TOTAL VARIATION DISTANCE BETWEEN PRODUCT DISTRIBUTIONS: AN ANALYTIC PROXY

ARIEL AVITAL, ARYEH KONTOROVICH, ROMAN VERSHYNIN, AND GUANGYI ZOU

Abstract. We characterize, up to universal constants, the total variation distance between finite products of arbitrary probability measures by a simple formula. A theorem of Latała reduces this expression to one scalar equation. As an application, we characterize the sample complexity of equal-prior binary hypothesis testing uniformly in the weak-detection regime, where squared Hellinger distance alone does not determine the answer, and also characterize the TV distance between multinomial distributions.

1. Main results

Total variation measures the accuracy of distributional approximations [Le 60] and determines the optimal error in equal-prior binary hypothesis testing [LCY00]. It does not tensorize exactly [Kon25]; analytic bounds for Bernoulli products and sums appear in [Kon26b, Kon26a, AKS26].

Exact evaluation is #P-hard even for Bernoulli products [BGM+23], but randomized and deterministic polynomial-time relative-approximation schemes are available for finite-alphabet products [FGJW23, FLL24]. For fixed accuracy and confidence, the randomized algorithm of Anand, Benford, and Guo [ABG26] runs in time linear in the marginal-table size in the unit-cost arithmetic model.

Our goal is analytic: we give a dimension-free constant-factor formula for products of arbitrary probability measures. It reduces to one scalar equation.

Recall that, for probability measures P, Q on a measurable space $(\mathsf{X}, \mathcal{F})$, the total variation is defined as

$$\mathrm{TV}(P, Q) := \sup_{B \in \mathcal{F}} |P(B) - Q(B)|.$$

Fix $n \geq 1$. For probability measures P_i, Q_i on measurable spaces $(\mathsf{X}_i, \mathcal{F}_i)$, our goal is to compute the TV distance between the product measures

$$\mathbf{P} := \bigotimes_{i=1}^{n} P_i, \qquad \mathbf{Q} := \bigotimes_{i=1}^{n} Q_i.$$

To do this, consider the midpoint measure defined by

$$M_i := \frac{P_i + Q_i}{2}, \qquad \mathbf{M} := \bigotimes_{i=1}^{n} M_i.$$

Define the *midpoint score* in coordinate i as the Radon–Nikodym derivative[1]

$$u_i := \frac{\mathrm{d}(P_i - Q_i)}{\mathrm{d}(P_i + Q_i)}.$$

Let $X_1, \ldots, X_n$ be independent random variables with distribution $X_i \sim M_i$, and set

$$U_i = u_i(X_i). \tag{1.1}$$

A.K. was supported by the Israel Science Foundation (ISF 581/25) and the Binational Science Foundation (BSF 2024243). R.V. was supported by NSF Grant DMS 2451011 and U.S. Air Force Grant FA9550-25-1-0294.

[1] If P_i, Q_i have densities or probability mass functions p_i, q_i, then $u_i(x) = (p_i(x) - q_i(x))/(p_i(x) + q_i(x))$, with $0/0 := 0$. The Radon–Nikodym notation extends this definition to arbitrary probability measures.

Theorem 1.1. *We have*

$$\mathrm{TV}(\mathbf{P},\mathbf{Q}) \asymp \min\Big\{1, \mathbb{E}\Big(\sum_{i=1}^n U_i^2\Big)^{1/2}\Big\}, \tag{1.2}$$

where the equivalence is up to a universal constant factor.[2]

This result can be further simplified. The expectation in (1.2) can be characterized, up to universal constants, by a one-dimensional equation:

Theorem 1.2. *Assume that* $\mathbf{P} \neq \mathbf{Q}$. *Let* $T > 0$ *be the unique solution of*

$$G(T) = \frac{1}{2}, \qquad \text{where} \qquad G(t) := \sum_{i=1}^n \log \mathbb{E}\sqrt{1 + \frac{U_i^2}{t^2}}.$$

Otherwise, set $T = 0$. *Then*

$$\mathrm{TV}(\mathbf{P},\mathbf{Q}) \asymp \min\{1, T\}. \tag{1.3}$$

2. The proofs

Proof of Theorem 1.1. By definition, we have M_i-almost surely:

$$\frac{\mathrm{d}P_i}{\mathrm{d}M_i} = 1 + u_i, \qquad \frac{\mathrm{d}Q_i}{\mathrm{d}M_i} = 1 - u_i.$$

This allows us to express the TV distance between the marginals in terms of U_i:

$$\mathrm{TV}(P_i, Q_i) = \frac{1}{2}\int \Big|\frac{\mathrm{d}P_i}{\mathrm{d}M_i} - \frac{\mathrm{d}Q_i}{\mathrm{d}M_i}\Big| \mathrm{d}M_i = \int |u_i| \mathrm{d}M_i = \mathbb{E}|U_i|. \tag{2.1}$$

Similarly, we have

$$\mathbb{E}U_i = 0, \qquad |U_i| \le 1 \quad (M_i\text{-almost surely.}) \tag{2.2}$$

Next, let's express the TV distance between the product distributions in terms of U_i. For any $x = (x_1, \dots, x_n)$, we have, $\mathbf{M}$-almost surely:

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{M}}(x) = \prod_{i=1}^n \big(1 + u_i(x_i)\big), \qquad \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{M}}(x) = \prod_{i=1}^n \big(1 - u_i(x_i)\big). \tag{2.3}$$

Expanding the products, we see that the difference of these two product ratios is their odd part:

$$\mathrm{TV}(\mathbf{P},\mathbf{Q}) = \frac{1}{2}\int \Big|\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{M}} - \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{M}}\Big| \mathrm{d}\mathbf{M} = \mathbb{E}|\Phi(U)|, \tag{2.4}$$

where

$$\Phi(U) := \frac{1}{2}\Big\{\prod_i (1 + U_i) - \prod_i (1 - U_i)\Big\} = \sum_{\substack{I \subseteq [n] \\ |I| \text{ odd}}} \prod_{i \in I} U_i. \tag{2.5}$$

Let us linearize Φ by writing

$$\Phi(U) = S + R, \qquad S := \sum_{i=1}^n U_i.$$

Marcinkiewicz–Zygmund inequality with exponent 1 (see, e.g., [Gut13, Chapter 3, Section 8]) allows us to evaluate the main, linear term:

$$\mathbb{E}|S| \asymp A, \qquad \text{where } A := \mathbb{E}\Big(\sum_{i=1}^n U_i^2\Big)^{1/2}. \tag{2.6}$$

[2]We write $a \lesssim b$ if $a \le Cb$ for a universal constant C, write $a \gtrsim b$ if $b \lesssim a$, and write $a \asymp b$ if both bounds hold.

Next, we want to bound the remainder term R. It contains the odd terms of degrees at least three. Distinct square-free monomials are orthogonal, since $\mathbb{E}U_i = 0$ by (2.2). Hence

$$\mathbb{E}R^2 = \sum_{\substack{k\ge 3\\ k \text{ odd}}} \sum_{|I|=k} \prod_{i\in I} \mathbb{E}U_i^2 \le \sum_{\substack{k\ge 3\\ k \text{ odd}}} \frac{\rho^k}{k!}, \qquad \text{where } \rho := \sum_{i=1}^n \mathbb{E}U_i^2. \tag{2.7}$$

In particular, if $0 \le \rho \le 1$, we have $\mathbb{E}R^2 \lesssim \rho^3$, so we obtain a bound on the remainder:

$$\mathbb{E}|R| \lesssim \rho^{3/2}, \quad \text{if } 0 \le \rho \le 1. \tag{2.8}$$

We want to say this this is negligible compared to the main term, whose expected magnitude is about A as we saw in (2.6). So let's bound A below:

$$\begin{aligned}
A &\ge 1 - \mathbb{E}e^{-\sum_i U_i^2} && (\text{by } \sqrt{z} \ge 1 - e^{-z})\\
&= 1 - \prod_i \mathbb{E}e^{-U_i^2} && (\text{by independence})\\
&\ge 1 - \prod_i \mathbb{E}(1 - \frac{1}{2}U_i^2) && (\text{since } 0 \le U_i^2 \le 1 \text{ by (2.2)})\\
&= 1 - \prod_i (1 - \frac{1}{2}\mathbb{E}U_i^2) \ge 1 - \prod_i e^{-\frac{1}{2}\mathbb{E}U_i^2} && (\text{by } 1 - x \le e^{-x})\\
&= 1 - e^{-\rho/2} \gtrsim \min\{1, \rho\}. && (2.9)
\end{aligned}$$

Combining this with (2.8), we obtain

$$\mathbb{E}|R| \lesssim \sqrt{\rho}\,\mathbb{E}|S|, \qquad \text{if } 0 \le \rho \le 1.$$

Small total variance. Suppose that $\rho \le \rho_0$, where $\rho_0 \in (0,1)$ is a sufficiently small universal constant. Then $\mathbb{E}|R| \le \frac{1}{2}\mathbb{E}|S|$, and hence, by (2.4) and (2.6),

$$\mathrm{TV}(\mathbf{P}, \mathbf{Q}) = \mathbb{E}|\Phi(U)| = \mathbb{E}|S + R| \asymp \mathbb{E}|S| \asymp A.$$

Moreover, Jensen's inequality implies that $A \le \sqrt{\rho} < 1$. This proves (1.2) in this regime.

Large total variance. If $\rho > \rho_0$, we no longer linearize Φ. Using $\min\{a,b\} \le \sqrt{ab}$ and the product structure, we obtain

$$\begin{aligned}
1 - \mathrm{TV}(\mathbf{P}, \mathbf{Q}) &= \int \min\Big\{\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{M}}, \frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{M}}\Big\}\,\mathrm{d}\mathbf{M}\\
&\le \int \sqrt{\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{M}}\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{M}}}\,\mathrm{d}\mathbf{M} = \prod_i \mathbb{E}\sqrt{1 - U_i^2} && (\text{by (2.3) and definition of } U_i)\\
&\le \prod_i \Big(1 - \frac{1}{2}\mathbb{E}U_i^2\Big) \le \prod_i e^{-\frac{1}{2}\mathbb{E}U_i^2} = e^{-\rho/2}.
\end{aligned}$$

Thus

$$\mathrm{TV}(\mathbf{P}, \mathbf{Q}) \ge 1 - e^{-\rho/2}. \tag{2.10}$$

This bound holds for every $\rho \ge 0$. In the present regime, it gives $\mathrm{TV}(\mathbf{P}, \mathbf{Q}) \gtrsim 1$, while (2.9) gives $A \gtrsim 1$. Conversely, $\mathrm{TV}(\mathbf{P}, \mathbf{Q}) \le 1$. Therefore

$$\mathrm{TV}(\mathbf{P}, \mathbf{Q}) \asymp \min\{1, A\}.$$

This completes the proof of Theorem 1.1. □

Proof of Theorem 1.2. The proof is a straightforward application of Latała's moment theorem [Lat97].

G is continuous and strictly decreasing. Moreover, $G(t) \to 0$ as $t \to \infty$ and $G(t) \to \infty$ as $t \downarrow 0$. (To check the latter fact, bound each summand of G below by $\mathbb{E}|U_i|/t$.) Therefore, by the intermediate value theorem, T exists and is unique.

Applying Latała's theorem [Lat97, Theorem 1] to the independent nonnegative random variables U_i^2, with $p = 1/2$, gives

$$\mathbb{E}\Big(\sum_{i=1}^n U_i^2\Big)^{1/2} \asymp T.$$

Combined with Theorem 1.1, this gives (1.3). □

3. Examples and applications

3.1. **Continuous and discrete distributions.** Theorem 1.2 applies for completely general distributions P_i and Q_i. Let's now specialize this result for continuous and for discrete distributions on $\mathbb{R}^d$.

If P_i and Q_i have densities p_i and q_i, then, unwrapping the definition of U_i in (1.1), we get

$$G(t) = \sum_{i=1}^n \log\left[\frac{1}{2}\int \sqrt{(p_i(x)+q_i(x))^2 + \frac{(p_i(x)-q_i(x))^2}{t^2}}\,\mathrm{d}x\right].$$

Similarly, if P_i and Q_i have probability mass functions p_i and q_i, then the same formula holds with the integral replaced by the sum over all x in the joint support.

In either case, $\mathrm{TV}(\bigotimes_i P_i, \bigotimes_i Q_i) \asymp \min\{1, T\}$ where T is the unique solution of $G(T) = \frac{1}{2}$.

3.2. **Tensorization bounds.** Suppose we wish to express the TV distance between the product measures $\bigotimes_i P_i$ and $\bigotimes_i Q_i$ in terms of the TV distances between their marginals P_i and Q_i. In general, this is not possible: TV distance does not tensorise, even approximately. The best we can do is the following pair of bounds, due to A. Kontorovich [Kon25, Theorem 1.1], which we can now readily recover from Theorem 1.1.

Corollary 3.1 (Kontorovich)**.** *For any probability measures P_i, Q_i on the same measurable space,*

$$\min\Big\{1, \Big(\sum_{i=1}^n \mathrm{TV}(P_i,Q_i)^2\Big)^{1/2}\Big\} \lesssim \mathrm{TV}\Big(\bigotimes_{i=1}^n P_i, \bigotimes_{i=1}^n Q_i\Big) \lesssim \min\Big\{1, \sum_{i=1}^n \mathrm{TV}(P_i,Q_i)\Big\}. \tag{3.1}$$

Proof. We just apply Theorem 1.1 and then use two elementary bounds on $\mathbb{E}\big(\sum_{i=1}^n U_i^2\big)^{1/2}$. For the upper bound, since $\|x\|_2 \le \|x\|_1$ for any vector $x \in \mathbb{R}$, we have

$$\mathbb{E}\Big(\sum_{i=1}^n U_i^2\Big)^{1/2} \le \mathbb{E}\sum_{i=1}^n |U_i| = \sum_i \mathrm{TV}(P_i,Q_i), \tag{3.2}$$

where we used (2.1) in the last step. For the lower bound, use the convexity of ℓ^2-norm and Jensen's inequality:

$$\mathbb{E}\Big(\sum_{i=1}^n U_i^2\Big)^{1/2} \ge \Big(\sum_{i=1}^n (\mathbb{E}|U_i|)^2\Big)^{1/2} = \Big(\sum_i \mathrm{TV}(P_i,Q_i)^2\Big)^{1/2}. \quad \square$$

Remark 3.2 (Optimality). Both bounds in (3.1) can be attained up to universal constants even in the i.i.d. case; see the Bernoulli examples in [Kon25, Section 2.4].

3.3. **Identical distributions.** Theorem 1.2 becomes especially transparent if all marginal measures are the same:

$$P_i = P, \quad Q_i = Q \quad \text{for all } i = 1, \dots, n.$$

Then $U_i \overset{d}{=} U$, and we get

$$\mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \asymp \min\{1, T\}, \tag{3.3}$$

where $T > 0$ is the unique solution of

$$\mathbb{E}\sqrt{1 + \frac{U^2}{T^2}} = e^{1/(2n)}. \tag{3.4}$$

To simplify this equation, subtract 1 from both sides and use that $\sqrt{1+x^2} - 1 \asymp \min\{x^2, x\}$ for $x \geq 0$. This gives

$$\Psi(T) \asymp \frac{1}{n}, \qquad \text{where} \qquad \Psi(t) := \mathbb{E}\min\left\{\frac{U^2}{t^2}, \frac{|U|}{t}\right\}. \tag{3.5}$$

The elementary scaling bounds

$$a^{-2}\Psi(t) \leq \Psi(at) \leq a^{-1}\Psi(t), \qquad a \geq 1, \tag{3.6}$$

show that $\Psi(t) \asymp 1/n$ if and only if $t \asymp T$. Thus any approximate solution of (3.5) may replace T in (3.3).

3.4. **Hypothesis testing.** Under equal priors, the optimal probability of correctly identifying whether n independent samples come from P or Q is

$$\frac{1}{2}\Big(1 + \mathrm{TV}(P^{\otimes n}, Q^{\otimes n})\Big),$$

see [LCY00]. Thus, the sample complexity for success probability at least $(1+\delta)/2$ is

$$n_\delta(P, Q) := \min\big\{n \geq 1 : \mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \geq \delta\big\}.$$

For fixed $\delta \in (0, 1)$, the classical characterization (see [TS21] and Remark 3.4 below) is in terms of the Hellinger distance:

$$n_\delta(P, Q) \asymp_\delta H(P, Q)^{-2} \tag{3.7}$$

where

$$H(P, Q)^2 = 1 - \mathbb{E}\sqrt{1 - U^2} \asymp \mathbb{E}U^2.$$

Our refinement is uniform as $\delta \downarrow 0$: the *weak-detection regime*, in which Hellinger distance alone does not characterize the sample complexity [PJL25, Section 8].

Corollary 3.3 (Uniform weak detection)**.** *For $P \neq Q$ and $0 < \delta \leq 1/2$,*

$$n_\delta(P, Q) \asymp \max\left\{1, \frac{\delta^2}{\mathbb{E}\min\{U^2, \delta|U|\}}\right\}. \tag{3.8}$$

Proof. Let Ψ be as in (3.5), and define T_n by $\Psi(T_n) = 1/n$. By (3.3), (3.5), and (3.6),

$$c\min\{1, T_n\} \leq \mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \leq C\min\{1, T_n\}$$

for universal $0 < c \leq 1/2$ and $C \geq 1$. For $0 < \delta \leq c$, monotonicity of Ψ gives

$$\max\{1, \Psi(\delta/C)^{-1}\} \lesssim n_\delta(P, Q) \lesssim \max\{1, \Psi(\delta/c)^{-1}\}. \tag{3.9}$$

By (3.6), both endpoints are comparable to $\max\{1, \Psi(\delta)^{-1}\}$, proving (3.8).

The lower bound in (3.9) remains valid for $c < \delta \leq 1/2$. In this range, $\Psi(\delta) \asymp \mathbb{E}U^2$, and (2.10) gives

$$\mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \geq 1 - \exp(-n\mathbb{E}U^2/2).$$

Thus any $n \geq 2\log 2/\mathbb{E}U^2$ satisfies $\mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \geq 1/2 \geq \delta$. Taking n of this order gives

$$n_\delta(P, Q) \lesssim \frac{1}{\mathbb{E}U^2},$$

which implies the upper bound in (3.8). □

The denominator separates small scores, which contribute quadratically, from large scores, which contribute linearly:

$$\mathbb{E}\min\{U^2, \delta|U|\} = \mathbb{E}[U^2\mathbf{1}_{\{|U|\le\delta\}}] + \delta\,\mathbb{E}[|U|\mathbf{1}_{\{|U|>\delta\}}].$$

This distinction is lost when the denominator is replaced by $\mathbb{E}U^2 \asymp H^2(P,Q)$.

Remark 3.4 (Higher confidence). For completeness, the remaining range has a classical characterization. Put

$$a := -\log \mathbb{E}\sqrt{1-U^2} = -\log\big(1 - H^2(P,Q)\big).$$

Hellinger tensorization and Cauchy–Schwarz give

$$1 - e^{-na} \le \mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \le \sqrt{1-e^{-2na}},$$

and hence, uniformly for $1/2 < \delta < 1$,

$$n_\delta(P,Q) \asymp \max\left\{1, \frac{\log(1/(1-\delta))}{a}\right\}.$$

For mutually singular P, Q, take $a = \infty$; one sample suffices. In the other end of the spectrum, if P, Q are close (for example, if $H^2(P,Q) \le 1/2$), we have $a \asymp H^2(P,Q)$, so we recover (3.7).

Together with (3.8), this covers every success probability strictly between 1/2 (a random guess) and 1 (full confidence).

3.5. **An analytic solution.** To facilitate further applications, let us solve the approximate equation (3.5) analytically.

Lemma 3.5. *We have*

$$T \asymp \inf_{A\in\mathcal{F}}\left\{n\mathbb{E}[|U|\mathbf{1}_A] + \sqrt{n\mathbb{E}[U^2\mathbf{1}_{A^c}]}\right\}. \tag{3.10}$$

Proof. By (3.5), for every $A \in \mathcal{F}$, we have

$$\frac{1}{n} \asymp \mathbb{E}\min\left\{\frac{U^2}{T^2}, \frac{|U|}{T}\right\} \lesssim \frac{1}{T}\mathbb{E}[|U|\mathbf{1}_A] + \frac{1}{T^2}\mathbb{E}[U^2\mathbf{1}_{A^c}].$$

Thus

$$T \lesssim \max\left\{n\mathbb{E}[|U|\mathbf{1}_A], \sqrt{n\mathbb{E}[U^2\mathbf{1}_{A^c}]}\right\} \lesssim n\mathbb{E}[|U|\mathbf{1}_A] + \sqrt{n\mathbb{E}[U^2\mathbf{1}_{A^c}]}.$$

Meanwhile, taking $A = \{|U| > T\}$, by (3.5), we obtain

$$\frac{1}{n} \asymp \mathbb{E}\min\left\{\frac{U^2}{T^2}, \frac{|U|}{T}\right\} = \frac{1}{T}\mathbb{E}[|U|\mathbf{1}_A] + \frac{1}{T^2}\mathbb{E}[U^2\mathbf{1}_{A^c}].$$

Hence,

$$T \gtrsim \max\left\{n\mathbb{E}[|U|\mathbf{1}_A], \sqrt{n\mathbb{E}[U^2\mathbf{1}_{A^c}]}\right\} \gtrsim n\mathbb{E}[|U|\mathbf{1}_A] + \sqrt{n\mathbb{E}[U^2\mathbf{1}_{A^c}]}.$$

Thus we finish the proof. □

3.6. **Multinomial laws.** For a probability measure P on $[k] := \{1, \dots, k\}$, let $\mathrm{Mult}(n, P)$ denote the law of the histogram of n independent samples from P.

For two such measures P, Q, every word with histogram z has probability $\prod_j P(j)^{z_j}$ under $P^{\otimes n}$ and $\prod_j Q(j)^{z_j}$ under $Q^{\otimes n}$. Their difference has the same sign for all words with the same histogram. Thus taking histograms preserves TV, and Theorem 1.2 gives

$$\mathrm{TV}\big(\mathrm{Mult}(n, P), \mathrm{Mult}(n, Q)\big) = \mathrm{TV}(P^{\otimes n}, Q^{\otimes n}) \asymp \min\{1, T\}. \tag{3.11}$$

By (3.10), we have

$$T \asymp \min_{A \subseteq [k]} \left\{ n \sum_{j \in A} |P(j) - Q(j)| + \sqrt{n \sum_{j \in A^c} \frac{(P(j) - Q(j))^2}{P(j) + Q(j)}} \right\}. \tag{3.12}$$

For Bernoulli case, let $p \neq q$. In (3.12), both absolute differences equal $|p - q|$, and the two denominators are $p + q$ and $2 - p - q$. Together with (3.11), this yields

$$\mathrm{TV}\big(\mathrm{Ber}(p)^{\otimes n}, \mathrm{Ber}(q)^{\otimes n}\big) \asymp \min \left\{ 1,\ n|p - q|,\ \frac{\sqrt{n}\,|p - q|}{\sqrt{(p + q)(2 - p - q)}} \right\}. \tag{3.13}$$

The case $p = q$ is immediate. This recovers the estimate of Liu et al. [LSY+20, Theorem 5]; see [AJ06] for exact formulas.

3.7. **Heterogeneous categorical sums.** Let P_i, Q_i be probability measures on $[k]$, with masses $p_{ij} := P_i(\{j\})$ and $q_{ij} := Q_i(\{j\})$. For independent $X_i \sim P_i$ and $Y_i \sim Q_i$, set

$$N_P := \sum_{i=1}^{n} e_{X_i}, \qquad N_Q := \sum_{i=1}^{n} e_{Y_i},$$

where e_j is the jth standard basis vector of $\mathbb{R}^k$.

Common tilt. Suppose that, for some $h \in \mathbb{R}^k$,

$$p_{ij} = \frac{q_{ij} e^{h_j}}{Z_i}, \qquad Z_i := \sum_{\ell=1}^{k} q_{i\ell} e^{h_\ell}. \tag{3.14}$$

The two product laws have the same support. For a word x in this support with histogram z,

$$\prod_{i=1}^{n} \frac{p_{i,x_i}}{q_{i,x_i}} = \frac{\exp\{\sum_j h_j z_j\}}{\prod_i Z_i}.$$

Hence their probability differences have the same sign for all words with the same histogram. Summing within each histogram therefore preserves total variation. Theorem 1.2 gives

$$\mathrm{TV}(N_P, N_Q) = \mathrm{TV}\Big(\bigotimes_i P_i, \bigotimes_i Q_i\Big) \asymp \min\{1, T\}, \tag{3.15}$$

Remark 3.6. If all p_{ij}, q_{ij} are positive, then (3.14) holds if and only if the product probability ratio depends only on the histogram. To see the converse, swapping categories j, ℓ between distinct trials i, r gives

$$\frac{p_{ij}}{q_{ij}} \frac{p_{r\ell}}{q_{r\ell}} = \frac{p_{i\ell}}{q_{i\ell}} \frac{p_{rj}}{q_{rj}}.$$

Thus the ratio matrix has rank one: $p_{ij}/q_{ij} = a_i b_j$ with $a_i, b_j > 0$. Row normalization gives $a_i = (\sum_j q_{ij} b_j)^{-1}$, yielding (3.14) with $h_j = \log b_j$.

Use of generative AI. Generative-AI tools were used during exploratory development and editorial preparation of this draft. The authors take responsibility for all mathematical statements and references.

Overlap with an earlier preprint. Some results overlap with [Kon26a]. The latter preprint is not currently under submission and will not be submitted for publication at any future time.

References

[ABG26] Konrad Anand, Alistair Benford, and Heng Guo. Linear time approximation of the TV distance between product distributions. arXiv:2607.27088, 2026.

[AJ06] José A. Adell and P. Jodrá. Exact Kolmogorov and total variation distances between some familiar discrete distributions. *J. Inequal. Appl.*, pages Art. ID 64307, 8, 2006.

[AKS26] Ariel Avital, Aryeh Kontorovich, and George Salafatinos. TV over Bernoulli products: the small parameter regime. *arXiv preprint arXiv:2602.21828*, 2026.

[BGM+23] Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, A. Pavan, and N. V. Vinodchandran. On approximating total variation distance. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 3479–3487. International Joint Conferences on Artificial Intelligence Organization, August 2023. Main Track.

[FGJW23] Weiming Feng, Heng Guo, Mark Jerrum, and Jiaheng Wang. A simple polynomial-time approximation algorithm for the total variation distance between two product distributions. *TheoretiCS*, 2, 2023.

[FLL24] Weiming Feng, Liqiang Liu, and Tianren Liu. On deterministically approximating total variation distance. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1766–1791, 2024.

[Gut13] Allan Gut. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, second edition, 2013.

[Kon25] Aryeh Kontorovich. On the tensorization of the variational distance. *Electronic Communications in Probability*, 30:1–10, 2025. Paper No. 32.

[Kon26a] Aryeh Kontorovich. A homogenization principle for total variation. arXiv preprint arXiv:2604.03882, 2026.

[Kon26b] Aryeh Kontorovich. TV homogenization inequalities. arXiv preprint arXiv:2601.04079, 2026.

[Lat97] Rafał Latała. Estimation of moments of sums of independent real random variables. *Ann. Probab.*, 25(3):1502–1513, 1997.

[LCY00] Lucien Le Cam and Grace Lo Yang. *Asymptotics in statistics*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2000.

[Le 60] Lucien Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10(4):1181–1197, 1960.

[LSY+20] Yuhan Liu, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Michael Riley. Learning discrete distributions: user vs item-level privacy. In *Advances in Neural Information Processing Systems*, volume 33, pages 20965–20976, 2020.

[PJL25] Ankit Pensia, Varun Jog, and Po-Ling Loh. The sample complexity of simple binary hypothesis testing. arXiv:2403.16981v2, 2025.

[TS21] Ananda Theertha Suresh. Robust hypothesis testing and distribution estimation in Hellinger distance. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2962–2970, 2021.

A.A and A.K.: Ben-Gurion University of the Negev

R.V. and G.Z: University of California, Irvine