The logarithmic cost of componentwise constraints in finite de Finetti approximation

Abstract

We study relative-entropy approximation of k-site marginals of exchangeable n-site laws on a d-symbol alphabet by mixtures of iid laws. Each one-site factor must lie in a prescribed affine slice of the probability simplex. We assume that, under the exchangeable law, each site's conditional distribution given the other sites lies in this slice. For $k \le n/2$, conditioning gives feasible error $O(k^2 \log d/n)$, compared with the unrestricted guarantee $O(k^2/n)$. We prove that the logarithmic factor is necessary for a bound uniform over affine slices. For even n, we construct examples with $d = \Theta(n^8)$ and $k = \lfloor n^{1/4} \rfloor$ whose optimal feasible error is $\Theta(\log n/\sqrt{n})$. Thus the logarithmic separation persists while the error vanishes. The construction places perfect nonsignalling finite-field CHSH boxes on a random perfect matching. An internal matching edge forces an event that is much rarer under every feasible product mixture; its likelihood ratio supplies the logarithmic factor. For the affine slice specifying an input marginal of support size s, a response-table construction gives $O(sk^2/n)$ error for $k \le \frac{1}{2}\lfloor n/s \rfloor$, independently of output-alphabet sizes.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction

Suppose a one-site law describes a channel with a prescribed input marginal: it has the form $q(x, a) = \pi(x)W(a \mid x)$, and only the channel W may vary. A de Finetti approximation is useful here only if each product component has the same input law π. The usual empirical-measure construction does not have this property. If π is uniform on $r > n$ inputs, an empirical law formed from n sites has at most n input values in its support, so every realised empirical law violates the constraint. Averaging over empirical laws does not repair the individual channels. This elementary support mismatch already prevents direct use of the standard finite de Finetti approximation.

More generally, we require the one-site factors of an iid mixture to lie in an affine slice F of the probability simplex. Such requirements arise when de Finetti theorems are used to round symmetric extensions to feasible points in constrained optimisation [7, 10]. The extension must satisfy more than a one-site constraint: we impose its lifted version, which says that a site's conditional law remains in F after observing the other sites. Conditioning then gives a natural way to construct feasible mixture components. For the marginal on k retained sites of an exchangeable n-site law, the conditional-information proof of Berta, Gavalakis, and Kontoyiannis [1] yields error at most $k(k-1)\log d/[2(n-k+1)]$ on a d-symbol alphabet. By contrast, without component constraints the empirical construction gives

$$\log \frac{n^k}{(n)_k} \le \frac{k(k-1)}{2(n-k+1)},$$

where $(n)_k = n(n-1)\cdots(n-k+1)$ is the falling factorial. This bound is independent of d [2, 3]. Is the entropy cost $\log d$ inherent in keeping the components feasible, or could another construction retain the unrestricted rate?

We show that the logarithmic factor is necessary. The obstruction already occurs for a prescribed uniform input marginal on $\mathbb{F}_r$, with one-site alphabet $\mathbb{F}_r \times \mathbb{F}_r$ and $d = r^2$. Theorem 3.1 gives an explicit lower bound of order $k(k-1)\log d/(n-k+1)$ for even n, $2 \le k \le \sqrt{n}$, and primes r growing sufficiently quickly with n. A concrete choice makes the separation particularly transparent: take $r = \Theta(n^4)$, with the constant specified by the theorem, and $k = \lfloor n^{1/4} \rfloor$. Then

$$d = \Theta(n^8), \qquad \text{unrestricted error} = O(n^{-1/2}), \qquad \text{optimal feasible error} = \Theta(\log n/\sqrt{n}).$$

Feasibility thus forces an error larger than the unrestricted guarantee by at least a logarithmic factor, even though the constrained optimum tends to zero. This gives a statistical obstruction to every uniform $o(\log d)$ replacement in the conditioning bound, without a computational hardness assumption.

The construction hides correlations on a uniformly random perfect matching of the n sites. At each edge we use a pair of uniform inputs $x, y \in \mathbb{F}_r$ and outputs satisfying $a + b = xy$. Each input remains uniform after the other endpoint is observed, so the matching law satisfies the lifted constraint. In a k-site marginal, an internal matching edge occurs with probability of order k^2/n and forces at least one pair to satisfy the game equation. Two feasible independent sites, however, satisfy it with probability only $O(r^{-1/2})$, by an elementary point-line incidence count. Hence the probability that any pair wins is at most $O(k^2/\sqrt{r})$, uniformly over all feasible iid mixtures. With $r = \Theta(n^4)$, this is smaller by a factor of order n than the probability forced by a matching edge. Binary relative-entropy data processing charges the event's probability times the logarithm of this ratio, producing $k^2 \log n/n$. The finite-field CHSH game and its incidence interpretation are due to earlier work [5]; the matching extension makes their separation compatible with exchangeability at an arbitrarily larger number of sites.

There is a useful converse lesson for the channel example. If the input law has support size s, one can use s sites to record one full table of possible outputs, one for each input. An n-site symmetric extension provides $M = \lfloor n/s \rfloor$ such rows. Sampling the rows with replacement gives feasible iid components, while sampling without replacement recovers the required marginal. Adapting the construction of Christandl and Toner [4] to relative entropy gives error at most $\log(M^k/(M)_k)$, simultaneously for all $k \le M$. For $k \le M/2$ this is $O(sk^2/n)$, with no output-alphabet cost. The lower-bound examples have $s = r > n$; their growing input support prevents even one complete table from fitting inside the extension.

The conditioning method is closely related to the quantum and nonsignalling de Finetti arguments of Brandão and Harrow [6] and the constrained quantum results of Berta et al. [7]. Other constrained reductions provide domination inequalities [8, 9], and fixed-point constraints lead to rates governed by the structure of channel algebras [11]. Our lower bound concerns relative entropy to componentwise-feasible mixtures. Its witness uses a rare event with a large likelihood ratio, so conclusions for total variation or optimisation complexity require additional arguments.

2 Feasible mixtures and the conditioning bound

Let $\mathcal{Z}$ be a finite alphabet, with $|\mathcal{Z}| = d$, and write $\Delta(\mathcal{Z})$ for its probability simplex. Given a linear map $\Gamma : \mathbb{R}^{\mathcal{Z}} \to \mathbb{R}^t$ and a vector $\eta \in \mathbb{R}^t$, consider the nonempty affine slice

$$F = F_{\Gamma,\eta} := \{q \in \Delta(\mathcal{Z}) : \Gamma(q) = \eta\}.$$

For an exchangeable law P_n on $\mathcal{Z}^n$, $n \ge 1$, let P_j denote its j-site marginal, with $P_0 = 1$. We impose the *lifted affine constraint*

$$(\mathrm{id}^{\otimes(n-1)} \otimes \Gamma)(P_n) = P_{n-1} \otimes \eta. \tag{CP}$$

Exchangeability gives the corresponding identity at every coordinate. This is stronger than the one-site condition $\Gamma(P_1) = \eta$: it requires feasibility of the conditional law of a site after all other sites have been observed. We call (Γ, η, P_n) *admissible* when these conditions hold.

For a Borel probability measure μ on F, define

$$M_{k,\mu} := \int_F q^{\otimes k} \, \mathrm{d}\mu(q), \qquad \varepsilon_{\mathrm{con}}(P_n, F; k) := \inf_{\mu \in \mathcal{P}(F)} D(P_k \| M_{k,\mu}), \quad 1 \le k \le n,$$

where $\mathcal{P}(F)$ denotes the set of such measures. All logarithms are natural, and

$$D(P\|Q) := \sum_{z : P(z) > 0} P(z) \log \frac{P(z)}{Q(z)},$$

with value $+\infty$ if P is not absolutely continuous with respect to Q. The worst-case constrained error is

$$B_{\mathrm{con}}(d, n, k) := \sup_{(\Gamma, \eta, P_n) \text{ admissible}} \varepsilon_{\mathrm{con}}(P_n, F_{\Gamma,\eta}; k),$$

where the supremum includes all nonempty affine slices of a d-symbol simplex. The restriction $\mu \in \mathcal{P}(F)$ applies to every factor q in the mixture, rather than only to the averaged one-site law $\int q \, \mathrm{d}\mu(q)$.

Without the support restriction on μ, there is an alphabet-independent bound

$$\inf_{\mu \in \mathcal{P}(\Delta(\mathcal{Z}))} D(P_k \| M_{k,\mu}) \le L(n,k) := \log \frac{n^k}{(n)_k}, \qquad (n)_k := n(n-1)\cdots(n-k+1). \tag{1}$$

Song, Attiah, and Yu [2] proved this bound and its worst-case sharpness; Gavalakis, Johnson, and Kontoyiannis [3] placed it in a general sampling-without-replacement framework. In particular, $L(n,k) \le k(k-1)/[2(n-k+1)]$. The empirical-measure mixture underlying this bound need not be supported on F.

A different construction conditions on a suitably chosen block of coordinates and mixes the resulting one-site posterior laws. The entropy estimate of Berta, Gavalakis, and Kontoyiannis [1], applied to posterior laws that remain in F under (CP), gives

$$B_{\mathrm{con}}(d, n, k) \le \frac{k(k-1)}{2(n-k+1)} \log d, \qquad 1 \le k < n. \tag{2}$$

Compatibility with linear constraints is discussed in [1]. The following proof tracks the posterior laws explicitly. Its entropy average is also where the factor $\log d$ enters; the lower bound will show that a uniform argument cannot remove this factor.

Lemma 2.1 (Posterior preservation)**.** *Suppose P_n is exchangeable and satisfies* (CP)*. For every $i \in [n] := \{1, \dots, n\}$ and every $I \subseteq [n] \setminus \{i\}$,*

$$\Gamma(P_{Z_i | Z_I}) = \eta \qquad \textit{almost surely.}$$

In particular, all these posterior laws belong to F.

Proof. Move the constrained coordinate in (CP) to position i by exchangeability, and sum over all coordinates outside $I \cup \{i\}$. For every value z_I, the resulting identity is

$$\Gamma(P(Z_I = z_I, Z_i = \cdot)) = P(Z_I = z_I)\eta.$$

Whenever $P(Z_I = z_I) > 0$, division by this probability proves the claim. □

Proposition 2.2 (Conditioning upper bound)**.** *For every admissible* (Γ, η, P_n) *and every* $1 \le k < n$*, there is* $\mu \in \mathcal{P}(F)$ *such that*

$$D(P_k \| M_{k,\mu}) \le \frac{k(k-1)}{2(n-k+1)} \log d. \tag{3}$$

Proof. We give the conditioning argument of [1], keeping track of its support. Write $Z_a^b = (Z_a, \dots, Z_b)$, with an empty tuple when $a > b$. Let H denote Shannon entropy and let $I(U;V \mid W) = H(U \mid W) - H(U \mid V,W)$ denote conditional mutual information. For $k \le \ell \le n$ and $1 \le i \le k$, exchangeability gives

$$I(Z_1^{i-1}; Z_i \mid Z_{k+1}^{\ell}) = I(Z_1^{i-1}; Z_\ell \mid Z_k^{\ell-1}).$$

Indeed, both expressions involve the same first $i-1$ coordinates, one distinguished additional coordinate, and $\ell - k$ conditioning coordinates, all distinct. Summing over ℓ and using the chain rule,

$$\begin{aligned} \sum_{\ell=k}^{n} \sum_{i=1}^{k} I(Z_1^{i-1}; Z_i \mid Z_{k+1}^{\ell}) &= \sum_{i=1}^{k} I(Z_1^{i-1}; Z_k^n) \\ &\le \sum_{i=1}^{k} H(Z_1^{i-1}) \le \frac{k(k-1)}{2} \log d. \end{aligned} \tag{4}$$

Choose $\ell_* \in \{k, \dots, n\}$ for which the inner sum is at most its average, and set $Y = Z_{k+1}^{\ell_*}$. Conditional on Y, the first k coordinates have a common one-site law $q_Y = P_{Z_1|Y}$. Let μ be the law of q_Y. The chain rule for relative entropy and joint convexity give

$$\begin{aligned} D(P_k \| M_{k,\mu}) &\le \mathbb{E}_Y D(P_{Z_1^k|Y} \| q_Y^{\otimes k}) \\ &= \sum_{i=1}^{k} I(Z_1^{i-1}; Z_i \mid Y) \\ &\le \frac{k(k-1)}{2(n-k+1)} \log d. \end{aligned}$$

Finally, Lemma 2.1 gives $q_Y \in F$ almost surely, so this mixing measure is feasible. □

A direct repair of each empirical law can lose information that the mixture never needed to preserve. For example, let $d \ge n$, $F = \{u_d\}$, and $P_n = u_d^{\otimes n}$. The optimal feasible approximation is exact. On a sample with n distinct symbols, however, its empirical law $\widehat{q}$ satisfies

$$D(\widehat{q} \| u_d) = \log(d/n).$$

Thus a proof that charges this repair cost for individual components can pay a positive error on an iid input. Posterior conditioning avoids that cost in this example: every posterior is already u_d. Establishing a lower bound therefore requires a law that is separated from *all* feasible mixtures, not merely from repaired empirical components.

3 The random-matching lower bound

The lower bound uses the slice of laws with a fixed uniform input marginal. Let $\mathbb{F}_r$ be the field with r elements, let u_r be its uniform law, and write X for the first coordinate of $\mathcal{Z} = \mathbb{F}_r \times \mathbb{F}_r$. The parameter θ measures how much of the game-value separation remains after accounting for the probability that two observed sites are matched together.

Theorem 3.1 (Logarithmic alphabet-size lower bound)**.** *Fix $0 < \theta < 1/2$. Let $n \geq 6$ be even, and let r be a prime such that*

$$r^{1/2-\theta} \geq 2e(n-1).$$

On the alphabet $\mathcal{Z} = \mathbb{F}_r \times \mathbb{F}_r$, of size $d = r^2$, set

$$F := \{q \in \Delta(\mathcal{Z}) : q_X = u_r\}.$$

There is an exchangeable law P_n satisfying (CP) *for this fixed-input-marginal constraint such that, for every integer $2 \leq k \leq \sqrt{n}$ and every $\mu \in \mathcal{P}(F)$,*

$$D(P_k \| M_{k,\mu}) \geq \frac{3\theta\, k(k-1)\log d}{16(n-1)} \geq \frac{3\theta}{32}\,\frac{k(k-1)\log d}{n-k+1}. \tag{5}$$

For $\theta = 1/4$, the second lower bound has constant $3/128$ and requires $r \geq [2e(n-1)]^4$. Comparison with (2) already gives matching rates in this range. To see how these rates interact when the error is small, take a prime r within a constant factor of this threshold and $k = \lfloor n^{1/4} \rfloor$: the upper and lower bounds are both of order $\log n/\sqrt{n}$. We give the general polynomial-alphabet sequences after the proof and then improve their exponent using the stronger prime-field game bound of [5].

3.1 A matching extension that preserves the constraint

The construction needs an edge law with two properties: a pair drawn from it must always win the game, and revealing one endpoint must leave the other input uniform. The first creates the separating event; the second allows the edge to be placed inside an admissible extension. Write $z = (x, a) \in \mathcal{Z} = \mathbb{F}_r \times \mathbb{F}_r$. The linear map and target vector defining the slice in Theorem 3.1 are

$$(\Gamma q)(x) = \sum_{a \in \mathbb{F}_r} q(x, a), \qquad \eta = u_r.$$

Define the symmetric two-site law

$$R((x, a), (x', b)) := r^{-3}\mathbf{1}\{a + b = xx'\}. \tag{6}$$

There are r^3 satisfying quadruples, so R is normalized, and both one-site marginals are uniform on $\mathcal{Z}$. For every (x, a) and x', there is exactly one admissible b, whence

$$\sum_b R((x, a), (x', b)) = r^{-3} = R_1(x, a)u_r(x'). \tag{7}$$

Thus the second input is uniform and independent of the entire first site; the symmetric identity holds at the other endpoint. Equivalently, R is the joint law of independent uniform inputs and a perfect nonsignalling CHSH_r box. No quantum realization is required.

Let M be a uniformly random perfect matching of $[n]$. Conditional on M, place an independent copy of R on each edge, and define

$$P_n := \mathbb{E}_{\mathsf{M}} \bigotimes_{\{i,j\} \in \mathsf{M}} R_{ij}. \tag{8}$$

A matching makes each site interact with only one partner. Randomising the matching hides that partner and restores permutation symmetry, while the linear conditional-input identity survives the average. In particular, we choose the extension once, before specifying k.

Proposition 3.2 (Admissibility of the matching law)**.** *The law P_n in* (8) *is exchangeable and satisfies* (CP) *for $F = \{q : q_X = u_r\}$.*

Proof. The uniform law on perfect matchings is permutation invariant, and R is symmetric under exchange of its endpoints. Hence P_n is exchangeable.

Fix a matching M, let j be the partner of coordinate n, and denote the associated product law by P_n^{M}. Applying Γ at coordinate n, identity (7) replaces $R_{j,n}$ by its marginal at j, tensored with u_r. All other edge factors are unchanged. Therefore

$$(\mathrm{id}^{\otimes(n-1)} \otimes \Gamma)(P_n^{\mathsf{M}}) = (P_n^{\mathsf{M}})_{[n-1]} \otimes u_r.$$

Averaging over M proves (CP). □

In probabilistic notation, this identity says that

$$P_{Z_{-i},X_i} = P_{Z_{-i}} \otimes u_r \qquad (i \in [n]), \tag{9}$$

where Z_{-i} is the tuple of all sites except i. The correlations are sparse: writing $u_{\mathcal{Z}}$ for the uniform law on $\mathcal{Z}$,

$$P_2 = \frac{1}{n-1} R + \frac{n-2}{n-1} u_{\mathcal{Z}}^{\otimes 2}. \tag{10}$$

The two observed sites are matched together with probability $1/(n-1)$. Otherwise they belong to different independent edge factors, and marginalizing their partners leaves two independent uniform variables.

3.2 An incidence bound for feasible product laws

For two distinct sites, let

$$W_{ij} := \{A_i + A_j = X_i X_j\}.$$

The classical value of the finite-field game can be written as

$$\omega_r := \frac{1}{r^2} \max_{f,g:\mathbb{F}_r \to \mathbb{F}_r} \big|\{(x,y) \in \mathbb{F}_r^2 : f(x) + g(y) = xy\}\big|. \tag{11}$$

A feasible product law cannot exploit the other site's input: each output is generated by a channel acting on its own uniformly distributed input. Randomisation in these channels is a mixture of deterministic functions. For a deterministic pair, winning pairs of inputs are incidences between a graph and lines with distinct slopes. The field property limits how often two graph points can lie on these lines, which gives the required bound. This is the same incidence interpretation used in [5].

Lemma 3.3 (Finite-field incidence bound)**.** *For every $q, q' \in F$,*

$$(q \otimes q')(W_{12}) \le \omega_r \le \beta_r := \frac{1 + \sqrt{4r-3}}{2r} \le \frac{3}{2\sqrt{r}}. \tag{12}$$

Proof. Write $q(x,a) = r^{-1} s(a \mid x)$ and $q'(y,b) = r^{-1} t(b \mid y)$, where s, t are stochastic channels. For functions $f, g : \mathbb{F}_r \to \mathbb{F}_r$, set

$$\lambda_s(f) := \prod_{x \in \mathbb{F}_r} s(f(x) \mid x), \qquad \lambda_t(g) := \prod_{y \in \mathbb{F}_r} t(g(y) \mid y).$$

These are probability distributions, and

$$s(a \mid x) = \sum_f \lambda_s(f)\mathbf{1}\{f(x) = a\}, \qquad t(b \mid y) = \sum_g \lambda_t(g)\mathbf{1}\{g(y) = b\}.$$

Consequently,

$$(q \otimes q')(W_{12}) = \sum_{f,g} \lambda_s(f)\lambda_t(g)\frac{I(f,g)}{r^2},$$

where $I(f,g)$ is the number of pairs satisfying $f(x) + g(y) = xy$. This proves the first inequality in (12).

To bound $I(f,g)$, consider the graph points $\mathcal{G}_g = \{(y, g(y)) : y \in \mathbb{F}_r\}$ and the lines

$$\ell_x := \{(y,z) : z = xy - f(x)\}, \qquad x \in \mathbb{F}_r.$$

Put $t_x = |\ell_x \cap \mathcal{G}_g|$, so that $I(f,g) = \sum_x t_x$. Two distinct graph points have different first coordinates, and any line through both has the uniquely determined slope

$$x = \frac{g(y) - g(y')}{y - y'}.$$

Hence a pair of distinct graph points belongs to at most one of the lines ℓ_x. Counting ordered pairs gives

$$\sum_x t_x(t_x - 1) \le r(r-1).$$

Cauchy–Schwarz now yields

$$I(f,g)^2 \le r\sum_x t_x^2 = rI(f,g) + r\sum_x t_x(t_x-1) \le rI(f,g) + r^2(r-1).$$

Solving this quadratic inequality and dividing by r^2 gives $\omega_r \le (1 + \sqrt{4r-3})/(2r)$. Finally, $\sqrt{4r-3} \le 2\sqrt{r}$ and $1/(2r) \le 1/(2\sqrt{r})$, proving the last inequality. □

3.3 A rare-event relative-entropy witness

A single observed edge would distinguish the box from every feasible product law, but its location is unknown. We therefore test whether *any* observed pair wins. This incurs a factor $\binom{k}{2}$ in the product-law upper bound and gains the same factor from the number of possible internal matching edges. Their ratio depends on $(n-1)\omega_r$, not on k. We keep the game value symbolic so that either incidence estimate can be inserted into the same proof.

Proposition 3.4 (Matching-event witness)**.** *Let $n \ge 6$ be even and $2 \le k \le \sqrt{n}$. For the law P_n in* (8)*, suppose $\omega_r \le \beta < 3/[4(n-1)]$. Then, for every $\mu \in \mathcal{P}(F)$,*

$$D(P_k \| M_{k,\mu}) \ge \frac{3k(k-1)}{8(n-1)} \log \frac{3}{4e(n-1)\beta}. \tag{13}$$

Proof. Set $m = \binom{k}{2}$ and $E = \bigcup_{1 \le i < j \le k} W_{ij}$. For each feasible q, the union bound and Lemma 3.3 give $(q^{\otimes k})(E) \le m\omega_r$. Integrating over μ,

$$M_{k,\mu}(E) \le m\beta. \tag{14}$$

This bound is uniform in the mixing measure; it uses neither independence between the events W_{ij} nor a bound on the number of mixture components.

Let J be the number of matching edges contained in $[k]$. An internal edge satisfies the game relation with probability one, so $P_k(E) \geq \Pr\{J \geq 1\}$. A specified edge belongs to a uniform perfect matching with probability $1/(n-1)$, and two specified disjoint edges both belong with probability $1/[(n-1)(n-3)]$. Overlapping edges cannot both belong. There are $3\binom{k}{4}$ unordered pairs of disjoint candidate edges. The first two inclusion-exclusion terms imply

$$\begin{aligned}\Pr\{J \geq 1\} &\geq \frac{m}{n-1} - \frac{3\binom{k}{4}}{(n-1)(n-3)} \\ &= \frac{m}{n-1}\left(1 - \frac{(k-2)(k-3)}{4(n-3)}\right) \geq \frac{3m}{4(n-1)}. \end{aligned} \tag{15}$$

For the last inequality, use $(k-2)(k-3) \leq k^2 - 3 \leq n-3$.

Write

$$p = P_k(E), \qquad q = M_{k,\mu}(E), \qquad p_0 = \frac{3m}{4(n-1)}, \qquad q_0 = m\beta.$$

Here $p_0/q_0 = 3/[4(n-1)\beta]$: the number of candidate pairs cancels from the likelihood ratio. The assumptions give $0 < q_0 < p_0 < 1$, and the event bounds give $p \geq p_0$ and $q \leq q_0$. If $q = 0$, the relative entropy is infinite. Otherwise data processing under $\mathbf{1}_E$ yields

$$D(P_k \| M_{k,\mu}) \geq d_{\mathrm{Ber}}(p \| q),$$

where d_{Ber} is the relative entropy between Bernoulli laws. Its derivative with respect to its second argument is $(q-p)/[q(1-q)]$, so it is decreasing for $0 < q \leq p$. Therefore

$$\begin{aligned} d_{\mathrm{Ber}}(p \| q) &\geq d_{\mathrm{Ber}}(p \| q_0) \\ &\geq p \log \frac{p}{q_0} - p = p \log \frac{p}{eq_0}. \end{aligned}$$

Here the second term in binary relative entropy is bounded by

$$(1-p) \log \frac{1-p}{1-q_0} \geq (1-p)\log(1-p) \geq -p,$$

with the usual limiting convention at $p = 1$. The function $x \mapsto x \log(x/(eq_0))$ has derivative $\log(x/q_0)$, which is nonnegative for $x \geq q_0$. It follows that

$$D(P_k \| M_{k,\mu}) \geq p_0 \log \frac{p_0}{eq_0} = \frac{3m}{4(n-1)} \log \frac{3}{4e(n-1)\beta},$$

which is (13). □

Proof of Theorem 3.1. Use the matching law (8), which is admissible by Proposition 3.2. By Lemma 3.3, we may take $\beta = 3/(2\sqrt{r})$ in Proposition 3.4. The parameter assumption implies

$$\frac{\sqrt{r}}{2e(n-1)} \geq r^{\theta} > 1,$$

so this choice satisfies the hypothesis of the witness bound, and

$$\log \frac{3}{4e(n-1)\beta} = \log \frac{\sqrt{r}}{2e(n-1)} \geq \theta \log r.$$

Since $\log d = 2\log r$, (13) gives

$$D(P_k \| M_{k,\mu}) \geq \frac{3\theta k(k-1)\log d}{16(n-1)}.$$

Finally, $k \leq \sqrt{n}$ implies $2(n-k+1) \geq n-1$, yielding the second inequality in (5). □

Two features clarify the scope of the witness. First, identical one-site factors are unnecessary: the same lower bound holds for any mixture of laws $q_1 \otimes \cdots \otimes q_k$ with each $q_i \in F$, because Lemma 3.3 applies to every pair q_i, q_j. Second, every marginal in Theorem 3.1 has full support. Indeed, $k \leq n/2$, so a perfect matching with no internal edge in $[k]$ exists and has positive probability. Conditional on any such matching, the marginal is $u_{\mathcal{Z}}^{\otimes k}$. This also gives a full-support feasible product law, so the lower bound is not forced by an unavoidable support mismatch.

The witness combines a probability of order k^2/n with a logarithm of $\sqrt{r}/n$. A polynomial alphabet is enough to make this logarithm a fixed fraction of $\log d$. Taking $k = o(\sqrt{n/\log d})$ then makes the matching upper bound vanish. Fix $0 < \theta < 1/2$. For sufficiently large even n, let r_n be the least prime at least

$$R_n := \left\lceil [2e(n-1)]^{1/(1/2-\theta)} \right\rceil.$$

Bertrand's postulate gives $R_n \leq r_n \leq 2R_n$. Thus

$$r_n = \Theta\Big(n^{2/(1-2\theta)}\Big), \qquad d_n = r_n^2 = \Theta\Big(n^{4/(1-2\theta)}\Big). \tag{16}$$

Take $k_n = \lfloor n^{1/4} \rfloor$, and denote the corresponding matching law and affine slice by P_n and F_n. Define

$$t_n := \frac{k_n(k_n-1)\log d_n}{n-k_n+1} = \Theta\Big(\frac{\log n}{\sqrt{n}}\Big) \longrightarrow 0. \tag{17}$$

The constants in these asymptotic statements may depend on the fixed parameter θ. Combining Proposition 2.2 and Theorem 3.1 gives

$$\frac{3\theta}{32} t_n \leq \varepsilon_{\mathrm{con}}(P_n, F_n; k_n) \leq B_{\mathrm{con}}(d_n, n, k_n) \leq \frac{1}{2} t_n. \tag{18}$$

In particular, the optimal error of the constructed law, not merely the lower bound, tends to zero.

For any fixed $\delta > 0$, choosing $\theta = \delta/[2(4+\delta)]$ gives

$$d_n = \Theta(n^{4+\delta}), \qquad \varepsilon_{\mathrm{con}}(P_n, F_n; k_n) \geq \frac{3\delta}{64(4+\delta)} t_n. \tag{19}$$

This explicit family already rules out every sublogarithmic replacement in the uniform bound.

Corollary 3.5 (No uniform sublogarithmic replacement)**.** *There do not exist a constant $C \in (0,\infty)$ and a function $h : \mathbb{N} \to [0,\infty)$ with $h(d) = o(\log d)$ such that*

$$\varepsilon_{\mathrm{con}}(P_n, F; k) \leq C\, \frac{k(k-1)h(d)}{n-k+1}$$

for every admissible (Γ, η, P_n) on every finite alphabet and every $1 < k < n$.

Proof. Apply the proposed bound to (18) for any fixed $\theta \in (0, 1/2)$. Cancelling $k_n(k_n-1)/(n-k_n+1) > 0$ gives

$$\frac{3\theta}{32} \leq C \frac{h(d_n)}{\log d_n} \longrightarrow 0,$$

a contradiction. □

The incidence estimate determines how many input symbols are needed to make $n\omega_r$ polynomially small. Improving that estimate reduces the alphabet size without changing the witness. Bavarian and Shor [5, Theorem 1.3] prove that there are universal constants $C_0 \geq 1$ and $\epsilon > 0$ such that

$$\omega_r \leq C_0 r^{-1/2-\epsilon} \qquad \text{for every prime } r. \tag{20}$$

Decreasing ϵ if necessary, we may assume $0 < \epsilon < 1/2$. Inserting this estimate into the matching-event bound gives a polynomial alphabet exponent strictly below four.

Corollary 3.6 (A polynomial alphabet exponent below four)**.** *There exist universal constants $\gamma \in (2,4)$ and $c > 0$, and admissible matching laws with $d_n = \Theta(n^\gamma)$ and $k_n = \lfloor n^{1/4} \rfloor$, such that, for all sufficiently large even n,*

$$c\, t_n \leq \varepsilon_{\mathrm{con}}(P_n, F_n; k_n) \leq B_{\mathrm{con}}(d_n, n, k_n) \leq \frac{1}{2} t_n, \qquad t_n = \frac{k_n(k_n-1)\log d_n}{n - k_n + 1} = \Theta\Big(\frac{\log n}{\sqrt{n}}\Big).$$

Proof. Set $\alpha = 1/2 + \epsilon/2$ and $\gamma = 2/\alpha = 4/(1+\epsilon)$. Choose r_n to be the least prime at least $\lceil n^{1/\alpha} \rceil$. Bertrand's postulate gives $r_n = \Theta(n^{1/\alpha})$, hence $d_n = r_n^2 = \Theta(n^\gamma)$. Use the matching law over $\mathbb{F}_{r_n}$, and put $\beta = C_0 r_n^{-1/2-\epsilon}$. Since $n \leq r_n^\alpha$,

$$(n-1)\beta \leq C_0 r_n^{-\epsilon/2} \longrightarrow 0,$$

so Proposition 3.4 applies for all sufficiently large n. Moreover,

$$\begin{aligned} \log \frac{3}{4e(n-1)\beta} &\geq \frac{\epsilon}{2} \log r_n + \log \frac{3}{4eC_0} \\ &\geq \frac{\epsilon}{4} \log r_n \end{aligned}$$

for all sufficiently large n. Consequently,

$$D(P_{k_n} \| M_{k_n,\mu}) \geq \frac{3\epsilon\, k_n(k_n-1)\log d_n}{64(n-1)} \geq \frac{3\epsilon}{128} t_n.$$

Take $c = 3\epsilon/128$, and apply Proposition 2.2 for the upper bound. The claimed asymptotic size of t_n follows from $d_n = \Theta(n^\gamma)$ and $k_n = \lfloor n^{1/4} \rfloor$. □

4 Fixed input support and feasible response tables

The lower bound fixes the input marginal but lets its support grow with n. We now exploit the additional structure available when that support is small. The right empirical objects are complete response tables, each of which prescribes an output for every possible input. Every table defines a feasible deterministic channel, so their empirical mixtures retain the prescribed input law exactly. Let

$$\mathcal{Z} = \bigsqcup_{x \in \mathcal{X}} \{x\} \times \mathcal{A}_x, \qquad F_\pi := \{q \in \Delta(\mathcal{Z}) : q_X = \pi\},$$

where $\mathcal{X}$ is finite, $\pi \in \Delta(\mathcal{X})$, and each $\mathcal{A}_x$ is a nonempty finite output alphabet. Write $s = |\operatorname{supp} \pi|$. A table requires s physical sites, one per input; thus there are $M = \lfloor n/s \rfloor$ complete rows. The construction of Christandl and Toner [4], combined with relative-entropy data processing, turns the unrestricted sampling bound $L(n,k)$ into $L(M,k)$.

Proposition 4.1 (Fixed input support)**.** *Suppose P_n is exchangeable and satisfies* (CP) *for F_π, and let $n \geq s$. Set $M = \lfloor n/s \rfloor$. There is a single probability measure $\mu \in \mathcal{P}(F_\pi)$ such that, simultaneously for every $1 \leq k \leq M$,*

$$D(P_k \| M_{k,\mu}) \leq L(M,k) = \log \frac{M^k}{(M)_k} \leq \frac{k(k-1)}{2(M-k+1)}. \tag{21}$$

Proof. We first identify the conditional structure forced by the lifted constraint. Discard zero-mass inputs, which occur with probability zero at every site, and assume $\pi(x) > 0$ for all $x \in \mathcal{X}$. Thus $|\mathcal{X}| = s$. The lifted constraint at coordinate i is

$$P_{Z_{-i},X_i} = P_{Z_{-i}} \otimes \pi. \tag{22}$$

After summing over all outputs, this shows that X_i has law π and is independent of X_{-i}. Applying the identity at $i = n$, then to the inherited identities on the first $n-1$ coordinates, proves inductively that $X^n \sim \pi^{\otimes n}$.

For $S \subseteq [n]$, define

$$T_S(a_S \mid x_S) := P(A_S = a_S \mid X_S = x_S).$$

Divide (22) by the positive input mass $\pi^{\otimes n}(x^n)$. This gives

$$\sum_{a_i} T_{[n]}(a^n \mid x^n) = T_{[n]\setminus\{i\}}(a_{-i} \mid x_{-i}),$$

independently of x_i. Repeated summation gives

$$\sum_{a_{S^c}} T_{[n]}(a^n \mid x^n) = T_S(a_S \mid x_S) \qquad (S \subseteq [n]). \tag{23}$$

This is the nonsignalling condition: unobserved inputs do not affect observed outputs. Exchangeability of P_n, together with the iid input law, also makes these conditional laws permutation covariant. Conversely, these nonsignalling identities with iid inputs imply (22). Thus the lifted fixed-input constraint has an exact interpretation in terms of symmetric nonsignalling conditional laws.

We can now construct a random table with M complete response rows. Choose an injection $\iota : [M] \times \mathcal{X} \to [n]$, and fix an input string $\bar{x} \in \mathcal{X}^n$ with $\bar{x}_{\iota(j,x)} = x$. Assign arbitrary inputs to unused sites. Draw $\bar{A} \sim T_{[n]}(\cdot \mid \bar{x})$ and form the table

$$\Lambda = (\lambda_1, \ldots, \lambda_M), \qquad \lambda_j(x) := \bar{A}_{\iota(j,x)}.$$

Each row specifies one output in $\mathcal{A}_x$ for each input x. By (23), the law ν of Λ is independent of the inputs at unused sites. For a realized table $\lambda = (\lambda_1, \ldots, \lambda_M)$, set

$$q_\lambda(x,a) := \pi(x) \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}\{\lambda_j(x) = a\}. \tag{24}$$

This is a probability law with input marginal π. Let μ be the law of q_Λ, so $\mu \in \mathcal{P}(F_\pi)$.

The two distributions to be compared differ only in how they sample table rows. Fix $k \leq M$. Let $U^{(k)}_{\mathrm{dist}}$ be uniform on the ordered k-tuples of distinct elements of $[M]$, and let $U^{(k)}_{\mathrm{iid}}$ be uniform on $[M]^k$. Define the map

$$\Phi_k(\lambda, i^k, x^k) := \big((x_\ell, \lambda_{i_\ell}(x_\ell))\big)_{\ell=1}^k.$$

Under $\nu \otimes U_{\mathrm{dist}}^{(k)} \otimes \pi^{\otimes k}$, the law of this map is P_k. To see this, fix i^k and x^k. Distinct rows select the distinct physical sites $\iota(i_\ell, x_\ell)$, $1 \le \ell \le k$, whose prescribed inputs are x^k. Marginalizing the unselected outputs by (23) and permuting these sites to the first k positions gives exactly $T_{[k]}(\,\cdot \mid x^k)$. Under $\nu \otimes U_{\mathrm{iid}}^{(k)} \otimes \pi^{\otimes k}$, by contrast, the displayed pairs are conditionally iid given the table, with law q_λ, so their joint law is $M_{k,\mu}$.

Data processing under Φ_k therefore gives

$$\begin{aligned} D(P_k \| M_{k,\mu}) &\le D\Big(\nu \otimes U_{\mathrm{dist}}^{(k)} \otimes \pi^{\otimes k} \,\Big\|\, \nu \otimes U_{\mathrm{iid}}^{(k)} \otimes \pi^{\otimes k}\Big) \\ &= D\Big(U_{\mathrm{dist}}^{(k)} \Big\| U_{\mathrm{iid}}^{(k)}\Big) = \log \frac{M^k}{(M)_k}. \end{aligned}$$

Indeed, the likelihood ratio on distinct tuples is the constant $M^k/(M)_k$. The construction of ν and μ did not depend on k, proving the simultaneous assertion. Finally,

$$\log \frac{M^k}{(M)_k} = \sum_{j=0}^{k-1} -\log\left(1 - \frac{j}{M}\right) \le \sum_{j=0}^{k-1} \frac{j}{M-j} \le \frac{k(k-1)}{2(M-k+1)}.$$

□

When $s = 1$, the rows are ordinary one-site observations and (21) becomes the unrestricted bound (1). For fixed s and $k \le M/2$, it is $O(sk^2/n)$, independently of all output-alphabet sizes and of the smallest positive input probability. The matching examples instead have $s = r_n > n$, so no complete response row fits inside their extension. This accounts for the different behaviour of the two constructions: conditioning pays for the entropy of a site, while the table method pays for the number of sites needed to specify one feasible channel. Determining which other affine slices admit comparably efficient representations of feasible components remains open.

A Finite fields versus residue rings

The construction and the elementary incidence proof extend without change from prime fields to arbitrary finite fields. Primes are sufficient for the explicit sequences and are essential to the application of (20) in the form used above. By contrast, replacing a field by a composite residue ring invalidates the pair-counting argument: $y - y'$ need not be invertible, so two graph points can have more than one admissible slope.

Two small examples verify this distinction directly. Over $\mathbb{Z}/4\mathbb{Z}$, with function values indexed by $0, 1, 2, 3$, set

$$f = (0, 0, 0, 2), \qquad g = (0, 1, 0, 3).$$

The sets of winning second inputs for first inputs $0, 1, 2, 3$ are, respectively,

$$\{0, 2\}, \quad \{0, 1, 3\}, \quad \{0, 2\}, \quad \{1, 2, 3\}.$$

Thus $I(f, g) = 2 + 3 + 2 + 3 = 10$, exceeding the putative field bound $\lfloor 2(1 + \sqrt{13})\rfloor = 9$.

Over $\mathbb{Z}/6\mathbb{Z}$, set

$$f = (0, 0, 0, 0, 0, 2), \qquad g = (0, 3, 2, 0, 0, 5).$$

Here the complete winning sets are

x	0	1	2	3	4	5
$\{y : f(x) + g(y) = xy\}$	$\{0, 3, 4\}$	$\{0, 2, 5\}$	$\{0, 3\}$	$\{0, 1, 4\}$	$\{0, 2, 3\}$	$\{1, 2, 4, 5\}$.

The total is $I(f, g) = 3 + 3 + 2 + 3 + 3 + 4 = 18$, whereas the corresponding field expression would give $\lfloor 3(1 + \sqrt{21})\rfloor = 16$. These explicit lists require no external computation or auxiliary data.

References

[1] M. Berta, L. Gavalakis, and I. Kontoyiannis, A third information-theoretic approach to finite de Finetti theorems, in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024. doi:10.1109/ISIT57864.2024.10619572; arXiv:2304.05360.

[2] R. Song, K. M. Attiah, and W. Yu, Coded downlink massive random access and a finite de Finetti theorem, *IEEE Transactions on Information Theory* **71**(9), 6932–6949, 2025. doi:10.1109/TIT.2025.3564114; arXiv:2405.08301.

[3] L. Gavalakis, O. Johnson, and I. Kontoyiannis, Finite de Finetti bounds in relative entropy, arXiv:2407.12921, 2024.

[4] M. Christandl and B. Toner, Finite de Finetti theorem for conditional probability distributions describing physical theories, *Journal of Mathematical Physics* **50**(4), 042104, 2009. doi:10.1063/1.3114986; arXiv:0712.0916.

[5] M. Bavarian and P. W. Shor, Information causality, Szemerédi–Trotter and algebraic variants of CHSH, in *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)*, 2015. doi:10.1145/2688073.2688112; arXiv:1311.5186.

[6] F. G. S. L. Brandão and A. W. Harrow, Quantum de Finetti theorems under local measurements with applications, *Communications in Mathematical Physics* **353**, 469–506, 2017. doi:10.1007/s00220-017-2880-3; arXiv:1210.6367.

[7] M. Berta, F. Borderi, O. Fawzi, and V. B. Scholz, Semidefinite programming hierarchies for constrained bilinear optimization, *Mathematical Programming* **194**, 781–829, 2022. doi:10.1007/s10107-021-01650-1; arXiv:1810.12197.

[8] R. Arnon-Friedman and R. Renner, de Finetti reductions for correlations, *Journal of Mathematical Physics* **56**(5), 052203, 2015. doi:10.1063/1.4921341; arXiv:1308.0312.

[9] C. Lancien and A. Winter, Flexible constrained de Finetti reductions and applications, *Journal of Mathematical Physics* **58**(9), 092203, 2017. doi:10.1063/1.5003633; arXiv:1605.09013.

[10] J. A. Zeiss, G. Koßmann, R. Schwonnek, and M. Plávala, Finite de Finetti for convex bodies and polynomial optimization, arXiv:2601.15184, 2026.

[11] G. Koßmann and J. A. Zeiss, Fixed points in de Finetti hierarchies, arXiv:2607.23689, 2026.