we have

$$\int_{B_r(0)} g(x_0+v)\, dv = \mathrm{Vol}(B_r) g(x_0)[1 + \bar{A}_r] \quad \text{where} \quad \bar{A}_r := \frac{1}{g(x_0)} \int f(x_0, y) A_r(y)\, \nu(dy) ,$$

Using the upper bound $|A_r(y)| \le C_1 r^2 \zeta_2(y)$, we have

$$|\bar{A}_r| \le \frac{1}{g(x_0)} \int f(x_0, y)|A_r(y)|\, \nu(dy) \le C_1 r^2 \int p_{x_0}(y)\zeta_2(y)\, \nu(dy) \le C_1 r^2 \bar{\zeta}_2 .$$

Combining the denominator and the numerator, we get

$$h_r(y) = p_{x_0}(y)\, \frac{1 + A_r(y)}{1 + \bar{A}_r} .$$

Now for simplicity, define

$$\Delta_r(y) := \frac{h_r(y)}{p_{x_0}(y)} - 1 = \frac{A_r(y) - \bar{A}_r}{1 + \bar{A}_r} .$$

From the upper bound of $|\bar{A}_r|$, there exists $\varepsilon_1 \in (0, \varepsilon_0)$ such that, for any $r \in (0, \varepsilon_1)$, $|\bar{A}_r| \leq \frac{1}{2}$, so

$$|\Delta_r(y)| \;\leq\; 2\big(|A_r(y)| + |\bar{A}_r|\big) \;\leq\; 2C_1\, r^2\, (\zeta_2(y) + \bar{\zeta}_2) =: C_2\, r^2 u(y) \tag{S.1.12}$$

where $C_2 := 2C_1 > 0$ and $u(y) := \zeta_2(y) + \bar{\zeta}_2$. Then, the $L^p(p_{x_0} \cdot \nu)$-integrability condition of ζ_2 implies that $u \in L^p(p_{x_0} \cdot \nu)$, i.e.,

$$\|u\|^p_{L^p(p_{x_0}\cdot\nu)} := \int |\zeta_2(y) + \bar{\zeta}_2|^p \; p_{x_0}(y) \; \nu(dy) \leq 2^{p-1} \int (|\zeta_2(y)|^p + |\bar{\zeta}_2|^p)\, p_{x_0}(y)\, \nu(dy) < \infty \;, \tag{S.1.13}$$

as $(\frac{a+b}{2})^p \leq \frac{1}{2}(a^p + b^p)$ holds by convexity for any $p \geq 1$ and $a, b \geq 0$.

Rate of $\mathrm{H}(P_r, P)$**.** Under our assumptions, $h_0(y) := \lim_{r\to 0} h_r(y) = p_{x_0}(y)$ for all $y \in \mathcal{Y}$. We use the inequality: for $s > -1$, $(\sqrt{1+s} - 1)^2 \leq \min\{s^2, |s|\}$, which gives

$$\Big(\sqrt{h_r(y)} - \sqrt{h_0(y)}\Big)^2 = \Big(\sqrt{1 + \Delta_r(y)} - 1\Big)^2 h_0(y) \leq \min\{\Delta_r^2(y), |\Delta_r(y)|\}\; h_0(y) \;. \tag{S.1.14}$$

Case $1 \leq p \leq 2$. For simplicity, let $\tau_r := r^{-2}$. Define $\mathcal{A}_r := \{y \in \mathcal{Y} : u(y) \leq \tau_r\}$ and $\mathcal{A}_r^c := \mathcal{Y} \backslash \mathcal{A}_r$. By (S.1.14), we have

$$\begin{aligned}
\mathrm{H}(P_r, P)^2 &\overset{(1)}{\leq} \frac{1}{2} \int \min\{\Delta_r^2(y), |\Delta_r(y)|\}\; h_0(y)\; \nu(dy) \\
&\overset{(2)}{\leq} \frac{1}{2} \int_{\mathcal{A}_r} \Delta_r^2(y)\; h_0(y)\; \nu(dy) + \frac{1}{2} \int_{\mathcal{A}_r^c} |\Delta_r(y)|\; h_0(y)\; \nu(dy) \\
&\overset{(3)}{\leq} \frac{C_2^2 r^4}{2} \int_{\mathcal{A}_r} u(y)^2\; h_0(y)\; \nu(dy) + \frac{C_2 r^2}{2} \int_{\mathcal{A}_r^c} u(y)\; h_0(y)\; \nu(dy) \\
&\overset{(4)}{\leq} \frac{C_2^2 r^4}{2} \int_{\mathcal{A}_r} \tau_r^{2-p} u^p(y)\; h_0(y)\; \nu(dy) + \frac{C_2 r^2}{2} \int_{\mathcal{A}_r^c} \tau_r^{1-p} u^p(y)\; h_0(y)\; \nu(dy) \\
&\overset{(5)}{\leq} \frac{C_2^2 r^{2p}}{2} \int u^p(y)\; h_0(y)\; \nu(dy) + \frac{C_2 r^{2p}}{2} \int u^p(y)\; h_0(y)\; \nu(dy) = r^{2p} \frac{C_2^2 + C_2}{2} \|u\|^p_{L^p(p_{x_0}\cdot\nu)} \;,
\end{aligned}$$

where (1) follows from (S.1.14), we have (2) by splitting $\mathcal{Y}$ into $\mathcal{A}_r$ and $\mathcal{A}_r^c$ and by the definition of minimum, (3) follows from (S.1.12), (4) follows from

$$\begin{aligned}
&\frac{u(y)}{\tau_r} \leq 1 \;\to\; \left(\frac{u(y)}{\tau_r}\right)^{2-p} \leq 1 \;\to\; u(y)^2 \leq \tau_r^{2-p} u^p(y) \quad \text{for any } y \in \mathcal{A}_r \;, \\
&\frac{u(y)}{\tau_r} > 1 \;\to\; \left(\frac{u(y)}{\tau_r}\right)^{1-p} \leq 1 \;\to\; u(y) \leq \tau_r^{1-p} u^p(y) \quad \text{for any } y \in \mathcal{A}_r^c \;,
\end{aligned}$$

and lastly (5) holds by replacing $\mathcal{A}_r$ and $\mathcal{A}_r^c$ with $\mathcal{Y}$ respectively and by substituting $\tau_r = r^{-2}$. By (S.1.13), $\mathrm{H}(P_r, P) = O(r^p)$.

Case $p \geq 2$. Similar as before, by using (S.1.14) and $(\sqrt{1+s}-1)^2 \leq s^2$, we have

$$\mathrm{H}(P_r, P)^2 \leq \frac{1}{2}\int \Delta_r^2(y)\; h_0(y)\; \nu(dy) \leq \frac{C_2^2 r^4}{2}\int u(y)^2\; h_0(y)\; \nu(dy) = \frac{r^4 C_2^2}{2}\|u\|^2_{L^2(p_{x_0}\cdot\nu)}\;,$$

where the second inequality follows from (S.1.12). Since $p \geq 2$ and $u \in L^p(p_{x_0}\cdot\nu)$ as shown in (S.1.13), the Lyapunov's inequality gives $u \in L^2(p_{x_0}\cdot\nu)$ as well. Thus, $\mathrm{H}(P_r, P) = O(r^2)$. Combining the two cases, we conclude $\mathrm{H}(P_r, P) = O\big(r^{\min\{p,2\}}\big)$.

Rate of $\mathrm{TV}(P_r, P)$**.** The total variation distance between P_r and P is

$$\begin{aligned}\mathrm{TV}(P_r, P) &= \tfrac{1}{2}\int |h_r(y) - h_0(y)|\,\nu(dy) = \tfrac{1}{2}\int |\Delta_r(y)| h_0(y)\,\nu(dy)\\ &\leq \tfrac{C_2}{2} r^2 \int u(y) h_0(y)\,\nu(dy) = \tfrac{C_2}{2} r^2 \|u\|_{L^1(p_{x_0}\cdot\nu)}\;.\end{aligned}$$

Since $\|u\|_{L^1(p_{x_0}\cdot\nu)} < \infty$, we have $\mathrm{TV}(P_r, P) = O(r^2)$ as $r \to 0$. ■

Lemma S.6. *Assumption 3 implies Assumption 1 with* $x_0 \in \mathrm{int}(\mathcal{X})$ *and Assumption 2. The converse is not true.*

Proof. By Assumption 3, there exist $\tilde{\zeta}_{2,v}$ and $C_0, \varepsilon_0 > 0$ uniformly for all $v \in \mathbf{R}^d$ with $\|v\| \leq \varepsilon_0$ and all $y \in \mathbf{R}^m$,

$$f(x_0 + v, y) = f(x_0, y)\big[1 + \zeta_1(y)^\top v + \|v\|^2 \tilde{\zeta}_{2,v}(y)\big], \qquad |\tilde{\zeta}_{2,v}(y)| \leq C_0 \zeta_2(y)\;. \tag{S.1.15}$$

Integrating with respect to y gives

$$g(x_0 + v) = g(x_0)\big[1 + \bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2\big] \tag{S.1.16}$$

where

$$\bar{\zeta}_1 := \int \zeta_1(y) p_{x_0}(y)\nu(dy), \qquad \bar{\zeta}_{2,v} := \int \tilde{\zeta}_{2,v}(y)\, p_{x_0}(y)\,\nu(dy)\;.$$

To see Assumption 1 holds, note that since $x_0 \in \mathrm{int}(\mathcal{X})$, $B_r \subseteq \mathcal{X}$ for all small r, and so $\mathrm{Vol}(B_r \cap \mathcal{X}) = \mathrm{Vol}(B_r)$ is of order r^d. In addition, since $\zeta_2(y) \in L^2(p_{x_0}\cdot\nu)$, $\bar{\zeta}_{2,v}$ is bounded uniformly for $\|v\| \leq \varepsilon_0$:

$$|\bar{\zeta}_{2,v}| \leq C_0 \int \zeta_2(y) p_{x_0}(y)\nu(dy) =: \bar{\zeta}_2 < \infty\;.$$

Let $\varepsilon_1 := \min\{1, \varepsilon_0\}$. Then, for all v with $\|v\| \leq \varepsilon_1$

$$|g(x_0 + v) - g(x_0)| \overset{(a)}{=} |g(x_0)\bar{\zeta}_1^\top v + g(x_0)\bar{\zeta}_{2,v}\|v\|^2| \overset{(b)}{\leq} g(x_0)\{\|\bar{\zeta}_1\| \cdot \|v\| + |\bar{\zeta}_{2,v}|\|v\|^2\} \overset{(c)}{\leq} C_g\|v\|$$

where (a) follows from (S.1.16), (b) holds by Cauchy-Schwarz and the triangle inequality, and (c) holds because $\|v\|^2 \leq \|v\|$ and by setting $C_g := g(x_0)(\|\bar{\zeta}_1\| + \bar{\zeta}_2)$.

Next to show that Assumption 2 holds, let

$$\dot{\ell}_{x_0}(y) := \zeta_1(y) - \bar{\zeta}_1, \qquad a_v(y) := \zeta_1(y)^\top v + \tilde{\zeta}_{2,v}(y)\|v\|^2, \qquad b_v := \sqrt{\frac{1}{1+\bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2}} \ .$$

Combining (S.1.15) and (S.1.16), we obtain $\sqrt{p_{x_0+v}(y)} = \sqrt{p_{x_0}(y)}\, b_v \sqrt{1+a_v(y)}$ holds. Define

$$\Delta_v(y) := \frac{\sqrt{p_{x_0+v}(y)} - \sqrt{p_{x_0}(y)}}{\|v\|} - \frac{1}{2}\dot{\ell}_{x_0}(y)^\top \frac{v}{\|v\|}\sqrt{p_{x_0}(y)} = \sqrt{p_{x_0}(y)}\{T_{1,v}(y) + T_{2,v}(y)\} \ ,$$

where

$$T_{1,v}(y) := \frac{b_v}{\|v\|}\Big(\sqrt{1+a_v(y)} - 1 - \tfrac{1}{2}a_v(y)\Big) \tag{S.1.17}$$

$$T_{2,v}(y) := \frac{b_v - 1}{\|v\|} + \frac{1}{2}\bar{\zeta}_1^\top \frac{v}{\|v\|} + \frac{b_v}{2\|v\|}a_v(y) - \frac{1}{2}\zeta_1(y)^\top \frac{v}{\|v\|} \ . \tag{S.1.18}$$

Below, we show that the integrals of the squares of $\sqrt{p_{x_0}}\ T_{2,v}$ and $\sqrt{p_{x_0}}\ T_{1,v}$ are $o(1)$.

Term $T_{2,v}$. Applying the Taylor expansion $(1+t)^{-\frac{1}{2}} = 1 - \frac{1}{2}t + O(t^2)$ around 0 to b_v gives

$$b_v = 1 - \tfrac{1}{2}(\bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2) + O((\bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2)^2) = 1 - \tfrac{1}{2}(\bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2) + O(\|v\|^2) \ .$$

The second equality holds because, for some constants $\bar{C}, \bar{\varepsilon} > 0$ uniformly for $\|v\| < \min(\varepsilon_1, \bar{\varepsilon})$,

$$O((\bar{\zeta}_1^\top v + \bar{\zeta}_{2,v}\|v\|^2)^2) \le 2\bar{C}((\bar{\zeta}_1^\top v)^2 + (\bar{\zeta}_{2,v}\|v\|^2)^2) \le 2\bar{C}(\|\bar{\zeta}_1\|^2\|v\|^2 + |\bar{\zeta}_{2,v}|^2\|v\|^4) = O(\|v\|^2)$$

where the two inequalities follow from $(a+b)^2 \le 2(a^2+b^2)$, the Cauchy-Schwarz inequality, and the last equality holds because $\|\zeta_1\|^2 < \infty$ and $|\bar{\zeta}_{2,v}|^2 < \bar{\zeta}_2^2$ uniformly for $\|v\| < \min(\varepsilon_1, \bar{\varepsilon})$. Thus, there exist $C_1, C_2 > 0$, and $\varepsilon_2 \in (0, \varepsilon_1)$ such that, uniformly for all v with $\|v\| < \varepsilon_2$

$$\left|\frac{b_v - 1}{\|v\|} + \frac{1}{2}\bar{\zeta}_1^\top \frac{v}{\|v\|}\right| \le \tfrac{1}{2}|\bar{\zeta}_{2,v}|\|v\| + O(\|v\|) \le C_1\|v\| \ , \tag{S.1.19}$$

$$|b_v - 1| \le \tfrac{1}{2}(\|\bar{\zeta}_1\|\|v\| + |\bar{\zeta}_{2,v}|\|v\|^2) + O(\|v\|^2) \le C_2\|v\| \ , \tag{S.1.20}$$

$$b_v^2 \le \max\{(1+C_2\varepsilon_2)^2, (1-C_2\varepsilon_2)^2\} =: \bar{b} \ . \tag{S.1.21}$$

The term $T_{2,v}(y)^2$ is bounded uniformly for all v with $\|v\| < \varepsilon_2$ as follows:

$$\begin{aligned}
T_{2,v}(y)^2 &\le 2\left(\frac{b_v - 1}{\|v\|} + \frac{1}{2}\bar{\zeta}_1^\top \frac{v}{\|v\|}\right)^2 + 2\left(\frac{b_v}{2\|v\|}a_v(y) - \frac{1}{2}\zeta_1(y)^\top \frac{v}{\|v\|}\right)^2 \\
&\le 2C_1^2\|v\|^2 + 2\left(\frac{(b_v-1)}{2\|v\|}\zeta_1(y)^\top v + \frac{b_v\|v\|}{2}\tilde{\zeta}_{2,v}(y)\right)^2 \\
&\le 2C_1^2\|v\|^2 + (b_v-1)^2\|\zeta_1(y)\|^2 + b_v^2\|v\|^2\tilde{\zeta}_{2,v}(y)^2 \\
&\le C_3\|v\|^2\left\{1 + \|\zeta_1(y)\|^2 + C_0^2\zeta_2(y)^2\right\}
\end{aligned}$$

where the first inequality holds by $(a+b)^2 \le 2a^2+2b^2$, the second inequality holds by (S.1.19) and by plugging in $a_v(y)$, the third inequality holds by $(a+b)^2 \le 2a^2+2b^2$ and the Cauchy-Schwarz inequality, and the final inequality follows from (S.1.20), (S.1.21), setting $C_3 := \max\{2C_1^2, C_2^2, \bar{b}\}$, and using $\tilde{\zeta}_{2,v}(y)^2 \le C_0^2\zeta_2(y)^2$. Therefore, uniformly for all v with $\|v\| < \varepsilon_2$

$$\int T_{2,v}(y)^2 p_{x_0}(y)\,\nu(dy) \le C_3\left\{1 + \int \Big(\|\zeta_1(y)\|^2 + C_0^2\zeta_2(y)^2\Big)p_{x_0}(y)\nu(dy)\right\}\|v\|^2 =: C_4\|v\|^2 \ .$$

Term $T_{1,v}$. For $u \ge -1$ the following inequality holds:

$$\left|\sqrt{1+u} - 1 - \tfrac{1}{2}u\right| \le C_5|u|\min\{|u|, 1\} \tag{S.1.22}$$

for some constant $C_5 > 0$. We apply this with $u = a_v(y)$ since $a_v(y) = f(x_0+v, y)/f(x_0, y) - 1 > -1$ for all y such that $p_{x_0}(y) > 0$. Partition $\mathcal{Y}$ into

$$E_1 := \big\{y \in \mathcal{Y} : |a_v(y)| \le \delta_v\big\}, \qquad E_2 := \big\{y \in \mathcal{Y} : |a_v(y)| > \delta_v\big\}, \qquad \text{with} \qquad \delta_v := \sqrt{\|v\|}\ .$$

On E_1, by applying (S.1.22) and $\min\{|a_v(y)|, 1\} \le |a_v(y)|$

$$T_{1,v}(y)^2 = \frac{b_v^2}{\|v\|^2}\Big|\sqrt{1+a_v(y)} - 1 - \tfrac{1}{2}a_v(y)\Big|^2 \le \frac{b_v^2 C_5^2}{\|v\|^2}|a_v(y)|^4 \le C_5^2 b_v^2\delta_v^2\Big|\frac{a_v(y)}{\|v\|}\Big|^2 \ .$$

By the triangle inequality and Cauchy-Schwarz,

$$\left|\frac{a_v(y)}{\|v\|}\right| = \left|\zeta_1(y)^\top\frac{v}{\|v\|} + \tilde{\zeta}_{2,v}(y)\|v\|\right| \le \|\zeta_1(y)\| + |\tilde{\zeta}_{2,v}(y)|\|v\| \le \|\zeta_1(y)\| + C_0|\zeta_2(y)|\|v\| \ , \tag{S.1.23}$$

uniformly for all v with $\|v\| < \varepsilon_2$. Then for all v with $\|v\| < \varepsilon_2$

$$\int_{E_1} T_{1,v}(y)^2 p_{x_0}(y)\nu(dy) \le C_5^2 b_v^2\|v\|\int 2\big(\|\zeta_1(y)\|^2 + C_0^2|\zeta_2(y)|^2\|v\|^2\big)p_{x_0}(y)\nu(dy) \le C_6\|v\| + C_7\|v\|^3$$

where we use by $(a+b)^2 \le 2a^2+2b^2$ for the first inequality and the second inequality follows from (S.1.21) and setting $C_6 := 2C_5^2\bar{b}\int\|\zeta_1(y)\|^2 p_{x_0}(y)\,\nu(dy)$ and $C_7 := 2C_5^2\bar{b}C_0^2\int|\zeta_2(y)|^2 p_{x_0}(y)\,\nu(dy)$.

Since $\min\{|a_v(y)|, 1\} \le 1$, we have, for all v with $\|v\| < \varepsilon_2$,

$$T_{1,v}(y)^2 \le C_5^2 b_v^2\Big|\frac{a_v(y)}{\|v\|}\Big|^2 \le 2C_5^2\bar{b}\{\|\zeta_1(y)\|^2 + C_0^2|\zeta_2(y)|^2\|v\|^2\}\ ,$$

where the second inequality holds by applying $(a+b)^2 \le 2a^2+2b^2$ to (S.1.23) and by using $b_v^2 \le \bar{b}$ in (S.1.21). Integrating this on E_2 yields that uniformly for all v with $\|v\| < \varepsilon_2$,

$$\int_{E_2} T_{1,v}(y)^2 p_{x_0}(y)\,\nu(dy) \le 2C_5^2\bar{b}\int_{E_2}\|\zeta_1(y)\|^2 p_{x_0}(y)\,\nu(dy) + C_8\|v\|^2\ ,$$

where $C_8 := 2C_5^2\bar{b}C_0^2 \int |\zeta_2(y)|^2 p_{x_0}(y)\,\nu(dy)$. We finally claim that

$$\int_{E_2} \|\zeta_1(y)\|^2 p_{x_0}(y)\,\nu(dy) = o(1)\ . \tag{S.1.24}$$

To see this, fix y with $p_{x_0}(y) > 0$. It holds from (S.1.23) that $|a_v(y)|/\delta_v \le \|\zeta_1(y)\|\|v\|^{1/2} + C_0|\zeta_2(y)|\|v\|^{3/2} \to 0$ as $\|v\| \to 0$. Therefore, $1[y \in E_2] \to 0$. By dominated convergence, (S.1.24) holds where we used the integrability of $\|\zeta_1\|^2$. Therefore, for an arbitrary $\eta > 0$, there exists $\varepsilon_3 \in (0, \varepsilon_2)$ such that uniformly for all v with $\|v\| < \varepsilon_3$,

$$\int_{E_2} T_{1,v}(y)^2 p_{x_0}(y)\nu(dy) \le \eta + C_8\|v\|^2\ .$$

Combining the two regions, we have, uniformly for all v with $\|v\| < \varepsilon_3$,

$$\int T_{1,v}(y)^2 p_{x_0}(y)\nu(dy) \le C_6\|v\| + C_7\|v\|^3 + C_8\|v\|^2 + \eta\ .$$

To summarize, it holds that, uniformly for all v with $\|v\| < \varepsilon_3$,

$$\int \Delta_v(y)^2\nu(dy) \le 2\int T_{1,v}^2(y)p_{x_0}(y)\nu(dy) + 2\int T_{2,v}^2(y)p_{x_0}(y)\nu(dy) \le 2C_6\|v\| + C_9\|v\|^2 + 2C_7\|v\|^3 + 2\eta\ .$$

where $C_9 := 2(C_4 + C_8)$. As $\eta \to 0$, the upper bound converges to 0 as $\|v\| \to 0$, so quadratic mean differentiability holds. Since

$$\int \|\dot{\ell}_{x_0}(y)\|^2 p_{x_0}(y)\nu(dy) \le 2\int \|\zeta_1(y)\|^2 p_{x_0}(y)\nu(dy) + 2\|\bar{\zeta}_1\|^2 < \infty\ ,$$

the integrability condition also holds. Therefore, Assumption 3 implies 2.

Finally, unlike Assumption 2, Assumption 3 imposes a support restriction. If $f(x_0, y_*) = 0$ for some y_*, then $f(x_0 + \varepsilon, y_*) = 0$ for all $\|\varepsilon\| \le \varepsilon_0$. Thus, the converse implication does not hold in general. To provide a simple example, let $d = 1$, $x_0 = 0$, and $y \in \{0, 1\}$, and let ν be the counting measure on $\{0, 1\}$. For $x \in \mathbf{R}$ with $|x|$ small, set

$$p_x(1) := |x|^3, \qquad p_x(0) := 1 - |x|^3\ .$$

This is a valid pmf (nonnegative and summing to one) for small $|x|$. To see that Assumption 2 is satisfied, take $\dot{\ell}_{x_0}(y) := 0 \in \mathbf{R}$. Then, writing the $L^2(\nu)$ expression as a sum,

$$\sum_{y\in\{0,1\}} \left[\frac{\sqrt{p_x(y)} - \sqrt{p_{x_0}(y)}}{|x|} - \tfrac{1}{2}\dot{\ell}_{x_0}(y)\sqrt{p_{x_0}(y)}\right]^2 = \left(\frac{\sqrt{1-|x|^3} - 1}{|x|}\right)^2 + \left(\frac{|x|^{3/2}}{|x|}\right)^2\ .$$

Using $\sqrt{1-t} = 1 - \frac{1}{2}t + O(t^2)$ as $t \to 0$ with $t = |x|^3$,

$$\left(\frac{\sqrt{1-|x|^3}-1}{|x|}\right)^2 = \left(\frac{-\frac{1}{2}|x|^3 + O(|x|^6)}{|x|}\right)^2 = O(|x|^4) \to 0\ , \qquad \left(\frac{|x|^{3/2}}{|x|}\right)^2 = |x| \to 0\ .$$

Hence the QMD condition holds. Moreover, $\int \|\dot{\ell}_{x_0}(y)\|^2 \, p_{x_0}(y)\, \nu(dy) = 0 < \infty$. However, at $y = 1$ we have $f(x_0, 1) = p_{x_0}(1) = 0$, but for every small $x \neq 0$, $f(x, 1) = p_x(1) = |x|^3 > 0$, which is a violation of Assumption 3. ■

Appendix S.2 Results under Alternative Assumptions

The main results of the paper are derived under differentiability in quadratic mean (QMD); see Assumption 2. In this appendix, we study an alternative, perhaps more standard differentiability condition: a Taylor expansion of the joint density $f(\cdot, y)$ around x_0 with a Hölder remainder (Assumption S.4), which allows for a wide range of smoothness levels and remainder controls. This condition may appear similar in spirit to the multiplicative expansion in Assumption 3, since the latter is closely related to the case $(\kappa_s, \kappa_r) = (1, 1)$. We will show, however, that Assumption 3 is much stronger: beyond smoothness and integrability, it imposes a restrictive multiplicative structure (and associated local support restrictions) that is not required under QMD or Assumption S.4. We begin by stating the smoothness with Hölder remainder condition that underpins the results in this section.

Assumption S.4 (Smoothness with Hölder remainder at x_0)**.** *Fix $\kappa_s \in \mathbf{N}_0$ and $\kappa_r \in (0, 1]$. There exists $\tilde{r} > 0$ such that, for every $y \in \mathbf{R}^m$ and every t with $\|t\| \le \tilde{r}$ and $x_0 + t \in \mathcal{X}$,*

$$f(x_0 + t, y) = \sum_{\|\beta\|_1 \le \kappa_s} \frac{\partial_x^\beta f(x_0, y)}{\beta!}\, t^\beta + \zeta^{\text{rem}}_{\kappa_s+\kappa_r}(t, y), \qquad |\zeta^{\text{rem}}_{\kappa_s+\kappa_r}(t, y)| \le M(y)\, \|t\|^{\kappa_s+\kappa_r}\ . \tag{S.2.1}$$

Here $\beta = (\beta_1, \ldots, \beta_d)$ is a multi-index,

$$\partial_x^\beta f(x_0, y) := \frac{\partial^{\|\beta\|_1} f}{\partial x_1^{\beta_1} \cdots \partial x_d^{\beta_d}}(x_0, y), \qquad t^\beta := \prod_{j=1}^d t_j^{\beta_j}, \qquad \beta! := \prod_{j=1}^d \beta_j!\ .$$

Moreover,

$$\int M(y)\, \nu(dy) < \infty, \qquad \int \left|\partial_x^\beta f(x_0, y)\right| \nu(dy) < \infty \quad \textit{for all} \quad \|\beta\|_1 \le \kappa_s\ . \tag{S.2.2}$$

Assumption S.4 has several ingredients. First, it requires that the joint density $f(\cdot, y)$ admits a Taylor series expansion of order κ_s around x_0, with a remainder that is controlled by Hölder continuity of order κ_r. Second, the bound on the remainder term through the envelope $M(y)$

allows the quality of approximation to depend on y, but ensures uniform control in t once $M(y)$ is integrable. Finally, the integrability conditions guarantee that both the derivatives and the envelope are well behaved relative to the dominating measure ν, so that expansions can be integrated and used in distributional approximations. Together, these conditions provide a flexible smoothness requirement that interpolates between differentiability of various orders and Hölder continuity.

Several features of Assumption S.4 are worth highlighting.

Monotonicity in κ_{s}. For any fixed $\kappa_{\mathrm{r}} \in (0,1]$, Assumption S.4 becomes stronger as κ_{s} increases. In particular, if it holds with $(\kappa_{\mathrm{s}}+1, \kappa_{\mathrm{r}})$, then it automatically holds with $(\kappa_{\mathrm{s}}, \kappa_{\mathrm{r}})$.

The case $\kappa_{\mathrm{s}} = 0$. When $\kappa_{\mathrm{s}} = 0$, the condition reduces to Hölder continuity,

$$|f(x,y) - f(x_0,y)| \;\leq\; M(y)\,\|x - x_0\|^{\kappa_{\mathrm{r}}} \;,$$

with $\kappa_{\mathrm{r}} \in (0,1]$. In particular, $\kappa_{\mathrm{r}} = 1$ yields (local) Lipschitz continuity.

The case $\kappa_{\mathrm{s}} = 1$. When $\kappa_{\mathrm{s}} = 1$, Assumption S.4 takes the form

$$\left| f(x_0+t,y) - f(x_0,y) - \nabla_x f(x_0,y)^\top t \right| \;\leq\; M(y)\,\|t\|^{1+\kappa_{\mathrm{r}}} \;,$$

i.e., pointwise differentiability of f in x at x_0 with a Hölder remainder of order κ_{r}. Heuristically, while Assumption 2 imposes differentiability in quadratic mean of $x \mapsto \sqrt{p_x}$ at x_0, Assumption S.4 requires pointwise differentiability of the *joint* density. Neither assumption implies the other.

The main result of this section is as follows.

Theorem S.1. *Let Assumptions 1 and S.4 hold. Then,*

$$\mathrm{TV}(P_r, P) = O(r^{\min\{1,\kappa_s+\kappa_r\}}) \qquad \textit{and} \qquad \mathrm{H}(P_r, P) = O(r^{\min\{1/2,(\kappa_s+\kappa_r)/2\}}) \;.$$

Moreover, if x_0 *is an interior point of* $\mathcal{X}$*,*

$$\mathrm{TV}(P_r, P) = O(r^{\min\{2,\kappa_s+\kappa_r\}}) \qquad \textit{and} \qquad \mathrm{H}(P_r, P) = O(r^{\min\{1,(\kappa_s+\kappa_r)/2\}}) \;.$$

Theorem S.1 provides the local approximation rates implied by Assumption S.4. These rates contrast with those obtained under the QMD condition in Theorem 2: the two sets of primitives generate distinct local exponents (a_h, a_{tv}), which then determine the joint rates through Theorem 1. Under Hölder smoothness, the exponents depend on the combined smoothness order $\kappa_{\mathrm{s}} + \kappa_{\mathrm{r}}$ *and* on whether x_0 is an interior point of $\mathcal{X}$. In general,

$$a_h = \min\{1/2, (\kappa_{\mathrm{s}}+\kappa_{\mathrm{r}})/2\} \qquad \text{and} \qquad a_{tv} = \min\{1, \kappa_{\mathrm{s}}+\kappa_{\mathrm{r}}\} \;, \tag{S.2.3}$$

and if x_0 is an interior point of $\mathcal{X}$,

$$a_h = \min\{1, (\kappa_s + \kappa_r)/2\} \qquad \text{and} \qquad a_{tv} = \min\{2, \kappa_s + \kappa_r\} \ . \tag{S.2.4}$$

In both (S.2.3) and (S.2.4), one has $a_{tv} = 2a_h$, so TV converges (at most) twice as fast as Hellinger.

The effect of whether x_0 is an interior point of $\mathcal{X}$ or not is analogous to that in kernel regression estimation under twice continuous differentiability. When x_0 lies in the interior of $\mathcal{X}$, the first moment integrates to zero by symmetry and the second term of the Taylor series expansion becomes the leading term. The same phenomenon explains the differences between (S.2.3) and (S.2.4). As $\kappa_s + \kappa_r$ increases, the Hellinger rate saturates at r (interior case) or at $r^{1/2}$ (non-interior case), while the TV rate saturates at r^2 (interior case) or at r (non-interior case). Under QMD, the guaranteed polynomial exponent is one for both distances at interior and boundary points. At interior points, both distances are $o(r)$, but no larger polynomial exponent is guaranteed by these assumptions.

Remark S.5. *The rates in Theorem S.1 can be shown to be sharp by characterizing DGPs that satisfy the conditions of the theorem and yield the rates described in each case. This sharpness result thus shows that these local approximation rates cannot be improved under Assumption S.4. We omit the details here for brevity.* ■

Combining Theorem S.1 with Theorem 1 in the case with $\kappa_s + \kappa_r = 2$ and x_0 belongs in the interior of $\mathcal{X}$ yields

$$\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{1/d}\big), \qquad \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k(k/n)^{2/d}\big) \ . \tag{S.2.5}$$

Here $a_{tv} = 2$ and $a_h = 1$, so the direct bound $k(k/n)^{a_{tv}/d}$ from Theorem 1 determines the TV rate, rather than the indirect inequality $\mathrm{TV} \leq \sqrt{2}\,\mathrm{H}$. As in the local approximation case, the TV distance converges twice as fast as the Hellinger distance. Moreover, both expressions imply the same growth condition on k for joint convergence:

$$k(k/n)^{2/d} \to 0 \qquad \Longleftrightarrow \qquad k = o\big(n^{2/(2+d)}\big) \ .$$

For $d = 1$ this becomes $k = o(n^{2/3})$, matching the threshold obtained in Section 3.1. If instead $\kappa_s + \kappa_r = 2$ and x_0 does not belong in the interior of $\mathcal{X}$, the local exponents are $a_h = 1/2$ and $a_{tv} = 1$, and Theorem 1 yields the slower joint rates

$$\mathrm{H}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k^{1/2}(k/n)^{1/(2d)}\big), \qquad \mathrm{TV}\big(\mathcal{L}(S_n), \mathcal{L}(S)\big) = O\big(k(k/n)^{1/d}\big) \ ,$$

which require $k = o\big(n^{1/(1+d)}\big)$. More generally, as $\kappa_s + \kappa_r$ decreases, the admissible growth of k slows accordingly, and the slowdown is more pronounced in the non-interior case where the linear term does not cancel.

Remark S.6. *The L^1 integrability requirement $\int M(y)\,\nu(dy) < \infty$ in Assumption S.4 may be strengthened to an L^p condition for some $p > 1$. Such a strengthening does not, by itself, improve*

the convergence rates for $\mathrm{H}(P_r, P)$ *or* $\mathrm{TV}(P_r, P)$ *obtained in Theorem S.1. This parallels the result in Lemma S.5.* ■

S.2.1 Auxiliary Results under Alternative Assumptions

Assumption 3 in Falk et al. (2010) implies the local approximation error $\mathrm{H}(P_r, P) = O(r^2)$, a rate that is substantially faster than the feasible rates delivered by Theorem S.1, which are sharp. Because Assumption 3 is closely related to Assumption S.4 with $\kappa_{\mathrm{s}} = \kappa_{\mathrm{r}} = 1$, it is natural to examine more carefully which aspects of these conditions drive such a discrepancy in rates. To keep the discussion as simple as possible, in this section we focus on the univariate case $d = m = 1$ and x_0 in the interior of the support of X.

The main takeaway is that Assumption 3 imposes a multiplicative structure on the local approximation to $f(x, y)$ (or $p_x(y)$) that automatically restricts the behavior of $f(x, y)$ (or $p_x(y)$) in a neighborhood of the zero–density region

$$\mathcal{N} := \{\, y \in \mathcal{Y} : f(x_0, y) = 0 \,\} \,.$$

Using similar arguments to those in the proof of Theorem S.1, it is possible to show that if Assumption S.4 holds with $\kappa_{\mathrm{s}} = \kappa_{\mathrm{r}} = d = 1$ and $M(y)/f(x_0, y) \in L^2(p_{x_0} \cdot \nu)$, then

$$\mathrm{H}(P_r, P) = \begin{cases} O\big(r^2\big) & \text{if } \nu(\mathcal{N}) = 0 \,, \\ O(r) & \text{otherwise} \,, \end{cases} \tag{S.2.6}$$

so that the baseline linear rate improves to the quadratic rate when the integrability condition in Assumption S.4 is strengthened and $\nu(\mathcal{N}) = 0$ is imposed. These conditions are sufficient but not necessary. More precisely, under the strengthened integrability condition, $\mathrm{H}^2(P_r, P) \leq Cr^4 + \frac{1}{2}P_r(\mathcal{N})$, so any condition ensuring $P_r(\mathcal{N}) = O(r^4)$, such as $\nu(\mathcal{N}) = 0$ or $f(x, y) = 0$ for all $y \in \mathcal{N}$ and x near x_0, yields the quadratic rate. Without such control of the probability entering $\mathcal{N}$, these assumptions do not guarantee an improvement over the baseline rate. Effectively, these additional assumptions close the gap between Assumption 3 and Assumption S.4, yielding the same fast rate. This highlights the central role of the zero–density region $\mathcal{N}$ in determining whether higher integrability translates into faster convergence. The details behind the result in (S.2.6) are omitted for brevity.

S.2.2 Proof of Theorem S.1

By Assumption 1(ii), for any $r \in (0, r_0)$,

$$\mathrm{Vol}(B_r \cap \mathcal{X}) = \int_{B_r \cap \mathcal{X}} dx \geq C_{x_0} r^d > 0 \,. \tag{S.2.7}$$

Let r_{s} denote the constant $\tilde{r}$ in Assumption S.4, and let r_1 be as in Lemma 2. For the rest of the proof, let $r_2 = \min\{r_0, r_{\mathrm{s}}, r_1\} > 0$.

To get the result for $\mathrm{TV}(P_r, P)$, we derive expansions for any $r \in (0, r_2)$. Let

$$\mathcal{I}_{\kappa_s}^{(1)} := \big\{\beta \in \mathbf{N}_0^d : 1 \le \|\beta\|_1 \le \kappa_s\big\}, \qquad \mathcal{I}_{\kappa_s}^{(2)} := \big\{\beta \in \mathbf{N}_0^d : 2 \le \|\beta\|_1 \le \kappa_s\big\}$$

where $\mathbf{N}_0$ is the set of nonnegative integers.

For $\beta \in \mathcal{I}_{\kappa_s}^{(1)}$, define

$$c_\beta := \int \partial_x^\beta f(x_0, y)\, \nu(dy)\ , \qquad \text{so that} \qquad |c_\beta| \le \int |\partial_x^\beta f(x_0, y)|\, \nu(dy) < \infty$$

by (S.2.2). First, we have

$$\begin{aligned}\int_{B_r \cap \mathcal{X}} f(x,y)dx &\overset{(1)}{=} \int_{B_r \cap \mathcal{X}} \Big(f(x_0,y) + \sum_{\beta \in \mathcal{I}_{\kappa_s}^{(1)}} \frac{\partial_x^\beta f(x_0,y)}{\beta!} \prod_{j=1}^d (x_j - x_{0,j})^{\beta_j} + \zeta^{\mathrm{rem}}_{\kappa_s+\kappa_r}(x - x_0, y)\Big)dx \\ &\overset{(2)}{=} \mathrm{Vol}(B_r \cap \mathcal{X}) f(x_0,y) + \sum_{\beta \in \mathcal{I}_{\kappa_s}^{(1)}} \frac{\partial_x^\beta f(x_0,y)}{\beta!} J_\beta(r) + R(r,y)\ , \end{aligned} \tag{S.2.8}$$

where (1) holds by Assumption S.4, and (2) by setting

$$\begin{aligned} J_\beta(r) &:= \int_{B_r \cap \mathcal{X}} \prod_{j=1}^d (x_j - x_{0,j})^{\beta_j} dx \\ R(r,y) &:= \int_{B_r \cap \mathcal{X}} \zeta^{\mathrm{rem}}_{\kappa_s+\kappa_r}(x - x_0, y) dx\ . \end{aligned}$$

Second, for any $r \in (0, r_2)$, we have

$$\begin{aligned}&\int_{B_r \cap \mathcal{X}} g(x)dx = \int_{B_r \cap \mathcal{X}} \int f(x,y)\nu(dy)dx \\ &\overset{(1)}{=} \int_{B_r \cap \mathcal{X}} \Big(g(x_0) + \sum_{\beta \in \mathcal{I}_{\kappa_s}^{(1)}} \frac{c_\beta}{\beta!} \prod_{j=1}^d (x_j - x_{0,j})^{\beta_j} + \int \zeta^{\mathrm{rem}}_{\kappa_s+\kappa_r}(x - x_0, y)\nu(dy)\Big)dx \\ &\overset{(2)}{=} \mathrm{Vol}(B_r \cap \mathcal{X}) g(x_0) + \sum_{\beta \in \mathcal{I}_{\kappa_s}^{(1)}} \frac{c_\beta}{\beta!} J_\beta(r) + R(r)\ , \end{aligned} \tag{S.2.9}$$

where (1) holds by integrating the expansion in Assumption S.4 with respect to $\nu(dy)$, which (S.2.2) justifies, and by $g(x_0) = \int f(x_0, y)\nu(dy)$, and (2) by setting

$$R(r) := \int_{B_r \cap \mathcal{X}} \int \zeta^{\mathrm{rem}}_{\kappa_s+\kappa_r}(x - x_0, y)\ \nu(dy)dx\ .$$

Finally, for any $r \in (0, r_2)$, we have

$$\begin{aligned} h_r(y) - h_0(y) &= \frac{\int_{B_r \cap \mathcal{X}} f(x,y)dx - h_0(y) \int_{B_r \cap \mathcal{X}} g(x)dx}{\int_{B_r \cap \mathcal{X}} g(x)dx} \\ &\overset{(1)}{=} \frac{(\int_{B_r \cap \mathcal{X}} f(x,y)dx - h_0(y) \int_{B_r \cap \mathcal{X}} g(x)dx)/\mathrm{Vol}(B_r \cap \mathcal{X})}{\int_{B_r \cap \mathcal{X}} g(x)dx/\mathrm{Vol}(B_r \cap \mathcal{X})} \\ &\overset{(2)}{=} \frac{\sum_{\beta \in \mathcal{I}_{\kappa_s}^{(1)}} \frac{\partial_x^\beta f(x_0,y) - h_0(y) c_\beta}{\beta!} \frac{J_\beta(r)}{\mathrm{Vol}(B_r \cap \mathcal{X})} + \frac{R(r,y) - h_0(y) R(r)}{\mathrm{Vol}(B_r \cap \mathcal{X})}}{\int_{B_r \cap \mathcal{X}} g(x)dx/\mathrm{Vol}(B_r \cap \mathcal{X})} , \end{aligned} \tag{S.2.10}$$

where (1) holds by (S.2.7), which yields $\mathrm{Vol}(B_r \cap \mathcal{X}) > 0$, and (2) by (S.2.8), (S.2.9), and $f(x_0, y) = h_0(y)g(x_0)$.

For any $x \in \mathcal{X}$ with $\|x - x_0\| \le r$ and $r \in (0, r_2)$, (B-2) in the proof of Lemma 2 and $r_2 \le r_1$ imply $g(x) \ge g(x_0)/2 > 0$. Then, for any $r \in (0, r_2)$,

$$\int_{B_r \cap \mathcal{X}} g(x)dx \overset{(1)}{\ge} (g(x_0)/2)\mathrm{Vol}(B_r \cap \mathcal{X}) > 0 .$$

Then, for all $r \in (0, r_2)$, we get

$$\frac{\int_{B_r \cap \mathcal{X}} g(x)dx}{\mathrm{Vol}(B_r \cap \mathcal{X})} \ge g(x_0)/2 > 0 . \tag{S.2.11}$$

Now there are two cases, depending on whether x_0 is in the interior of $\mathcal{X}$ or not. First, consider the case where x_0 is in the interior of $\mathcal{X}$. By definition, this implies that there exists $r_3 > 0$ such that $B_r \subset \mathcal{X}$ for all $r \in (0, r_3)$. Then, for any $r \in (0, \min\{r_2, r_3\})$, we obtain

$$h_r(y) - h_0(y) \overset{(1)}{=} \frac{\sum_{\beta \in \mathcal{I}_{\kappa_s}^{(2)}} \frac{\partial_x^\beta f(x_0,y) - h_0(y) c_\beta}{\beta!} \frac{J_\beta(r)}{\mathrm{Vol}(B_r \cap \mathcal{X})} + \frac{R(r,y) - h_0(y) R(r)}{\mathrm{Vol}(B_r \cap \mathcal{X})}}{\int_{B_r \cap \mathcal{X}} g(x)dx/\mathrm{Vol}(B_r \cap \mathcal{X})} , \tag{S.2.12}$$

where (1) holds by (S.2.10) and $B_r \subset \mathcal{X}$ for all $r \in (0, r_3)$, and so $J_\beta(r) = \int_{B_r} \prod_{j=1}^d (x_j - x_{0,j})^{\beta_j} dx = 0$ for any $\beta \in \{\{0,1\}^d : \|\beta\|_1 = 1\}$. Then, for any $r \in (0, \min\{r_2, r_3\})$,

$$\begin{aligned} \mathrm{TV}(P_r, P) &= \frac{1}{2} \int |h_r(y) - h_0(y)| \nu(dy) \\ &\overset{(1)}{\le} \frac{\sum_{\beta \in \mathcal{I}_{\kappa_s}^{(2)}} \frac{2 \int |\partial_x^\beta f(x_0,y)| \nu(dy)}{\beta!} \frac{|J_\beta(r)|}{\mathrm{Vol}(B_r \cap \mathcal{X})} + \frac{2 \int |R(r,y)| \nu(dy)}{\mathrm{Vol}(B_r \cap \mathcal{X})}}{g(x_0)} \\ &\overset{(2)}{\le} \sum_{\beta \in \mathcal{I}_{\kappa_s}^{(2)}} O(r^{\|\beta\|_1}) + O(r^{\kappa_s + \kappa_r}) \\ &= O(r^{\min\{2, \kappa_s + \kappa_r\}}) , \end{aligned}$$

where (1) holds by (S.2.12), the triangle inequality, (S.2.11), $\int h_0(y)\nu(dy) = 1$, $h_0(y) \ge 0$, $|c_\beta| \le$

$\int |\partial_x^\beta f(x_0,y)|\nu(dy)$, and $|R(r)| \leq \int |R(r,y)|\nu(dy)$, and (2) by Assumption S.4,

$$|J_\beta(r)| = \Big| \int_{B_r \cap \mathcal{X}} \prod_{j=1}^{d} (x_j - x_{0,j})^{\beta_j} dx \Big| \leq r^{\|\beta\|_1} \mathrm{Vol}(B_r \cap \mathcal{X}) \ ,$$

and

$$\int |R(r,y)|\nu(dy) \leq \int M(y)\, \nu(dy)\ r^{\kappa_s+\kappa_r} \mathrm{Vol}(B_r \cap \mathcal{X}) \ .$$

Next, consider the case where x_0 is not in the interior of $\mathcal{X}$. The sole difference is that (S.2.12) includes the term $J_\beta(r)$ with $\beta \in \{\{0,1\}^d : \|\beta\|_1 = 1\}$, which now becomes

$$h_r(y) - h_0(y) \overset{(1)}{=} \frac{\sum_{\beta \in \mathcal{I}_{\kappa_s}^{(1)}} \frac{\partial_x^\beta f(x_0,y) - h_0(y) c_\beta}{\prod_{s=1}^d \beta!} \frac{J_\beta(r)}{\mathrm{Vol}(B_r \cap \mathcal{X})} + \frac{R(r,y) - h_0(y) R(r)}{\mathrm{Vol}(B_r \cap \mathcal{X})}}{\int_{B_r \cap \mathcal{X}} g(x) dx / \mathrm{Vol}(B_r \cap \mathcal{X})} \ . \tag{S.2.13}$$

By using this expression and repeating previous arguments, we get

$$\mathrm{TV}(P_r, P) = O(r^{\min\{1, \kappa_s+\kappa_r\}}) \ .$$

The convergence rate of the Hellinger distance follows from its relationship with the total variation distance, $\mathrm{H}^2(P_r, P) \leq \mathrm{TV}(P_r, P)$. ■