Exact minimax risk at the breakdown boundary under Huber contamination

Scott Schubr
Luminary Oddities

November 2026

Abstract

We determine the exact limiting minimax risk of estimation as the fraction of corrupted observations approaches one half from below. For two clean laws that become asymptotically disjoint, the minimax risk under independent Huber contamination converges to the Bayes risk of a binary Gaussian shift when the total variation distance between their contamination neighborhoods, multiplied by the square root of the sample size, converges. Write $\delta = 1 - 2\epsilon$ for the clean excess, where ϵ is the contamination fraction, and let d be the dimension or batch size. As d grows, we obtain exact limits when $\log(1/\delta)/d$ converges and, under independent contamination, the sample size n grows and $n\delta^2$ converges. We treat isotropic Gaussian variance with a bounded translation nuisance, regular exponential families on compact convex parameter sets, and categorical laws on $K = o(d)$ symbols observed in batches of size d. The limiting risk under independent contamination is the larger of two terms: one from overlap between competing clean laws and one from random fluctuations in the size of the clean majority. Under a fixed cap on replacements, only the overlap term remains.

For categorical batches, let n be the number of batches, d the batch size, K the alphabet size, and $\ell < n/2$ the replacement cap. Write $m = n - 2\ell$ and $S = \log(n/m)$. Under loss equal to the pth power of total variation distance, the minimax risk is asymptotic to $(S/(2d))^{p/2}$ when S tends to infinity, $S = o(d)$, and $K = o(mS)$. For the independent models studied here, a single estimator attains the known-level limit for absolute loss without knowing the contamination level. For powers above one, we determine the exact additional minimax risk over compact sets of positive critical levels $\tau = \lim n\delta^2$ in the joint critical regime and characterize when that cost is zero.

1 Introduction

Huber's contamination model $(1 - \epsilon)P + \epsilon H$ loses identifiability at $\epsilon = 1/2$. Chen, Gao and Ren [2, Section 3, Theorems 3.1–3.2] showed that, below that point, the minimax risk is governed up to constants by the clean risk and a modulus of continuity in total variation. Peng and Neykov [3] determined the minimax rate, again up to constants, for Gaussian means with a deterministic cap on corruptions throughout $[0, 1/2)$, exhibiting the factor $\log(1/(1-2\epsilon))$. In learning from untrusted batches, introduced by Qiao and Valiant [4] and developed by Jain and Orlitsky [5] and by Chen, Li and Moitra [6], the rates are likewise known up to constants; see Diakonikolas and Kane [8] for the algorithmic theory. We study exact limiting risks when the clean excess $\delta = 1-2\epsilon$ tends to zero. The rate bounds above explain how the worst-case error scales, and leave the limiting constants undetermined in the joint regimes considered here.

Open questions addressed. As the corruption fraction approaches one half, what is the exact limiting minimax risk as the number of observations and the information in each observation vary? How does the answer change if the adversary may replace at most $\ell < n/2$ observations? Finally, what is the cost of estimating without knowing the contamination level? We answer these questions for the models and scalings below.

In the growing-dimension or growing-batch regimes of Section 4, let $c = \lim d^{-1} \log(1/\delta)$ measure the corruption scale relative to dimension or batch size. For independent contamination, let $\tau = \lim n\delta^2$ measure the squared clean-majority signal. The exact risk is the larger of an overlap term and a term from fluctuations in the clean majority. With a fixed replacement cap, the overlap term gives the entire limit. For example, for a target set with $D > 0$, at $c = 0$ and $0 < \tau < \infty$ under absolute loss the capped risk vanishes while the independent risk converges to $D\Phi(-\sqrt{\tau})$, where D is the diameter of the target set and Φ is the standard normal distribution function. This difference determines whether estimation is consistent in that regime.

Main results.

1. *Two-point limit* (Theorem 2.1). For two asymptotically disjoint clean laws, the limit is $\mathcal{B}_p(\lambda)$ with $\lambda = \lim \sqrt{n}(\Delta_n)_+$, where o_n is the overlap of the clean laws and $(\Delta_n)_+$ is the exact total variation distance between their contamination neighborhoods, with $\Delta_n = \delta_n - \frac{1+\delta_n}{2} o_n$. A weighted vote extends this to overlapping laws with ties (Theorem 2.2). Exact overlap prefactors give a parity effect for symmetric Bernoulli batches and a second-order window (Section 2.5).

2. *Fixed cap on binary corruptions* (Theorems 3.1 and 3.2). This gives a Gaussian Bayes risk with signal equal to the positive part of the limit of $(m/2 - ne)/\sqrt{ne}$, and a Poisson convex program whose complete-loss threshold is a single Poisson tail.

3. *Global limits* (Theorems 4.1, 4.2, 4.3, 4.4, 4.5). Here the limit is $\max\{(M(c)/2)^p, D^p B_p(\tau)\}$ under independent contamination and $(M(c)/2)^p$ under a cap, for Gaussian variance, exponential families and categorical batches. We also give consistency thresholds for large alphabets, sharp at $c = 0$ under the cap and at $c = 0,\ \tau = \infty$ in the independent model.

4. *Shrinking categorical risk* (Theorem 5.1). When $S = \log(n/m) \to \infty$ with $S = o(d)$, the risk is of order $\min\{1, ((S + K/m)/d)^{p/2}\}$ for alphabet size K. If also $K = o(mS)$, then $R \sim (S/(2d))^{p/2}$.

5. *Adaptation* (Theorems 6.1, 6.2, 6.3). Exact adaptation holds for absolute loss in the independent models. For higher powers it can fail. We give the exact minimax regret over compact sets of positive critical levels and characterize when it vanishes.

Relation to earlier Luminary work. The article *Sharp minimax rates for Gaussian location mixtures with unknown covariance*, edition 3 [1], Section 7, treats the isotropic variance under Huber contamination for *fixed* n and ϵ (Theorem 7.1 there). It shows that the absolute risk is $(nd)^{-1/2} + \min\{1, \psi(\epsilon)/\sqrt{d}\} + \beta_n(\epsilon)$ up to constants and tends to exactly $(b-a)\beta_n(\epsilon)$ as $d \to \infty$, with $\beta_n(\epsilon)$ the probability of a contaminated majority, and it identifies the corresponding power-loss limit in Section 7.5 there. The present Theorem 4.1 is the complementary joint limit, in which $n \to \infty$ and $\epsilon \to 1/2$ together with d. The finite-n majority term of [1] becomes the Gaussian-shift term $D^p B_p(\tau)$, and a second, Chernoff-overlap term $M(c)$ appears. Both use the swapped-endpoint label experiment. Here it is combined with a two-candidate pilot, which is what makes the overlap

modulus attainable. At fixed n the corresponding binary label experiment has absolute-loss Bayes risk $(b-a)\beta_n(\epsilon)$, with ties counted at half weight. This is the fixed-n Gaussian variance limit of [1]. We use [1] only through citation.

Notation. Throughout, $p \geq 1$ is fixed. For $\lambda \in (0,\infty)$, $\mathcal{B}_p(\lambda)$ is the equal-prior Bayes risk for estimating $b \in \{0,1\}$ under $|x-b|^p$ from $Z \sim N((2b-1)\lambda, 1)$. Thus $\mathcal{B}_1(\lambda) = \Phi(-\lambda)$. For $p > 1$ the Bayes action is $(1+\exp[-2\lambda Z/(p-1)])^{-1}$. We set $\mathcal{B}_p(0) = 2^{-p}$ and $\mathcal{B}_p(\infty) = 0$, and write $B_p(\tau) = \mathcal{B}_p(\sqrt{\tau})$.

2 The two-point limit under independent contamination

This part treats two known clean laws. It isolates the mechanism behind every later limit: the neighborhoods either intersect, or their distance Δ produces a binary Gaussian experiment of signal $\sqrt{n}\Delta$.

2.1 Statement

For each n let $P_{0,n}, P_{1,n}$ be laws on a common space, and observe n iid draws from a member of $\mathcal{H}_{b,n} = \{(1-\epsilon_n)P_{b,n} + \epsilon_n H\}$, $b \in \{0,1\}$. The estimator may use ϵ_n and the two laws. Let $R_{n,p}$ be the minimax expected loss $|\widehat{b}-b|^p$. Put $\delta_n = 1-2\epsilon_n > 0$, $o_n = 1 - \mathrm{TV}(P_{0,n}, P_{1,n})$ and $\Delta_n = \delta_n - \frac{1+\delta_n}{2} o_n$.

Theorem 2.1. *If $n \to \infty$, $\delta_n \to 0$, $o_n \to 0$ and $\sqrt{n}(\Delta_n)_+ \to \lambda \in [0,\infty]$, then $R_{n,p} \to \mathcal{B}_p(\lambda)$. If the targets lie at distance $D > 0$ in a normed space and segment interpolation is allowed, the minimax risk is D^p times this.*

Huber [10] identified least favorable pairs and clipped likelihood ratios for robust tests, and Chen, Gao and Ren [2, Theorem 2.1] use likelihood sign votes throughout the contamination model. Theorem 2.1 evaluates the limiting risk when the clean laws and the level vary together. This includes the second-order regime in which overlap and clean majority are comparable.

2.2 Upper bound by sign votes

Suppress n. Set $\alpha = (1+\delta)/2$ and take densities p_b with respect to $P_0 + P_1$. The measure B with density $\alpha \max(p_0, p_1)$ has mass $\alpha(2-o) = 1+\Delta$. If $\Delta \leq 0$, completing B to a probability law gives a law lying in both neighborhoods, so the midpoint is minimax and $R = 2^{-p}$.

Let $\Delta > 0$ and $A = \{p_1 > p_0\}$. Under endpoint zero the probability of a vote in A is at most $q_0 = \alpha P_0(A) + \epsilon$, and under endpoint one it is at least $q_1 = \alpha P_1(A)$. Since $P_1(A) - P_0(A) = 1-o$, we get $q_1 - q_0 = \Delta$, and both tend to $1/2$. This difference is also the exact total-variation distance between the neighborhoods. It is a lower bound for that distance, and placing contamination proportional to $(p_1-p_0)_+$ at endpoint zero and to $(p_0-p_1)_+$ at endpoint one attains it. With K the vote count and $Z_n = 2(K-n\bar{q})/\sqrt{n}$, $\bar{q} = (q_0+q_1)/2$, use the sign of Z_n at $p=1$ and the logistic action $(1+\exp[-2\sqrt{n}\Delta Z_n/(p-1)])^{-1}$ at $p > 1$. By stochastic ordering each risk is bounded by its value for Bernoulli votes at q_b. The binomial CLT gives $Z_n \Rightarrow N((2b-1)\lambda, 1)$ when $\lambda < \infty$. When $\sqrt{n}\Delta \to \infty$ the score diverges in the correct direction.

2.3 Lower bound from two likelihood regions

Assume $\Delta > 0$, so $o \le 2\delta/(1+\delta)$. With $t = \sqrt{\delta}$ let

$$A_0 = \{p_0 > 0,\ p_1 \le (1-t)p_0\}, \quad A_1 = \{p_1 > 0,\ p_0 \le (1-t)p_1\}, \quad W = (A_0 \cup A_1)^c.$$

On A_b the density of B is αp_b. Off A_b the integral of p_b is at most $o/(1-t)$, so $b_i = B(A_i) \to 1/2$ and $w = B(W) \to 0$. Define $Q_0 = B - \frac{\Delta}{b_1} B|_{A_1}$ and $Q_1 = B - \frac{\Delta}{b_0} B|_{A_0}$. Both have mass one. Since $\Delta/b_i = O(\delta) < t$, each still dominates αp_b, so $Q_b \in \mathcal{H}_{b,n}$. The conditional shapes on A_0, A_1, W agree under Q_0, Q_1. Data can therefore be simulated from a shared erasure label with probability w and, otherwise, a binary label for A_1 with success probability $u_0 = (b_1 - \Delta)/(1-w)$ or $u_1 = b_1/(1-w)$. Given N nonerased labels, write $a = u_1 - u_0$. Then $\log(dQ_1^{\mathrm{lab}}/dQ_0^{\mathrm{lab}}) = a(K - N\bar{u})/(\bar{u}(1-\bar{u})) + O(Na^3)$ uniformly. For $\lambda > 0$ the CLT yields the Gaussian binary limit, and for $\lambda = 0$ Pinsker's inequality and the relative entropy bound $O(Na^2)$ give the midpoint. Since $N/n \to 1$, the lower bound $\mathcal{B}_p(\lambda)$ follows. For normed targets, any action has distances $d_0 + d_1 \ge D$ to the endpoints, and interpolation along the segment attains equality.

2.4 Overlapping clean laws and ties

Drop $o_n \to 0$ and $\delta_n \to 0$. Let $S_i = \{p_i > p_{1-i}\}$ and $b_{i,n} = \alpha_n P_{i,n}(S_i) \to b_i > 0$. On indices with $\Delta_n > 0$ assume the near-tie condition $\alpha_n P_{i,n}\{p_i > p_{1-i} > (1 - \sqrt{\Delta_n})p_i\} \to 0$. Exact ties are unrestricted.

Theorem 2.2. *If $\Delta_n \to 0$ and $\frac{\sqrt{n}(\Delta_n)_+}{2}\sqrt{b_0^{-1} + b_1^{-1}} \to \lambda$, then $R_{n,p} \to \mathcal{B}_p(\lambda)$. In particular this applies to every fixed distinct pair with $\mathrm{TV} = v > 0$ as $\alpha_n \to 1/(1+v)$.*

Proof. For fixed laws the near-tie condition holds by dominated convergence. For the upper bound give each row the score $-1/b_{0,n}$ on S_0, $1/b_{1,n}$ on S_1 and 0 on ties. The extremal contaminations put their mass on S_1 (endpoint zero) or S_0 (endpoint one). This gives cell probabilities $(b_{0,n}, b_{1,n} - \Delta, \cdot)$ and $(b_{0,n} - \Delta, b_{1,n}, \cdot)$, score means $-\Delta/b_{1,n}$ and $\Delta/b_{0,n}$, and limiting variance $b_0^{-1} + b_1^{-1}$. The normalized centered score converges to $N(\pm\lambda, 1)$, and the sign or logistic decision gives the bound. For the lower bound repeat the trimming with $t = \sqrt{\Delta}$. The strong masses tend to b_i, the erasure fraction tends to $1 - b_0 - b_1$, and the informative labels have probabilities tending to $u = b_1/(b_0 + b_1)$ with gap $\Delta/(1-w)$. The signal $\sqrt{N}\Delta/(2(1-w)\sqrt{u(1-u)})$ with $N/n \to b_0 + b_1$ equals λ in the limit. □

2.5 Batch-scale applications: parity and second order

Bernoulli batches. Fix $0 < g < 1$. Take d-fold products of Bernoulli laws with means $(1 \mp g)/2$, and write $C = -\frac{1}{2}\log(1-g^2)$ and $L = \log\frac{1+g}{1-g}$. Under the fair tilt, $o_d = e^{-dC}\mathbb{E}_{\mathrm{Bin}(d,1/2)} e^{-L|K-d/2|}$, an instance of the lattice setting of Bahadur and Ranga Rao [11]. The central binomial point masses are asymptotic to $\sqrt{2/(\pi d)}$, and summing over integer or half-integer offsets gives

$$\sqrt{d}\, e^{dC} o_d \to A_{\mathrm{even}} = \sqrt{2/\pi}\, g^{-1}, \qquad A_{\mathrm{odd}} = \sqrt{2/\pi}\,\sqrt{1-g^2}\, g^{-1}.$$

These are $\coth(L/2)$ and $\mathrm{csch}(L/2)$ times $\sqrt{2/\pi}$. If $n\delta^2 \to \tau < \infty$ and $\delta\sqrt{d}e^{dC} \to a \in (0, \infty)$, Theorem 2.1 gives the signal

$$\lambda = \sqrt{\tau}\Big(1 - \frac{A_{\mathrm{parity}}}{2a}\Big)_+.$$

For $A_{\mathrm{odd}}/2 < a < A_{\mathrm{even}}/2$ and $\tau > 0$, the odd-d subsequence therefore has positive signal, while the even subsequence stays indistinguishable, all within one Chernoff exponent. For total-variation loss between the categorical laws multiply by g^p.

Gaussian scale batches. For variances $0 < t < s$ put $r = ts\log(s/t)/(s-t)$, $x_+ = r/t$, $x_- = r/s$ and $I(x) = (x - 1 - \log x)/2$. The exponent is $C = I(x_+) = I(x_-)$. The overlap is the sum of the upper χ^2_d/d tail at x_+ and the lower tail at x_-. The Stirling form of the χ^2_d/d density and endpoint Laplace integration in the coordinate $u = \log x$, where $dI(e^u)/du = (e^u - 1)/2$, give

$$\sqrt{d}\, e^{dC} o_d \to \frac{1}{\sqrt{\pi}}\Big(\frac{1}{x_+ - 1} + \frac{1}{1 - x_-}\Big).$$

The same signal formula then holds, and under variance loss the risk is multiplied by $(s-t)^p$.

Cancellation window. Write $o_d = Ae^{-dC}d^{-1/2}(1 - \gamma/d + o(d^{-1}))$. For the Bernoulli example, $\gamma_{\text{even}} = g^{-2} - 3/4$ and $\gamma_{\text{odd}} = g^{-2} - 1/4$. This follows from the next Stirling term $-(1/4 + 2v^2)/d$ of the central binomial mass at offset v, together with the geometric averages $2z/(1-z)^2$ and $2z/(1-z)^2 + 1/4$ of v^2, $z = e^{-L}$. For the Gaussian example each tail has correction $\gamma(x) = 1/6 + 2x/(x-1)^2$ at $x = x_\pm$, and γ is their average with weights $1/|x_\pm - 1|$. If $\delta = (A/2)e^{-dC}d^{-1/2}(1 + \beta/d + o(d^{-1}))$ and $n\delta^2/d^2 \to \tau < \infty$, then $\Delta/\delta = (\beta + \gamma)/d + o(d^{-1})$, so

$$\lambda = \sqrt{\tau}(\beta + \gamma)_+.$$

All four prefactors and all four corrections were confirmed numerically to the displayed precision (see the verification record).

3 A fixed cap on binary corruptions

This part replaces independent contamination by an adversary who sees the data. Only the clean errors fluctuate, which changes the critical scale.

For each n the clean observations are iid Bernoulli with mean $e = e_n$ at $b = 0$ and $1 - e$ at $b = 1$. An adversary who sees the full clean sequence, may randomize and may depend on b replaces at most $\ell < n/2$ coordinates. Let $m = n - 2\ell$ and let $R^{\text{cap}}_{n,p}$ be the minimax loss. The same risk arises for any observations with two mutually singular conditional laws given the binary symbol, when these conditional laws are common to both endpoints. For the upper bound record the symbol. For the lower bound corrupt symbols and resample the observation from its conditional law.

Theorem 3.1. *If $e \to 0$, $ne \to \infty$ and $(m/2 - ne)/\sqrt{ne} \to \xi \in [-\infty, \infty]$, then $R^{\text{cap}}_{n,p} \to \mathcal{B}_p(\xi_+)$.*

Theorem 3.2. *If $ne \to \Lambda \in [0, \infty)$ and $m = k \geq 1$ eventually, let $Y \sim \text{Poisson}(\Lambda)$ and $w_j = \mathbb{P}(\min(Y, k) = j)$. Then*

$$\lim R^{\text{cap}}_{n,p} = \min_{0 \leq a_0 \leq \cdots \leq a_k \leq 1} \tfrac{1}{2} \sum_{j=0}^{k} w_j \big[a_j^p + (1 - a_{k-j})^p\big].$$

For $p = 1$ this equals

$$\min\{1/2,\ \mathbb{P}(Y \geq h+1)\}\ (k = 2h+1), \qquad \min\{1/2,\ \mathbb{P}(Y \geq h+1) + \tfrac{1}{2}\mathbb{P}(Y = h)\}\ (k = 2h).$$

For every $p \geq 1$ the limit equals 2^{-p} exactly when this absolute-loss value is $1/2$. For $k = 1$, $u = e^{-\Lambda}$: the limit is 2^{-p} if $u \leq 1/2$; otherwise it is $1 - u$ at $p = 1$ and, at $p > 1$,

$$\big[u^{-1/(p-1)} + (1-u)^{-1/(p-1)}\big]^{1-p}.$$

Jin and Lai [14, Theorem 1] give the finite equal-prior threshold value for bounded perturbations on a discrete line, Tondi, Merhav and Barni [15] give testing exponents under distortion-constrained corruption, and Pydi and Jog [16] characterize zero–one adversarial risk by optimal transport. Theorems 3.1 and 3.2 evaluate the expected power loss in the two count regimes, including nonzero limits and integer caps.

Reduction to counts. By convexity and exchangeability one may restrict to functions a of the corrupted one-count. With $X \sim \mathrm{Bin}(n, e)$, the reachable counts are $[\max(0, X - \ell), \min(n, X + \ell)]$ at $b = 0$ and the reflection at $b = 1$. A nondecreasing a has worst risks $\mathbb{E}a(\min(n, X + \ell))^p$ and $\mathbb{E}[1 - a(\max(0, n - X - \ell))]^p$. Averaging $a(k)$ with $1 - a(n - k)$ equalizes them, so the optimal monotone prior average bounds the minimax risk above. Conversely, given any a, define its nondecreasing lower envelope $a_\uparrow(k) = \min_{j \geq k} a(j)$. At $b = 0$ the reachable right endpoint $k = \min(n, X + \ell)$ gives loss $a(k)^p \geq a_\uparrow(k)^p$. At $b = 1$, on $\{X \leq \ell\}$ every count from $\max(0, n - X - \ell)$ to n is reachable, so the adversary can attain the loss given by the lower envelope at the left endpoint. The equal-prior risk of a is therefore bounded below by that of the monotone envelope less $\mathbb{P}(X > \ell)/2$. The minimax risk is at least the monotone optimum less this error. □

Proof of Theorem 3.1. For the upper bound use the midpoint if $\xi \leq 0$. Otherwise apply the Gaussian Bayes action to $Z = (K - n/2)/\sqrt{ne}$. At the extremal count for $b = 0$, $Z \Rightarrow N(-\xi, 1)$ when $0 < \xi < \infty$, and clipping is negligible. For $\xi = +\infty$, the majority rule has vanishing risk by Chebyshev's inequality, which also gives the exact limit by nonnegativity. For the remaining lower bounds use the reduced cap $\ell' = \min\{\ell, \lfloor n/2 - ne \rfloor\}$. Under endpoint zero choose uniformly ℓ' clean zeros to turn into ones, and under endpoint one choose ℓ' clean ones to turn into zeros. Use no replacements if there are too few eligible symbols, an event of probability $o(1)$ because $e \to 0$ and $\ell' \leq n/2$. The resulting laws are exchangeable and their one-counts are $X + \ell'$ and $n - X - \ell'$ apart from the exceptional event. The normalized counts lie on a shared lattice of mesh $(ne)^{-1/2}$ and have center shifts converging to $\pm\xi_+$. The binomial local limit theorem on $|X - ne| \leq C\sqrt{ne}$ and then $C \to \infty$ show convergence in total variation to the two rounded Gaussian laws on this lattice. Their binary Bayes risk tends to the continuous Gaussian risk: rounding cannot decrease the Bayes risk, and applying the Gaussian Bayes action to the rounded observation gives the matching upper bound by bounded convergence. For $p = 1$ the decision has at most one discontinuity, which is a null set for the Gaussian laws. □

Proof of Theorem 3.2. Here $\ell \to \infty$ and $\mathbb{P}(X > \ell) \to 0$, and the binomial count converges to Y in total variation. After subtracting ℓ, the extreme counts converge to Y and $k - Y$. Clamping a monotone action to $\{0, \ldots, k\}$ only helps, which gives the convex program, and uniform approximation lets the minimum pass to the limit. At $p = 1$ every monotone $[0, 1]$ sequence is a mixture of thresholds, so a threshold j is optimal, with objective $\frac{1}{2}T_j$ for $T_j = \mathbb{P}(Y \geq j) + \mathbb{P}(Y \geq k + 1 - j)$ (truncated law). Writing $p_j = \mathbb{P}(Y = j)$ for the untruncated Poisson probabilities, the endpoint increment is $T_1 - T_0 = \mathbb{P}(Y \geq k) - p_0 \geq p_k - p_0$. For $1 \leq j < k/2$, $T_{j+1} - T_j = p_{k-j} - p_j$. Its sign is that of Λ minus the geometric mean of $j + 1, \ldots, k - j$, which is nondecreasing in j by pairing factors equally distant from their midpoint. Thus the increments on the lower half can switch from positive to negative only. Reflection gives the same values on the upper half. The minimum is at a middle threshold or a constant action. At the constant action $1/2$, every feasible directional derivative of the power objective is a positive multiple of the absolute-loss one. By convexity, $1/2$ is optimal for power p exactly when it is optimal for $p = 1$. For $k = 1$ reflection gives $a_0 = x = 1 - a_1$ with $x \leq 1/2$, and the objective $ux^p + (1 - u)(1 - x)^p$ yields the formula, with transition at $\Lambda = \log 2$. □

4 Global limits for continuous parameters

This part passes from two points to continuous parameter sets, where the overlap modulus $M(c)$ enters. The engine is a pilot that reduces arbitrary contamination to at most two candidates, followed by one vote whose clean error is $o(\delta)$.

4.1 Statements

Let $d \to \infty$ and $\delta \downarrow 0$ with

$$d^{-1}\log(1/\delta) \to c \in [0,\infty], \qquad n\delta^2 \to \tau \in [0,\infty],$$

and $n \to \infty$ in the independent model. In the cap model, an adversary observing n clean iid points or batches replaces at most $\ell < n/2$ of them, and $\delta = (n-2\ell)/n$.

Gaussian variance. Fix $R > 0$ and $0 \le a < b$, and put $D = b-a$. Clean points are $U + \sqrt{t}Z$ with $t \in [a,b]$, $Z \sim N(0, I_d)$, and U independent of Z in the radius-R ball, with a law that may depend on d. Let $I(x) = (x - 1 - \log x)/2$ and $J(u) = I(u\log u/(u-1)) = I(\log u/(u-1))$ for $u > 1$. This J is strictly increasing from $J(1+) = 0$ to ∞. Put $u_c = J^{-1}(c)$ and $M(c) = \min\{D, b(1-u_c^{-1})\}$.

Theorem 4.1. *Under independent contamination the minimax power risk tends to*

$$\max\{(M(c)/2)^p,\ D^p B_p(\tau)\}.$$

This also holds when U is restricted to at most k atoms or to $U = 0$.

Theorem 4.2. *Under the fixed cap, with only $d^{-1}\log(1/\delta) \to c$ assumed, the limit is $(M(c)/2)^p$.*

Exponential families. Let P_θ have densities $\exp(\theta\cdot T - A(\theta))$, with A finite near a compact convex $\Theta \subset \mathbb{R}^q$ and $\nabla^2 A \succ 0$ there, and let $v : \Theta \to E$ be continuous into a normed space. A clean batch is $P_\theta^{\otimes d}$. Write $C(\theta,\vartheta) = \max_{s\in[0,1]}[(1-s)A(\theta) + sA(\vartheta) - A((1-s)\theta + s\vartheta)]$, $D = \max\|v(\theta) - v(\vartheta)\|$ and $M(c) = \max_{C(\theta,\vartheta)\le c}\|v(\theta) - v(\vartheta)\|$.

Theorem 4.3. *The minimax risk under $\|\widehat{v} - v(\theta)\|^p$ tends to $\max\{(M(c)/2)^p, D^p B_p(\tau)\}$ under independent contamination and to $(M(c)/2)^p$ under a cap.*

Theorem 4.4. *For arbitrary laws on $K = K_d$ symbols with $2 \le K = o(d)$ and loss $\mathrm{TV}(\widehat{P}, P)^p$, the same two limits hold with $D = 1$ and $M(c) = \sqrt{1 - e^{-2c}}$, attained by an estimator using polynomially many operations in n, d, K.*

Theorem 4.5. *The independent limit in Theorem 4.4 holds whenever $K = o(nd)$, and the cap limit whenever $K = o(dn\delta)$, by a finite search over a net. These conditions are sharp at $c = 0$ under the cap, and at $c = 0$, $\tau = \infty$ in the independent model. For constants $c_p > 0$ the minimax risks are at least $c_p \min\{1, (K/(dn\delta))^{p/2}\}$ under the cap and $c_p \min\{1, (K/(nd))^{p/2}\}$ in every model.*

Qiao and Valiant [4, Theorems 1.1–1.2] gave an $O(\epsilon/\sqrt{d})$ upper bound at small contamination with sufficiently many batches, and a matching-scale lower obstruction using symmetric Bernoulli witnesses (Lemmas 2.1 and B.1 of the arXiv version). Jain and Orlitsky [5] gave polynomial-time procedures with optimal rates up to constants, with explicit constants for adversarial fractions at most 0.4 in their Theorems 1–2. Theorems 4.3–4.5 identify the exact limiting constant as $\epsilon \to 1/2$ at exponential scale in the batch size, and they separate the independent and capped models at that level.

4.2 Lower bounds

Overlap. If clean one-batch laws with target gap L have overlap o_d with $o_d/\delta \to \infty$, then $(1-\epsilon)(2-o_d) \le 1$ eventually. The neighborhoods intersect, and the risk is at least $(L/2)^p$. Under a cap, maximally couple the n clean pairs independently. The agreement count has mean $no_d \gg n\delta \ge 1$, so at least $n - 2\ell$ pairs agree with probability $1 - o(1)$. Split the disagreements into two sets of size at most ℓ and replace one side on each set by its counterpart. Disintegrating this construction gives valid randomized adversaries whose outputs coincide with probability $1 - o(1)$, and the risk is again at least $(L/2)^p - o(1)$.

The overlap exponents are as follows. For Gaussian variances $t < s$ the densities cross at squared radius $dr(t,s)$, $r = ts\log(s/t)/(s-t)$, and $-d^{-1}\log o_d \to J(s/t)$: the upper bound comes from exponential Markov with $\mathbb{E}e^{vZ_1^2} = (1-2v)^{-1/2}$ and the lower bound from tilting to a nearby variance (see also Nielsen [12, Proposition 11]). For exponential families the exponent is $C(\theta,\vartheta)$, via $\min(x,y) \le x^{1-s}y^s$ and the tilted law $P_{(1-s)\theta+s\vartheta}$. Shortening a segment at one endpoint strictly lowers C, so gaps up to $M(c)$ are approached by pairs of exponent below c. For categorical laws use symmetric Bernoulli pairs $(1 \mp g)/2$ on two symbols, with exponent $-\frac{1}{2}\log(1-g^2)$. In the Gaussian case the pairs $s = b$, $t \uparrow b/u_c$ give gaps approaching $M(c)$.

Label term. In the independent model swap clean and contaminant laws of a diameter pair. The data are a parameter-free randomization of binary labels with upper-label probability $(1 \mp \delta)/2$, and the risk is at least D^p times the label Bayes risk. With log odds $\Lambda = (2K_1 - n)\log\frac{1+\delta}{1-\delta}$ the Bayes decision is the majority rule for $p = 1$ (with midpoint on ties) and $g_{n,\delta}(K_1) = (1 + e^{-\Lambda/(p-1)})^{-1}$ for $p > 1$. The binomial CLT gives the limit $B_p(\tau)$ for $0 < \tau < \infty$. Pinsker's inequality gives 2^{-p} at $\tau = 0$, and Chebyshev's inequality gives 0 at $\tau = \infty$.

4.3 The two-candidate reduction

Suppose a per-batch proxy lies within $\eta_d \to 0$ of the truth with probability $1 - o(1)$, uniformly. Use a pilot of $\lfloor\sqrt{n}\rfloor$ rows (independent model) or all rows (cap). Mark a center, either a point of a finite η_d-net or an observed proxy, if its $2\eta_d$-ball holds at least $2/5$ of the proxies. Choose marked centers greedily, adding a center only when its distance to each previously selected center exceeds $4\eta_d$. The selected balls are disjoint, so at most two centers are selected, and one lies within $5\eta_d$ of the truth with probability $1 - o(1)$. Under independent contamination this holds because the near-true mass is at least $1/2 - o(1)$. Under the cap, at least $n - \ell - o_{\mathbb{P}}(n)$ near-true proxies survive. With a slack $\kappa > 0$ the procedure is as follows. With one center, return its target. With two centers whose targets are within $M(c) + \kappa$, return the midpoint, with error at most $(M(c) + \kappa)/2 + o(1)$. With a far pair, vote.

For a far pair we exhibit a batch vote with clean wrong-side probability $e_d \le e^{-d(c+\gamma)}$ for a fixed $\gamma > 0$, so $e_d = o(\delta)$. In the independent model votes on the $m = n - n_0$ fresh rows are iid. Their success probability is at most $(1-\delta)/2 + e_d$ when center zero is near the truth and at least $(1+\delta)/2 - e_d$ otherwise. The label Bayes action, which is monotone and reflection-symmetric, then has risk bounded by D^p times the comparison-binomial risk. For $\tau < \infty$, $\sqrt{m}e_d = \sqrt{m}\,\delta(e_d/\delta) \to 0$, so the comparison risk converges to $B_p(\tau)$. For $\tau = \infty$, its mean vote fraction has the correct majority with margin at least $\delta/2 - e_d$. Since $m\delta^2 \to \infty$, Chebyshev's inequality gives the correct signed count gap of order $m\delta$ with probability $1 - o(1)$. The sign decision at $p = 1$ and the Bayes interpolation at $p > 1$ both have vanishing risk, since in the latter case the signed log odds have magnitude of order $m\delta^2$ on that event. Under the cap, take an ordinary majority. It suffices that the original clean mistakes are at most $n\delta/4$ simultaneously for all admissible pairs, since then the wrong votes number at most $\ell + n\delta/4 < n/2$. This holds by Markov's inequality and a union bound

when every wrong-vote event is contained in one of $e^{o(d)}$ deterministic events of probability at most e_d. Letting $\kappa \downarrow 0$ slowly, with the finite-d index $d^{-1}\log(1/\delta)$ in place of c, attains the constants.

4.4 The votes

Gaussian variance. The proxy is $\|X\|^2/d$ clipped to $[a, b]$, with $\eta = d^{-1/4}$. Before clipping, its mean differs from t by at most R^2/d and its variance is $O(1/d)$. Clipping can only decrease its distance from t. For far representatives $t_0 < t_1$ set $t_0' = t_0 + \kappa/8$ and $t_1' = t_1 - \kappa/8$ and vote on $\|X\|^2/d > r(t_0', t_1')$. Since the smallest ratio at gap g is $b/(b-g)$, the condition $M(c) < D$ gives $J(t_1'/t_0') \geq c + \gamma_\kappa$. The bounds $\sqrt{t}\|Z\| - R \leq \|X\| \leq \sqrt{t}\|Z\| + R$ move the thresholds by $O(d^{-1/2})$, and exponential Markov gives $e_d \leq e^{-d(c+\gamma_\kappa/2)}$. Under the cap, the possible cutoffs come from a deterministic mesh, and the extremal cutoffs (smallest for upper-tail errors, largest for lower-tail errors) dominate every admissible one.

Exponential families. The proxy is the nearest point to the batch mean of T in a $d^{-1/4}/4$-net of $\nabla A(\Theta)$, with metric $|\nabla A(\theta) - \nabla A(\vartheta)|$ (equivalent to the Euclidean one). The vote is the sign of the total log likelihood ratio of the centers. At the Chernoff parameter s_* the one-observation log-mgf near center zero is $-C(\theta_0, \theta_1) + o(1)$ uniformly, and compactness supplies the margin γ. There are polynomially many pairs, so the union bound is available under the cap.

Categorical laws. The proxy is the batch histogram, with $\mathbb{E}\,\mathrm{TV} \leq \frac{1}{2}\sqrt{K/d}$, and $\eta_d = (K/d)^{1/4}$. For centers P_0, P_1 let $A = \{P_1 > P_0\}$, $x = P_0(A) + \kappa/8$ and $y = P_1(A) - \kappa/8$, and vote by the Bernoulli likelihood comparison of the frequency of A. For $0 < x < y < 1$ with $g = y - x$,

$$-\log\bigl(\sqrt{xy} + \sqrt{(1-x)(1-y)}\bigr) \geq -\tfrac{1}{2}\log(1-g^2),$$

because at fixed g the affinity is concave in x and symmetric about $(1-g)/2$ (compare Sason [13, Proposition 2]). Exponential Markov with powers $\pm 1/2$ gives $e_d \leq (1-g^2)^{d/2}$. Under the cap the extremal cutoffs over the finitely many histogram pairs are indexed by A and the tail side, giving at most $2^{K+1} = e^{o(d)}$ events. The algorithm computes n histograms, at most n^2 distances and one vote, and it never enumerates subsets.

4.5 Larger alphabets and consistency thresholds

Proof of Theorem 4.5. Put $r = K/(nd)$ and $\eta = \max\{d^{-1/4}, r^{1/8}\}$, so that $d\eta^2 \to \infty$ and $K\log(3/\eta) = o(nd\eta^2)$. Take a total-variation η-net of the simplex with at most $(1 + 2/\eta)^K$ points. For each pair at distance above 4η fix a midpoint vote on its sign subset. By Hoeffding's inequality its correct side fails with probability at most $e^{-2d\eta^2}$ for laws within η of a center. Mark a center if, for every pair of rivals more than 4η away, it beats both on at least $2/5$ of the rows. Use the pilot $n_0 = \lceil \sqrt{n} + \sqrt{nK\log(3/\eta)/(d\eta^2)} \rceil = o(n)$ in the independent model and all rows under the cap. A near-true center is marked. In the independent model condition on the pilot clean labels; with probability $1 - o(1)$ at least $0.45n_0$ labels are genuine, by a binomial tail bound. Given such labels, failure to mark the near-true center requires at least $0.05n_0$ genuine losses in one of its prescribed events. A clean row loses with probability at most $2e^{-2d\eta^2}$, so the binomial upper tail bounds each event by $e^{-c_0 n_0 d\eta^2}$. There are at most the cube of the net size of such events, and $K\log(3/\eta) = o(n_0 d\eta^2)$ makes the union vanish. Under the cap a failure requires more than $0.1n$ losses on the original clean rows in a prescribed event, since at most $\ell < n/2$ successes were removed, and the same bound applies with n. Three pairwise separated marked centers would require three disjoint events each of mass $2/5$, which is impossible. The far votes are as before. Under the cap a fixed extreme event has more than $n\delta/4$ clean mistakes with probability at most $(4ee_d/\delta)^{n\delta/4} \leq e^{-\gamma' dn\delta}$, so the union over 2^{K+1} events vanishes when $K = o(dn\delta)$.

For the lower bounds, the clean bound comes from the uncontaminated submodel with $N = nd$ items. Divide $2\lfloor K/2 \rfloor$ symbols into pairs of equal total mass and perturb each split with an independent fair sign and amplitude $h = c\min(1, \sqrt{K/N})$. Flipping one sign changes the N-sample relative entropy by at most CNh^2/K, so for small c the corresponding total variation is at most $1/2$, and by the two-point testing bound conditional on the other signs each sign is misestimated with prior-averaged probability at least $1/4$. A wrong ordering of a pair costs at least a constant times h/K in total variation, so the expected total variation is at least a constant times h, and Jensen's inequality gives Bayes power risk at least $c_p h^p$. Under the cap use the posterior symmetrization of Peng and Neykov [3]. Reserve $n - 2\ell = n\delta$ clean batches Z. After seeing Z, draw P' from the posterior of P under the sign-pair prior. Divide the remaining 2ℓ positions into two blocks of ℓ, and let a fair coin choose one block to be replaced by fresh independent P' batches; this is an admissible corruption. Given Z, the parameters P, P' are independent with the same posterior, and the corrupted data have a law invariant under swapping their roles, so the risk is at least $2^{-p}\mathbb{E}\,\mathrm{TV}(P, P')^p$, which, since drawing a posterior parameter is a randomized estimator based on Z, is at least the clean Bayes risk with $N = dn\delta$ items. □

5 Shrinking risk from nearly half corrupted categorical batches

Under a fixed cap, the preceding limits vanish when $c = 0$. This part finds the leading constant of the vanishing risk under the cap, which requires exponential rather than in-probability control and a multiscale aggregation.

Let P be any law on $\{1, \dots, K\}$, with n clean $P^{\otimes d}$ batches and at most $\ell < n/2$ replaced. Put $m = n - 2\ell$ and $S = \log(n/m)$.

Theorem 5.1. *If $S \to \infty$ and $S = o(d)$, then*

$$R_{n,d,K,p} \asymp_p \min\Big\{1, \Big(\frac{S + K/m}{d}\Big)^{p/2}\Big\},$$

and if additionally $K = o(mS)$ then $R_{n,d,K,p} \sim (S/(2d))^{p/2}$. The constant is attained by a finite search with multiscale aggregation. A deterministic histogram procedure attains it in polynomial time when $K = o(S)$, and randomly pooled histograms attain it in randomized polynomial time when $K = o(S\log(nd))$ (within $K = o(mS)$).

Peng and Neykov [3] obtain the analogous near-breakdown order for Gaussian means under a cap. Theorem 5.1 gives the categorical order and, in the overlap-dominated range $K = o(mS)$, its exact constant.

Lower bounds. Let $r_* = \sqrt{S/(2d)}$, $0 < \kappa < 1$, and take symmetric Bernoulli laws with gap $g = (1-\kappa)\sqrt{2S/d}$. Then $o_d = (1-g^2)^{d/2}\mathbb{E}_{\mathrm{fair}}e^{-|L|/2}$, where the batch log likelihood ratio L has variance $O(S)$, so $|L| \le S^{3/4}$ with probability $1 - o(1)$. Since $dg^4 = o(S)$, $o_d \ge e^{-(1-\kappa)^2 S - o(S)} \gg \delta$. The coupling of Section 4 gives risk at least $(1 - o(1))(1 - \kappa)^p r_*^p$. The term $c_p \min\{1, (K/(md))^{p/2}\}$ is the cap bound of Theorem 4.5. □

One scale. Fix $0 < \alpha < 1/50$, a radius $r < 1$, and $\eta = \alpha r$. Suppose selection has returned at most two candidates, one within 5η of P. Return the single candidate, or the midpoint if the gap is at most $2(r-5\eta)$, with error at most r. Otherwise, for $A = \{P_1 > P_0\}$, vote with Bernoulli parameters $x = P_0(A) + 6\eta$ and $y = P_1(A) - 6\eta$, so that $y - x > 2r - 22\eta$. By the affinity inequality the clean

wrong-vote probability is at most $e(r) = \exp[-2(1-11\alpha)^2 dr^2]$. Group the possible wrong-vote events by A and tail side and dominate each group by its extreme threshold. Then

$$\mathbb{P}\{\text{clean mistakes} > m/4\} \le (4e\, e(r)/\delta)^{m/4}$$

for each of at most 2^{K+1} events. With at most $m/4$ clean mistakes, fewer than $n/2$ votes are wrong. Selection uses the finite net and the pair/triple marking of Theorem 4.5. Fix one net point Q within η of P. Against any rival separated by more than 4η, Q's midpoint vote has clean margin at least η and loses with probability at most $e^{-2d\eta^2}$, by Hoeffding. Group the possible losses of Q by subset and tail side, as in the far vote, and dominate each group by its extreme threshold; there are at most 2^{K+1} groups, each with the same probability bound. If every group has fewer than $0.04n$ losses on the original clean rows, Q jointly beats any pair of its rivals on more than $0.92n$ rows, so it is still marked after $\ell < n/2$ replacements. A binomial tail bound then shows that, for $d\eta^2$ large, selection fails with probability at most $2^{K+1}e^{-c_0 nd\eta^2}$ for an absolute $c_0 > 0$. When $\eta \ge \sqrt{K/d}$, histograms can serve as centers: the bounded-difference inequality gives per-row error probability $e^{-d\eta^2/2}$. For pooled histograms, at each scale take $q = \lceil C_2 K/(d\eta^2)\rceil \le n/6$ and $N_q = \lceil C_3 3^q dr^2\rceil$ random q-subsets of rows. Any admissible corruption leaves an untouched fraction of at least 3^{-q} of the q-subsets, and on the original data all but a fraction $3^{-q}/2$ of the pooled histograms are η-accurate except with probability $2\cdot 3^q e^{-cqd\eta^2}$. So a near-true candidate is sampled except with probability $O(e^{-c_\alpha dr^2})$. Run the same pairwise and triple-vote marking on this list. For the selection proof take the cutoff extrema over all possible pooled histograms within η of P and all their rivals, grouped by subset and side; this set is finite at a fixed scale and each extreme has clean loss probability at most $e^{-2d\eta^2}$, so the preceding union marks any near-true candidate that is sampled. The far-vote extrema and union apply to this candidate set as well, and the one-scale exception bound gains only a term $O(e^{-c_\alpha dr^2})$. In the exact regime $q = O(1 + K/S)$, so $K = O(S\log(nd))$ keeps $q = O(\log(nd))$ and N_q polynomial. □

Expected risk. Fix $\kappa > 0$, $r_0 = (1+\kappa)r_*$ and α with $(1-11\alpha)^2(1+\kappa)^2 > 1$. For $r \ge r_0$, $\log(4e\, e(r)/\delta) \le -c_\kappa dr^2$. Since $K = o(mS)$, the factor 2^{K+1} is absorbed and $\mathbb{P}\{\mathrm{TV}(A_r, P) > r\} \le Ce^{-c'_\kappa dr^2}$ uniformly for $r_0 \le r < 1$. For the order bound use a small fixed α and take $r_0 = C_0\sqrt{(S + K/m)/d}$. If $r_0 \ge 1$ the bounded loss proves the bound. Otherwise choose C_0 large enough that S and K are absorbed into the far-vote exponent $c\, mdr^2$ and the selection exponent $c_0 nd\eta^2$ uniformly for $r \ge r_0$; the finite-net procedure then obeys the same exponential estimate. Run the scales $r_j = 2^j r_0$ up to the first $r_J \ge 1$ and select the smallest j with $\mathrm{TV}(A_{r_j}, A_{r_k}) \le r_j + r_k$ for all $k > j$. If all estimates from i on are accurate, the error is at most $3r_i$. Hence the risk is at most $r_0^p + \sum_i (3r_{i+1})^p Ce^{-c' dr_i^2} = r_0^p(1 + o(1))$ because $dr_0^2 \to \infty$. Let $\kappa \downarrow 0$. For the polynomial constructions take $\kappa = a_d$ and $\alpha = a_d/100$ with

$$a_d = \max\Big\{\frac{K}{S\log(nd)}, \frac{K}{mS}, \frac{1}{S}\Big\}^{1/4} \to 0.$$

Use a bounded fallback if a_d exceeds a fixed small constant. Then $a_d^2 S \to \infty$, $K = o(a_d mS)$ and $K = o(a_d^2 nS)$. The near-center and pool exponents are at least a fixed multiple of $a_d^2 dr^2$, and the far-vote exponent before the subset union is at least a fixed multiple of $a_d mdr^2$; the unions 2^{K+1} are absorbed by the preceding inequalities, and the multiscale sum still vanishes relative to r_0^p since $a_d^2 dr_0^2 \to \infty$. Moreover $q \le 1 + Ca_d^2 \log(nd)$ and $q \le 1 + Ca_d^2 m$, so the runtime is polynomial and $q \le n/6$. For the deterministic histogram procedure include K/S in the maximum defining a_d when $K = o(S)$; this gives $K = o(a_d^2 S)$, so $\eta \ge \sqrt{K/d}$ as its one-scale histogram bound requires. There are $O(\log d)$ scales, each polynomial. □

6 Adaptation to the contamination level

The constructions above use ϵ. This part studies whether the known-level risk can be attained without ϵ. Under absolute loss one procedure attains it throughout the stated regimes. For higher powers the same requirement can impose an exact, quantifiable penalty.

Consider the independent models of Theorems 4.1, 4.3 and 4.5 (categorical with $K = o(nd)$), with $n, d \to \infty$.

Theorem 6.1. *In each model one absolute-loss estimator that does not use ϵ, c or τ attains $\max\{M(c)/2, D\Phi(-\sqrt{\tau})\}$ along every sequence. For categorical laws with $K = o(d)$ it uses polynomially many operations.*

Theorem 6.2. *Suppose the target set has diameter $D > 0$. Fix $p > 1$, $0 < \tau_1 < \tau_2 < \infty$, and $\log n = o(d)$. In each model no estimator that does not use the level attains the known-level limit along both $\delta_j = \sqrt{\tau_j/n}$. For categorical laws $K = 2$ already suffices.*

Chen, Gao and Ren [2] and Liu and Gao [9] obtain adaptation at the level of rates, and Jain, Orlitsky and Ravindrakumar [7] give a general meta-technique that removes knowledge of the corruption level at a constant factor (2 in their Lemmas 1–2). Theorems 6.1–6.3 concern equality of limiting constants.

Proof of Theorem 6.1. The pilots of Section 4.3 do not use ϵ. Each candidate pair carries a computable bound e on its clean wrong-vote probability. Pairs with target gap at most $12\eta_d$ return the midpoint before the vote is defined. In the Gaussian case set $z_0 = (\max\{0, \sqrt{r} - R/\sqrt{d}\})^2/x$ and $z_1 = (\sqrt{r} + R/\sqrt{d})^2/y$, for $x = t_0 + 6\eta_d$, $y = t_1 - 6\eta_d$ and $r = xy\log(y/x)/(y - x)$. Use $e = \exp[-d\min\{I(z_0), I(z_1)\}]$ when $z_0 > 1 > z_1$, and use $e = 1$ otherwise. The strict tail conditions hold eventually for pairs whose target gap is separated from $M(c)$ by a fixed positive margin. In the exponential case it is $\min\{1, e^{-d(C(\theta_0,\theta_1)-L\eta_d)}\}$, and in the categorical case $(\sqrt{xy}+\sqrt{(1-x)(1-y)})^d$. Put $T(e) = 0$ if $e\sqrt{m} \le d^{-1/8}$ and $T(e) = de$ otherwise. Choose representative one if the vote fraction $V > 1/2 + T$, representative zero if $V < 1/2 - T$, and the midpoint otherwise.

For every pair the normalized risk is at most $1/2 + o(1)$. If $T = 0$, the binomial at $1/2 + e$ is $o(1)$ from the fair one in total variation, by Pinsker and $me^2 \le d^{-1/4}$. If $T > 0$, Hoeffding bounds the wrong choice by $e^{-2m(d-1)^2e^2} \to 0$. For far pairs (gap above $M(c) + \kappa$), $e \le e^{-d(c+\gamma)}$, so $de = o(\delta)$. If $\tau < \infty$ then $T = 0$ eventually and the CLT gives $\Phi(-\sqrt{\tau})$. If $\tau = \infty$, the vote fraction on the correct side has mean margin at least $\delta/2 - e$, whereas $T \le de = o(\delta)$. The variance bound $1/(4m)$ gives probability $O((m\delta^2)^{-1})$ of a wrong decision or abstention on a far pair. Letting $\kappa \downarrow 0$ and combining with the known-level lower bound proves the claim. □

Proof of Theorem 6.2. Here $c = 0$ and $M(0) = 0$, so the known-level limits are $D^pB_p(\tau_j)$. On the swapped diameter submodel the data are a level- and endpoint-free randomization of labels with probabilities $(1 + (2b - 1)\delta_j)/2$. Scalarizing by $a = \min\{1, \|w - v_0\|/D\}$ only lowers both risks. Attaining the limit at level j forces the scalar action to approach the unique Bayes action, the logistic function of $Z_n = (2K_1 - n)/\sqrt{n}$ with slope $2\sqrt{\tau_j}/(p - 1)$, in probability. This uses strict convexity of the posterior loss and tightness of Z_n. The two label experiments are mutually contiguous, since the chi-square bound $(1 + C/n)^n$ holds in both directions. So both convergences hold under level one, but the two logistic functions differ off zero, which is a contradiction. □

For $D > 0$, $p > 1$, a nonempty compact $\mathcal{T} \subset (0, \infty)$, $\delta(\tau) = \sqrt{\tau/n}$ and $(\log n)/(2d) \to c$, let $H_p(\pi)$ be the Bayes risk for a fair label B and $\tau \sim \pi$ observed through $Z \sim N((2B-1)\sqrt{\tau}, 1)$. Put $a_c = (M(c)/(2D))^p$ and $b_c(\tau) = \max\{a_c, B_p(\tau)\}$.

Theorem 6.3. *For estimators that know $\mathcal{T}$,*

$$\lim_{n,d\to\infty}\left[\inf_{A}\sup_{\tau\in\mathcal{T}}\frac{R_{n,d}(A,\tau)-R^*_{n,d}(\tau)}{D^p}\right]=\max\Big\{0,\max_{\pi}\Big[H_p(\pi)-\int b_c\,d\pi\Big]\Big\}.$$

The regret is zero for a singleton $\mathcal{T}$. Otherwise, with $\tau_0=\min\mathcal{T}$, $\tau_+=\inf(\mathcal{T}\setminus\{\tau_0\})$, g_0 the Bayes rule at τ_0 and $L_\tau(g)$ the normalized endpoint risk of rule g at level τ, the regret is zero exactly when $a_c\ge L_{\tau_+}(g_0)$. In particular it is strictly positive at $c=0$ whenever $\mathcal{T}$ has two points.

Proof. The known-level convergence $R^*/D^p\to b_c$ is uniform on $\mathcal{T}$. Indeed, every sequence $\tau_{n,d}\in\mathcal{T}$ has a convergent subsequence, and $d^{-1}\log(1/\delta(\tau_{n,d}))=(\log n-\log\tau_{n,d})/(2d)\to c$. Apply the joint-limit theorem along that subsequence and use continuity of B_p.

For the lower bound use the swapped submodel with prior $B\times\pi$. Relative to fair labels the log likelihood equals $(2b-1)\sqrt{\tau}\,Z_n-\tau/2+O_{\mathcal{T}}((1+|Z_n|)/n)$, with $Z_n=(2K_1-n)/\sqrt{n}$. The expansion is uniform on bounded Z_n, and under the fair-label law the squared likelihood has expectation $(1+\tau/n)^n\le\exp(\max\mathcal{T})$. Integrating against π, the central limit theorem and uniform integrability imply convergence of both mixture likelihoods and of the Bayes risks. Here the minimized posterior loss, written against the fair-label law, is a continuous function of the two mixture likelihoods and is bounded above by their sum. The label Bayes risk therefore converges to $H_p(\pi)$.

For the upper bound write $W=\max_\pi[H_p(\pi)-\int b_c\,d\pi]$. A maximizer π_* exists by compactness and continuity. Its Bayes rule $g_{\pi_*}(z)=(1+(f_0^{\pi_*}(z)/f_1^{\pi_*}(z))^{1/(p-1)})^{-1}$ is increasing and reflection-symmetric. Uniqueness of this action, along with bounded loss, gives the directional derivative of the prior objective toward δ_τ as $L_\tau(g_{\pi_*})-b_c(\tau)-W$. This is nonpositive, so $L_\tau(g_{\pi_*})\le b_c(\tau)+W$.

Use the pilot of Theorem 6.1. Put $c_n=(\log n)/(2d)$ and fix $\kappa>0$. For a pair with target gap at most $M(c_n)+\kappa$, return its midpoint. On the event that a representative is near the truth, this has normalized risk at most $((M(c_n)+\kappa)/(2D))^p+o(1)$. For a far pair the Chernoff margin supplies a clean voting error $e=o(1/\sqrt{n})$ uniformly over $\mathcal{T}$. When $M(c)=D$, there are eventually no far pairs with this slack. Otherwise use the action $g_{\pi_*}((2K_1-m)/\sqrt{m})$ to interpolate between the representatives. The comparison binomial counts converge after standardization to the Gaussian laws uniformly in $\tau\in\mathcal{T}$ by Berry–Esseen, since the clean voting error is $o(1/\sqrt{n})$ and $m/n\to1$. Monotonicity gives the binomial risk comparison, and the bounded variation of the endpoint losses transfers uniform convergence of the count distribution functions to uniform convergence of the risks. The pilot error contributes $o(1)$ by bounded loss. The supremum excess risk is therefore at most $\max(0,W)$ after letting $\kappa\downarrow0$. In this argument one can use the prior maximizing the objective at the known finite index c_n: the Bayes terms $H_p(\pi)$ do not change, and the change in the benchmark b_c is uniform in τ and hence uniform over priors. For the threshold, $L_\tau(g_0)$ is continuous and nonincreasing in τ. If $a_c\ge L_{\tau_+}(g_0)$, then g_0 meets every benchmark. Otherwise $a_c<L_{\tau_+}(g_0)\le B_p(\tau_0)$, so $b_c(\tau_0)=B_p(\tau_0)$, and a sequence of rules meeting the benchmark at τ_0 asymptotically must converge to g_0 in probability on compact score sets, by the strict-convexity and tightness argument in the proof of Theorem 6.2; contiguity transfers this convergence to every level in $\mathcal{T}$. If $\tau_+>\tau_0$, then $\tau_+\in\mathcal{T}$ and $L_{\tau_+}(g_0)>\max\{a_c,B_p(\tau_+)\}=b_c(\tau_+)$, since g_0 differs from the unique Bayes rule at τ_+; the benchmark at τ_+ fails. If $\tau_+=\tau_0$, the threshold reads $a_c\ge B_p(\tau_0)$; when it fails, pick $\tau_j\in\mathcal{T}$, $\tau_j\downarrow\tau_0$, and use continuity of $\tau\mapsto L_\tau(g_0)$ to find τ_j with $L_{\tau_j}(g_0)>a_c$, while $L_{\tau_j}(g_0)>B_p(\tau_j)$ strictly for $\tau_j\ne\tau_0$; the benchmark at τ_j fails. At $c=0$, $a_c=0<L_{\tau_+}(g_0)$. □

7 Verification and literature

The constants of Sections 2–3 were checked numerically: the identity for J; the Gaussian overlap exponent and prefactor with $\gamma(x)$ to four digits at $d = 2 \times 10^5$; both Bernoulli parity prefactors and $\gamma_{\mathrm{even}}, \gamma_{\mathrm{odd}}$ at $d = 16000$; the formula $B_p(\tau) = \mathbb{E}g(Z_0)^p$; the $k = 1$ Poisson formula; the middle-threshold formula for $k \le 8$ by linear programming; the complete-loss threshold for $k \le 6$ and $p \in \{1.5, 2, 3\}$; the Bernoulli affinity inequality; and a positive adaptive regret at $c = 0$. Literature searches in November 2026 covered Chen–Gao–Ren [2], Peng–Neykov [3] (rates up to constants under a cap), Qiao–Valiant [4], Jain–Orlitsky [5], Chen–Li–Moitra [6], Jain–Orlitsky–Ravindrakumar [7] and Diakonikolas–Kane [8]. The comparisons with the cited work concern the specific joint limits studied here. Chen, Gao and Ren provide general modulus-based rate bounds, and Peng and Neykov identify the capped Gaussian mean rate up to constant factors. The untrusted-batch references supply rate bounds and algorithmic comparisons. The present results determine the joint limiting constants and the adaptation regret in the settings specified by the theorems.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Scott Schubr directed the research and is responsible for the content.

References

[1] S. Schubr. Sharp minimax rates for Gaussian location mixtures with unknown covariance, edition 3. Luminary Research Archive, November 2026. Section 7, *Robust isotropic noise calibration and the finite-sample majority limit*. https://doi.org/10.5281/zenodo.24447150.

[2] M. Chen, C. Gao and Z. Ren. A general decision theory for Huber's ϵ-contamination model. *Electronic Journal of Statistics* 10:3752–3774, 2016. https://doi.org/10.1214/16-EJS1216. Theorem numbers refer to arXiv:1511.04144v2, https://arxiv.org/abs/1511.04144v2.

[3] T. Peng and M. Neykov. Minimax rates for robust Gaussian mean estimation near the breakdown point under constraints. arXiv:2610.36283, 2026. https://arxiv.org/abs/2610.36283.

[4] M. Qiao and G. Valiant. Learning discrete distributions from untrusted batches. *ITCS*, LIPIcs 94:47:1–47:20, 2018. https://doi.org/10.4230/LIPIcs.ITCS.2018.47. arXiv version https://arxiv.org/abs/1711.08113.

[5] A. Jain and A. Orlitsky. Optimal robust learning of discrete distributions from batches. *ICML*, PMLR 119:4651–4660, 2020. https://proceedings.mlr.press/v119/jain20a.html.

[6] S. Chen, J. Li and A. Moitra. Efficiently learning structured distributions from untrusted batches. *STOC* 2020. https://arxiv.org/abs/1911.02035.

[7] A. Jain, A. Orlitsky and V. Ravindrakumar. Robust estimation algorithms don't need to know the corruption level. arXiv:2202.05453, 2022. https://arxiv.org/abs/2202.05453.

[8] I. Diakonikolas and D. M. Kane. *Algorithmic High-Dimensional Robust Statistics*. Cambridge University Press, 2023. https://doi.org/10.1017/9781108943161.

[9] H. Liu and C. Gao. Density estimation with contamination: minimax rates and theory of adaptation. *Electronic Journal of Statistics* 13:3613–3653, 2019. https://doi.org/10.1214/19-EJS1617.

[10] P. J. Huber. A robust version of the probability ratio test. *Annals of Mathematical Statistics* 36:1753–1758, 1965. https://doi.org/10.1214/aoms/1177699803.

[11] R. R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Annals of Mathematical Statistics* 31:1015–1027, 1960. https://doi.org/10.1214/aoms/1177705674.

[12] F. Nielsen. Revisiting Chernoff information with likelihood ratio exponential families. *Entropy* 24(10):1400, 2022. https://doi.org/10.3390/e24101400.

[13] I. Sason. Tight bounds for symmetric divergence measures and a refined bound for lossless source coding. *IEEE Transactions on Information Theory* 61:701–707, 2015. https://doi.org/10.1109/TIT.2014.2387065.

[14] Y. Jin and L. Lai. On the adversarial robustness of hypothesis testing. *IEEE Transactions on Signal Processing* 69:515–530, 2021. https://doi.org/10.1109/TSP.2020.3045206.

[15] B. Tondi, N. Merhav and M. Barni. Detection games under fully active adversaries. *Entropy* 21:23, 2019. https://doi.org/10.3390/e21010023.

[16] M. S. Pydi and V. Jog. Adversarial risk via optimal transport and optimal couplings. *ICML*, PMLR 119:7814–7823, 2020. https://proceedings.mlr.press/v119/pydi20a.html.