The Annals of Applied Probability
2020, Vol. 30, No. 2, 503–525
https://doi.org/10.1214/19-AAP1506

LOWER BOUNDS FOR TRACE RECONSTRUCTION

BY NINA HOLDEN[1] AND RUSSELL LYONS[2]

[1]*Department of Mathematics, Massachusetts Institute of Technology, ninahold@gmail.com*
[2]*Department of Mathematics, Indiana University, rdlyons@indiana.edu*

In the trace reconstruction problem, an unknown bit string $\mathbf{x} \in \{0,1\}^n$ is sent through a deletion channel where each bit is deleted independently with some probability $q \in (0,1)$, yielding a contracted string $\widetilde{\mathbf{x}}$. How many i.i.d. samples of $\widetilde{\mathbf{x}}$ are needed to reconstruct $\mathbf{x}$ with high probability? We prove that there exist $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ such that at least $cn^{5/4}/\sqrt{\log n}$ traces are required to distinguish between $\mathbf{x}$ and $\mathbf{y}$ for some absolute constant c, improving the previous lower bound of cn. Furthermore, our result improves the previously known lower bound for reconstruction of random strings from $c\log^2 n$ to $c\log^{9/4} n/\sqrt{\log\log n}$.

1. Introduction. In trace reconstruction, the goal is to reconstruct an unknown bit string $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{S}_n := \{0,1\}^n$ from noisy observations of $\mathbf{x}$. Here, we study the case where the noise is due to a deletion channel in which each bit is deleted independently with a fixed probability $q \in (0,1)$. More precisely, instead of observing $\mathbf{x}$, we observe many independent strings $\widetilde{\mathbf{x}}$ obtained by the following procedure for $k = 1, \dots, n$, starting from an empty string:

- (retention) With probability $p := 1 - q$, copy x_k to the end of $\widetilde{\mathbf{x}}$ and increase k by one.
- (deletion) With probability q, only increase k by one.

See Figure 1 for an illustration. We are *not* given the locations of the retained bits in the original string.

For $T \in \mathbb{N}$, we consider a collection $\mathfrak{X} = \{\widetilde{\mathbf{x}}^{(1)}, \dots, \widetilde{\mathbf{x}}^{(T)}\}$ of T independent outputs (called "traces") from the deletion channel. Our main question is the following: How many traces are needed to reconstruct $\mathbf{x}$ with high probability? A closely related question is, given strings $\mathbf{x}$ and $\mathbf{y}$, how many traces are needed to determine whether the input string was $\mathbf{x}$ or $\mathbf{y}$. See Section 1.2 for a more precise problem statement.

1.1. *History and results.* This problem was introduced by Batu, Kannan, Khanna and McGregor [1] as an abstraction and simplification of a fundamental problem in bioinformatics, where one desires to reconstruct a common ancestor of several organisms given genetic sequences from those organisms. Other kinds of changes can be present besides deletions, but deletions present a key difficulty. See [1] for more details.

De, O'Donnell and Servedio [5] and Nazarov and Peres [14] prove that any string $\mathbf{x} \in \{0,1\}^n$ can be reconstructed with $\exp(O(n^{1/3}))$ traces, using the single-bit statistics of the trace. This improves the earlier upper bound of $\exp(n^{1/2}\operatorname{polylog}(n))$ proved by Holenstein, Mitzenmacher, Panigrahy and Wieder [8] (see [13] for an alternative proof).

Previous to our paper, the best available lower bound for the number of traces needed for reconstruction was $\Omega(n)$. For example, the pair of strings $\mathbf{x}'_n = (0)^{n-1}1(0)^n \in \mathcal{S}_{2n}$ and $\mathbf{y}'_n = (0)^n 1(0)^{n-1} \in \mathcal{S}_{2n}$ (where $(b)^m$ means a string of m consecutive bs) requires $\Omega(n)$ traces to

Received August 2018; revised May 2019.
MSC2010 subject classifications. Primary 62C20, 68Q25, 68W32; secondary 68W40, 68Q87, 60K30.
Key words and phrases. Strings, deletion channel, sample complexity.

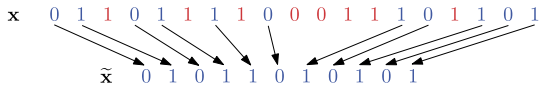

FIG. 1. *We obtain the trace* $\widetilde{\mathbf{x}}$ *by deleting* (*red*) *or copying* (*blue*) *each bit of* $\mathbf{x}$.

be distinguished ([1], Section 4.2, [13], Corollary 1). Our main result is an improvement of this lower bound. Define the strings $\mathbf{x}_n, \mathbf{y}_n \in \mathcal{S}_{4n}$ to be

$$\mathbf{x}_n := (01)^{n-1}10(01)^n, \qquad \mathbf{y}_n := (01)^n10(01)^{n-1}.$$

These strings are periodic with period 01 except for a single "defect" where the period is replaced by 10. They can be obtained from $\mathbf{x}'_n$ and $\mathbf{y}'_n$ by replacing each 0 with 01 and 1 with 10.

THEOREM 1.1. *For all* $q \in (0,1)$, *there is a constant* $c > 0$ *such that for all* $\varepsilon \in (0, 1/2)$ *and* $n \geq 2$, *at least* $T_n := \lceil c \log(1/\varepsilon) n^{5/4}/\sqrt{\log n}\,\rceil$ *traces are required to distinguish between* $\mathbf{x}_n$ *and* $\mathbf{y}_n$ *with probability at least* $1-\varepsilon$. *In particular,* T_n *traces are required to reconstruct all* n*-bit strings with probability at least* $1-\varepsilon$.

The following proposition is a partial converse to Theorem 1.1, and says that with $O(n^{3/2}\log n)$ traces we can distinguish between the strings $\mathbf{x}_n$ and $\mathbf{y}_n$.

PROPOSITION 1.2. *For all* $q \in (0,1)$, *there is a constant* $C > 0$ *such that for all* $\varepsilon \in (0, 1/2)$ *and* $n \geq 2$, $\lceil C \log(1/\varepsilon) n^{3/2} \log n \rceil$ *traces suffice to distinguish between* $\mathbf{x}_n$ *and* $\mathbf{y}_n$ *with probability at least* $1-\varepsilon$.

REMARK 1.3. Building on a prior version of this paper, Zachary Chase [3] recently strengthened our result by proving that Theorem 1.1 still holds with $T_n := \lceil cn^{3/2}/\log^{16} n \rceil$, where c depends on q and ε. This means that our upper bound in Proposition 1.2 is optimal up to a logarithmic factor. Using the stronger version of Theorem 1.1, Chase also improved the lower bound for random strings in Proposition 1.5 below to $\lceil c \log^{5/2} n/(\log\log n)^{16} \rceil$.

In general, the number of samples required to distinguish two probability measures is related to, but not determined by, the total variation distance between those measures; our Appendix reviews the precise relationships. Given a string $\mathbf{x}$ and a deletion probability $q \in (0,1)$, write $\Delta_{\mathbf{x}}$ for the law of the trace we obtain when applying the deletion channel with deletion probability q to $\mathbf{x}$. Note that the dependence on q is hidden in the notation $\Delta_{\mathbf{x}}$. The result of Proposition 1.2 follows from the lower bound on the total variation distance $d_{\mathrm{TV}}(\Delta_{\mathbf{x}_n}, \Delta_{\mathbf{y}_n})$ in the following proposition.

PROPOSITION 1.4. *For all* $q \in (0,1)$, *there is a constant* $C > 0$ *such that for all* $n \geq 1$, *the total variation distance* $d_{\mathrm{TV}}(\Delta_{\mathbf{x}_n}, \Delta_{\mathbf{y}_n})$ *between* $\Delta_{\mathbf{x}_n}$ *and* $\Delta_{\mathbf{y}_n}$ *satisfies*

$$C^{-1} n^{-3/4} (\log n)^{-1/2} \leq d_{\mathrm{TV}}(\Delta_{\mathbf{x}_n}, \Delta_{\mathbf{y}_n}) \leq C n^{-3/4}.$$

Above we considered reconstruction of arbitrary strings in $\mathcal{S}_n$. The number of traces required to reconstruct an arbitrary $\mathbf{x} \in \mathcal{S}_n$ is known as the *worst-case reconstruction problem*. We require that there exists a reconstruction algorithm such that no matter what the input string is, this string is found with high probability by the algorithm. One can also consider the *average-case reconstruction problem*. Letting μ_n denote the uniform probability measure on $\mathcal{S}_n$, we assume the input string $\mathbf{x}$ is sampled from μ_n. The question now is: What T ensures a large probability of reconstructing $\mathbf{x}$? We require that there exists a reconstruction

algorithm such that if $\mathbf{x} \sim \mu_n$, then the algorithm identifies $\mathbf{x}$ with high probability when we average over both the randomness of $\mathbf{x}$ and the randomness of the traces. In effect, this allows us to consider only $\mathbf{x} \in A_n$, where $A_n \subset \mathcal{S}_n$ is a set of large μ_n-measure and $\mathcal{S}_n \setminus A_n$ is a set of strings that are particularly difficult to reconstruct.

Using the lower bound of $\Omega(n)$ for worst-case strings, McGregor, Price and Vorotnikova [13] proved that $\Omega(\log^2 n)$ traces are needed to reconstruct random strings. Following [13], we state and prove a general result for transferring lower bounds for worst-case strings to lower bounds for random strings. We use this and Theorem 1.1 to prove Proposition 1.5, which improves the earlier lower bound for random strings.

PROPOSITION 1.5. *For all $q \in (0,1)$, there is a constant $c > 0$ such that for all large n, the probability of reconstructing random n-bit strings from $\lceil c \log^{9/4} n/\sqrt{\log\log n}\,\rceil$ traces is at most* $\exp(-n^{0.15})$.

Upper bounds for random strings are studied in [1, 7, 15]. In particular, it is proved in [7] that $e^{O(\log^{1/3} n)} = n^{o(1)}$ traces suffice for reconstruction of random strings with any deletion probability $q \in (0,1)$.

We use the following notation throughout the paper.

NOTATION 1.6. For two functions $f, g\colon \mathbb{N} \to [0,\infty)$, we write $f(n) = O(g(n))$ if there is a constant $C > 0$ such that for all sufficiently large n, $f(n) \le Cg(n)$; $f(n) = \Omega(g(n))$ if there is a constant $c > 0$ such that for all sufficiently large n, $f(n) \ge cg(n)$; $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$; and $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$. Unless otherwise specified, all constants $c, c_0, c_1, \ldots, C, C_0, C_1, \ldots$ and implicit constants in $\Omega(\cdot)$, $\Theta(\cdot)$, $O(\cdot)$ may depend on the deletion probability $q \in (0,1)$, but are independent of all other parameters.

For $\mathbf{x} \in \mathcal{S}_n$, let $\mathbf{P}_\mathbf{x}$ and $\mathbf{E}_\mathbf{x}$ denote probability and expectation, respectively, for the deletion channel with input string $\mathbf{x}$. The deletion probability is fixed and always denoted by q.

We remark that the trace reconstruction problem has a somewhat similar flavor to the problem of reconstructing a random scenery from the observations along a random walk path [2, 6, 10–12]. However, to our knowledge no nontrivial lower bounds have been proved for the scenery reconstruction problem.

In the remainder of the Introduction, we give a precise description of the trace reconstruction problem. We prove Theorem 1.1 and the upper bound of Proposition 1.4 in Section 2, Proposition 1.2 and the lower bound of Proposition 1.4 in Section 3 and Proposition 1.5 in Section 4. The Appendix contains some useful information about distances between probability measures and how they relate to the statistical problem of distinguishing two measures.

1.2. *The trace reconstruction problem.* Let $\mathcal{S} := \bigcup_{n\ge 0} \mathcal{S}_n$ denote the set of bit strings of finite length. Given $n \ge 0$ and $T \ge 0$, we say that (all) bit strings of length n can be *reconstructed* with probability at least $1-\varepsilon$ from T traces if there is a function[1] $G\colon \mathcal{S}^T \to \{0,1\}^n$ such that for all $\mathbf{x} \in \mathcal{S}_n$,

$$\mathbf{P}_\mathbf{x}\big[G(\mathfrak{X}) = \mathbf{x}\big] \ge 1-\varepsilon. \tag{1.1}$$

If (1.1) does not hold for any choice of G, then we say that more than T traces are required to reconstruct length-n bit strings with probability $1-\varepsilon$.

[1] Alternatively, we can replace $\mathcal{S}$ by $\bigcup_{k\le n} \mathcal{S}_k$ when specifying the domain of G.

Given $n \geq 0$, $T \geq 0$ and $\mathbf{x}, \mathbf{y} \in \mathcal{S}_n$, we say that we can *distinguish* between strings $\mathbf{x}$ and $\mathbf{y}$ with probability at least $1 - \varepsilon$ from T traces if there is a function $G\colon \mathcal{S}^T \to \{0,1\}^n$ such that

$$\mathbf{P}_{\mathbf{x}}\big[G(\mathfrak{X}) = \mathbf{x}\big] \geq 1 - \varepsilon \quad \text{and} \quad \mathbf{P}_{\mathbf{y}}\big[G(\mathfrak{X}) = \mathbf{y}\big] \geq 1 - \varepsilon. \tag{1.2}$$

If (1.2) does not hold for any choice of G, then we say that more than T traces are required to distinguish between $\mathbf{x}$ and $\mathbf{y}$ with probability $1 - \varepsilon$.

Recall that μ_n denotes the uniform probability measure on $\mathcal{S}_n$, that is, $\mu_n(\mathbf{x}) = 2^{-n}$ for all $\mathbf{x} \in \mathcal{S}_n$. We say that *random* bit strings of length n can be reconstructed with probability at least $1 - \varepsilon$ from T traces if there is a function $G\colon \mathcal{S}^T \to \{0,1\}^n$ such that

$$\sum_{\mathbf{x} \in \mathcal{S}_n} \mathbf{P}_{\mathbf{x}}\big[G(\mathfrak{X}) = \mathbf{x}\big] \cdot \mu_n(\mathbf{x}) \geq 1 - \varepsilon. \tag{1.3}$$

If (1.3) does not hold for any choice of G, then we say that more than T traces are required to reconstruct random length n bit strings with probability $1 - \varepsilon$.

Finally, we remark that one can also consider a variant of the problem where the function G may be randomized, as explained in Section 4, but this has no significant effect on our results.

2. Lower bound: Proof of Theorem 1.1. In this section, we will prove Theorem 1.1. We begin with a rough (and not entirely accurate) sketch of the proof. We will construct a coupling of the traces from $\mathbf{x}_n$ and $\mathbf{y}_n$ in two steps. The first step of the coupling is similar to what one does for $\mathbf{x}'_n$ and $\mathbf{y}'_n$, whose details can be found in [13], Corollary 1: Keep 01-blocks and 10-blocks intact, and for each 01-block decide only whether the block should be fully deleted (i.e., both bits are deleted) or not. Then the only thing we need to track is the numbers of blocks on either side of the defect that are not fully deleted. These are binomial random variables, and thus the total variation distance of the traces is at most that for binomial random variables, which is $\Theta(n^{-1/2})$. In fact, we will need to reserve some randomness, so in the first step, we delete each 01 independently with probability only $q^2/2$ (instead of q^2), which does not change the order of magnitude of the total variation distance.

We call the result of the first step a 2-partial trace. This is a string consisting of a sequence of 01-blocks, followed by a 10-block (i.e., the defect), followed by a sequence of 01-blocks. Consider the event that the first coupling step did not succeed in making the 2-partial traces the same for $\mathbf{x}_n$ and $\mathbf{y}_n$. On this event, in the second step of the coupling, we increase the success probability of coupling the final traces; this gives a better bound for the total variation distance. We do this by grouping the retained 01-blocks into 0101-blocks. Each 0101-block undergoes a deletion process that is modified because we are conditioning on the event that each of its constituent 01-blocks was not wholly deleted in the first step. By the triangle inequality, instead of coupling the 2-partial traces to each other, we may couple each to 2-partial traces with no defect. The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect, and at the same time, has the remarkable property that what becomes of the other 0101-blocks is unaffected, so that we can couple the defect to that special 0101-block. If we can achieve that, then we use the remaining randomness to couple the numbers of 0101-blocks that are not wholly deleted in the end (these are again binomial random variables). Using a result of Liggett [9], we can find that special 0101-block with high probability. Furthermore, how far the special 0101-block is from the center is controlled, which controls how far apart the binomial distributions are and leads to another factor of $O(n^{-1/4})$ in probability of failure to couple exactly. This is the most subtle part of our proof and requires careful attention to several dependencies.

Combining the two coupling stages gives that the total variation distance between the traces is $O(n^{-3/4})$. Knowing the total variation distance is not sufficient to determine the

number of traces required for reconstruction (it gives a lower bound of only $\Omega(n^{3/4})$ traces; see Appendix A.2). However, by throwing away a very small set of 2-partial traces, applying Lemma A.1, and using properties of the 2-partial traces, we can upgrade our bound on the total variation distance to show that the squared Hellinger distance between the traces is $O(n^{-5/4}\sqrt{\log n})$. This yields the desired lower bound of $\Omega(n^{5/4}/\sqrt{\log n})$ on the number of required traces.

The following lemma encapsulates the overall structure in the proof of Theorem 1.1. The proof of the lemma contains almost all of our work. When we write $\Delta_{\mathbf{x}_n} = \mu_1 + \mu_2 + \mu_3$ (and the corresponding sum for $\Delta_{\mathbf{y}_n}$), we are adding measures as functions on points; no convolution is involved. Recall the notation $\|\cdot\|_{\ell^\infty(\cdot)}$ from (A.3).

LEMMA 2.1. *For all $n \geq 2$, we have $\Delta_{\mathbf{x}_n} = \mu_1 + \mu_2 + \mu_3$ and $\Delta_{\mathbf{y}_n} = \mu_1 + \mu_2' + \mu_3'$, where for some constant C depending only on q,*

$$\mu_3(\mathcal{S}) = \mu_3'(\mathcal{S}) \leq n^{-10}, \tag{2.1}$$

$$\left\| \frac{\mu_2 - \mu_2'}{\mu_1 + \mu_2} \right\|_{\ell^\infty(\mu_1+\mu_2)} \leq Cn^{-1/2}\sqrt{\log n}, \tag{2.2}$$

$$\mu_1(x) + \mu_2(x) = 0 \iff \mu_1(x) + \mu_2'(x) = 0 \quad \text{for each } x \in \mathcal{S}, \tag{2.3}$$

and

$$d_{\mathrm{TV}}(\mu_1 + \mu_2, \mu_1 + \mu_2') \leq Cn^{-3/4}. \tag{2.4}$$

Note that (2.4) and (2.1) imply the upper bound in Proposition 1.4. Before proving Lemma 2.1, we will deduce Theorem 1.1 from the lemma, and we will state and prove Lemmas 2.2 and 2.4, which we use in the proof of Lemma 2.1.

PROOF OF THEOREM 1.1. By Lemma A.3, (A.7) and (2.1),

$$\begin{aligned} d_{\mathrm{H}}^2(\Delta_{\mathbf{x}_n}, \Delta_{\mathbf{y}_n}) &\leq d_{\mathrm{H}}^2(\mu_1 + \mu_2, \mu_1 + \mu_2') + d_{\mathrm{H}}^2(\mu_3, \mu_3') \\ &\leq d_{\mathrm{H}}^2(\mu_1 + \mu_2, \mu_1 + \mu_2') + 2n^{-10}. \end{aligned}$$

Let $\nu := \mu_1 + \mu_2$ and $\mu := \mu_1 + \mu_2'$. By Lemma A.1, (2.3), (2.2) and (2.4), we get

$$\begin{aligned} d_{\mathrm{H}}^2(\mu, \nu) &\leq \mu\{x; \nu(x) = 0\} + 2 \cdot \left\| \frac{\mu(x) - \nu(x)}{\nu(x)} \right\|_{\ell^\infty(\nu)} \cdot d_{\mathrm{TV}}(\mu, \nu) \\ &\leq 0 + 2Cn^{-1/2}\sqrt{\log n} \cdot Cn^{-3/4} \\ &= 2C^2 n^{-5/4}\sqrt{\log n}. \end{aligned}$$

Applying Lemma A.5, we obtain the theorem. □

Write $\mathrm{Bin}(n, s)$ for the binomial distribution corresponding to n trials with success probability s in each trial. We record the following routine calculations for later use.

LEMMA 2.2. *For $n \geq 1$ and $s \in (0, 1)$, let $X \sim \mathrm{Bin}(n, s)$ and $Y \sim \mathrm{Bin}(n-1, s)$. Then*

$$\big|\mathbf{P}[X = k] - \mathbf{P}[Y = k]\big| = \frac{|ns - k|}{n(1-s)} \cdot \mathbf{P}[X = k] \quad \text{for } k = 0, \ldots, n, \tag{2.5}$$

$$\mathbf{P}\big[|X - ns| > c\sqrt{n \log n}\big] \leq 2n^{-2c^2} \quad \text{for } c > 0, \tag{2.6}$$

and

$$d_{\mathrm{TV}}(X,Y) \le \sqrt{\frac{s}{4n(1-s)}}. \tag{2.7}$$

PROOF. Equation (2.5) follows by direct calculation:

$$\begin{aligned}\mathbf{P}[X=k]-\mathbf{P}[Y=k] &= \mathbf{P}[X=k]\cdot\left(1-\frac{n-k}{n(1-s)}\right)\\ &= \mathbf{P}[X=k]\cdot\frac{k-ns}{n(1-s)}.\end{aligned} \tag{2.8}$$

The estimate (2.6) is immediate by the inequality of Hoeffding–Azuma. We obtain (2.7) from (2.5):

$$\begin{aligned}d_{\mathrm{TV}}(X,Y) &= \frac{1}{2}\sum_{k=0}^{n}\bigl|\mathbf{P}[X=k]-\mathbf{P}[Y=k]\bigr| = \frac{1}{2}\sum_{k=0}^{n}\frac{|ns-k|}{n(1-s)}\cdot\mathbf{P}[X=k]\\ &= \frac{1}{2n(1-s)}\mathbf{E}\bigl[|ns-X|\bigr] \le \frac{\mathrm{Var}(X)^{1/2}}{2n(1-s)} = \sqrt{\frac{s}{4n(1-s)}}.\end{aligned}$$

□

The upcoming Lemma 2.4 will allow us to estimate the total variation distance between traces produced from a pair of strings with and without, respectively, a defect. A key role in its proof is played by the following theorem of Liggett [9] that concerns Bernoulli processes. Part (iii) of this theorem is not stated explicitly by Liggett, but follows from the proof of [9], Proposition 2.2 and Theorem 4.25.

THEOREM 2.3 ([9]). *Let $s \in (0,1)$, and let $(a_j)_{j\in\mathbb{Z}}$ be a bi-infinite sequence of i.i.d.* Bernoulli(s) *random variables. Then there is a random variable X supported on $\mathbb{N}_0 := \{0,1,\dots\}$ such that the following hold*:

(i) *The shifted string $(b_j)_{j\in\mathbb{Z}}$ for $b_j := a_{j-X}$ consists of i.i.d.* Bernoulli(s) *random variables, except that $b_0 = 1$ almost surely.*

(ii) *For a constant C depending only on s and for all $m \in \mathbb{N}$, $\mathbf{P}[X > m] \le Cm^{-1/2}$.*

(iii) *Conditional on X and the bits $(a_j)_{j\in\{-X,\dots,0\}}$, all the bits a_j for $j \notin \{-X,\dots,0\}$ are i.i.d.* Bernoulli(s) *random variables.*

Note that one *cannot* choose X so that $(a_j)_{j\ne -X}$ is a Bernoulli(s) process conditioned on X, because that would lead to the contradiction

$$\mathbf{E}\left[\mathbf{E}\left[\sum_{j\le 0} a_j 2^j \,\middle|\, X\right]\right] > \mathbf{E}\left[\sum_{j\le 0} a_j 2^j\right].$$

We will consider strings on the alphabet $\{\alpha,\beta,\gamma\}$. The βs will represent 0101-blocks that become the same as the defect becomes; the αs will represent 0101-blocks that are wholly deleted, and the γs will represent the rest. For a string $\mathbf{w} = (w_1,\dots,w_n) \in \{\alpha,\beta,\gamma\}^n$, let $R(\mathbf{w})$ denote the string obtained by deleting the αs and then contracting the string. In other words, $R(\mathbf{w})$ is obtained by repeating the following procedure for $k = 1,\dots,n$, starting with an empty string:

- If $w_k \in \{\beta,\gamma\}$, copy w_k to the end of $R(\mathbf{w})$ and increase k by one.
- If $w_k = \alpha$, only increase k by one.

LEMMA 2.4. *Let $C_0 > 1$ and $n \in \mathbb{N}$, and let $j_\ell, j_r \in \mathbb{N}$ satisfy $C_0^{-1} n < j_\ell, j_r < C_0 n$. Let $\mathbf{p} := (p_\alpha, p_\beta, p_\gamma) \in (0,1)^3$ be a probability vector on the triple (α, β, γ). Let $\mathbf{w} = (w_{-j_\ell}, \dots, w_{j_r}) \in \{\alpha, \beta, \gamma\}^{j_\ell + 1 + j_r}$ and $\mathbf{w}'' = (w''_{-j_\ell}, \dots, w''_{j_r}) \in \{\alpha, \beta, \gamma\}^{j_\ell + 1 + j_r}$ be strings of length $j_\ell + 1 + j_r$ on the alphabet $\{\alpha, \beta, \gamma\}$ such that the letters w_i and w''_i are i.i.d. with law $\mathbf{p}$:*

$$\mathbf{w} \sim \mathbf{p}^{j_\ell + 1 + j_r} \quad \text{and} \quad \mathbf{w}'' \sim \mathbf{p}^{j_\ell + 1 + j_r}. \tag{2.9}$$

Condition on the event that $w_0 = \beta$.

Then there is a constant C_1 depending only on C_0 and $\mathbf{p}$ such that the total variation distance between $R(\mathbf{w})$ and $R(\mathbf{w}'')$ is bounded above by $C_1 n^{-1/4}$.

PROOF. Throughout the proof, all constants may depend on $(C_0, p_\alpha, p_\beta, p_\gamma)$.

It will be more convenient in the proof to work with bi-infinite strings. Therefore, we assume throughout the proof that $\mathbf{w}$ and $\mathbf{w}''$ are bi-infinite strings $\mathbf{w} = (\dots, w_{-1}, w_0, w_1, \dots)$ and $\mathbf{w}'' = (\dots, w''_{-1}, w''_0, w''_1, \dots)$ with law $\mathbf{p}^{\mathbb{Z}}$ conditioned on the event that $w_0 = \beta$. We will show that the total variation distance between $R((w_{-j_\ell}, \dots, w_{j_r}))$ and $R((w''_{-j_\ell}, \dots, w''_{j_r}))$ is bounded above by $C_1 n^{-1/4}$.

By the result of Liggett stated in Theorem 2.3 above, we can find a random variable X supported on $\mathbb{N}_0$ and independent of $\mathbf{w}$ and a constant $C_2 > 0$ (depending on p_β) such that $\mathbf{P}[X \geq m] \leq C_2 (m+1)^{-1/2}$ for all $m \in \mathbb{N}$ and such that

$$\big(w''_{j-X}\big)_{j \in \mathbb{Z}} \stackrel{d}{=} \mathbf{w}. \tag{2.10}$$

Furthermore, by Theorem 2.3(iii) we may define X so that conditioned on X and the string $(w''_{-X}, \dots, w''_0)$, all letters except $w''_{-X}, \dots, w''_0$ are independent with law $\mathbf{p}$.

Let $B := [X < \lfloor \sqrt{n} \rfloor]$, so that $\mathbf{P}[B^{\mathrm{c}}] \leq C_2 n^{-1/4}$. On the event B, write the strings $\mathbf{w}$ and $\mathbf{w}''$ as concatenations of five strings each:

$$\mathbf{w} = \mathbf{w}_0 \mathbf{w}_1 \mathbf{w}_2 \mathbf{w}_3 \mathbf{w}_4 \quad \text{and} \quad \mathbf{w}'' = \mathbf{w}''_0 \mathbf{w}''_1 \mathbf{w}''_2 \mathbf{w}''_3 \mathbf{w}''_4,$$

where

$$\begin{aligned}
&\mathbf{w}_0 = (\dots, w_{-j_\ell - 1}), \qquad \mathbf{w}_1 = (w_{-j_\ell}, \dots, w_{-1}), \qquad \mathbf{w}_2 = (w_0, \dots, w_{\lfloor \sqrt{n} \rfloor - 1}),\\
&\mathbf{w}_3 = (w_{\lfloor \sqrt{n} \rfloor}, \dots, w_{j_r}), \qquad \mathbf{w}_4 = (w_{j_r + 1}, \dots),\\
&\mathbf{w}''_0 = \big(\dots, w''_{-j_\ell - 1}\big), \qquad \mathbf{w}''_1 = \big(w''_{-j_\ell}, \dots, w''_{-X-1}\big),\\
&\mathbf{w}''_2 = \big(w''_{-X}, \dots, w''_{-X + \lfloor \sqrt{n} \rfloor - 1}\big),\\
&\mathbf{w}''_3 = \big(w''_{-X + \lfloor \sqrt{n} \rfloor}, \dots, w''_{j_r}\big), \qquad \mathbf{w}''_4 = \big(w''_{j_r + 1}, \dots\big).
\end{aligned}$$

See Figure 2 for an illustration. On the event B^{c}, split the strings $\mathbf{w}$ and $\mathbf{w}''$ in the exact same way, except that $\mathbf{w}''_0 = (\dots, w''_{-X-1})$ and that $\mathbf{w}''_1$ is the empty string. By Theorem 2.3(iii),

$$\begin{aligned}
&\text{conditional on } X \text{ and } \mathbf{w}''_2 \text{ and on the event } B, \text{ the letters of the strings}\\
&\mathbf{w}''_0, \mathbf{w}''_1, \mathbf{w}''_3, \mathbf{w}''_4 \text{ are i.i.d. with law } \mathbf{p}.
\end{aligned} \tag{2.11}$$

Let Y_ℓ and Y''_ℓ denote the number of letters of $\mathbf{w}_1$ and $\mathbf{w}''_1$, respectively, that are *not* deleted, that is,

$$\begin{aligned}
Y_\ell &:= \#\big\{j \in \{-j_\ell, \dots, -1\}; w_j \neq \alpha\big\},\\
Y''_\ell &:= \#\big\{j \in \{-j_\ell, \dots, -X-1\}; w''_j \neq \alpha\big\}.
\end{aligned}$$

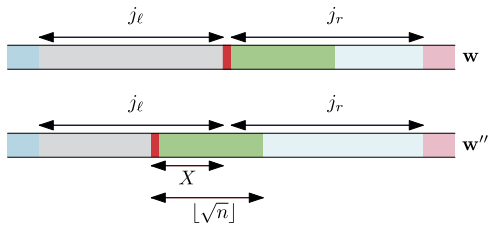

FIG. 2. *The figure illustrates* X, $\mathbf{w} = \mathbf{w}_0\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3\mathbf{w}_4$ *and* $\mathbf{w}'' = \mathbf{w}''_0\mathbf{w}''_1\mathbf{w}''_2\mathbf{w}''_3\mathbf{w}''_4$ *on the event* B. *The string* $\mathbf{w}_0$ (*resp.*, $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4$) *is shown in blue* (*resp.*, *gray*, *green & red*, *cyan and pink*), *and the color code for* $\mathbf{w}''_0, \mathbf{w}''_1, \mathbf{w}''_2, \mathbf{w}''_3, \mathbf{w}''_4$ *is similar. The locations of letters known equal to* β *are shown in red.*

Define Y_r and Y''_r similarly for $\mathbf{w}_3$ and $\mathbf{w}''_3$, that is,

$$Y_r := \#\{j \in \{\lfloor\sqrt{n}\rfloor, \dots, j_r\}; w_j \neq \alpha\},$$
$$Y''_r := \#\{j \in \{-X + \lfloor\sqrt{n}\rfloor, \dots, j_r\}; w''_j \neq \alpha\}.$$

For any given coupling of the strings $\mathbf{w}$ and $\mathbf{w}''$, define the event $B' := [(Y_\ell, Y_r) = (Y''_\ell, Y''_r)]$.

We now define a coupling of the two strings $\mathbf{w}$ and $\mathbf{w}''$ by sampling $\mathbf{w}$ and $\mathbf{w}''$ stepwise on the same probability space as follows. Roughly, we first sample the "central" strings $\mathbf{w}_2$ and $\mathbf{w}''_2$ so that they match, without specifying X. Then we sample X. Then, in case B occurs, we sample the binomial random variables Y_ℓ, Y''_ℓ, Y_r and Y''_r so that B' has as high probability as possible. Finally, we sample the rest of the information in the strings $\mathbf{w}$ and $\mathbf{w}''$. To be precise:

(i) Sample $\mathbf{w}_2$ and $\mathbf{w}''_2$ such that $\mathbf{w}_2 = \mathbf{w}''_2$ and the marginal law of each string is $\mathbf{p}^{\lfloor\sqrt{n}\rfloor}$. This is possible by (2.10).

(ii) Sample X conditioned on $\mathbf{w}_2$ and $\mathbf{w}''_2$. (We have not described explicitly this conditional distribution; also, note that X is not bounded.)

(iii) Sample Y_ℓ, Y''_ℓ, Y_r and Y''_r conditioned on $\mathbf{w}_2$, $\mathbf{w}''_2$ and X with a special joint distribution: First, Y_ℓ and Y_r are independent, as are Y''_ℓ and Y''_r. Second, by (2.11), conditioned on $\mathbf{w}_2$, $\mathbf{w}''_2$ and X, and on the event B, the random variables Y''_ℓ and Y''_r are binomial random variables $Y''_\ell \sim \mathrm{Bin}(j_\ell - X, 1 - p_\alpha)$ and $Y''_r \sim \mathrm{Bin}(j_r + X - \lfloor\sqrt{n}\rfloor + 1, 1 - p_\alpha)$. We couple Y_ℓ, Y''_ℓ, Y_r and Y''_r such that except on an event of conditional probability $d_{\mathrm{TV}}((Y_\ell, Y_r), (Y''_\ell, Y''_r))$ (where we consider the total variation distance conditional on $\mathbf{w}_2$, $\mathbf{w}''_2$, X and B), the event B' occurs. Third, on the event B^{c}, we take the independent coupling of Y_ℓ, Y''_ℓ, Y_r and Y''_r conditioned on $\mathbf{w}_2$, $\mathbf{w}''_2$ and X.

(iv) Sample the remaining randomness conditioned on $\mathbf{w}_2$, $\mathbf{w}''_2$, X, Y_ℓ, Y''_ℓ, Y_r and Y''_r: On the event $B' \cap B$, by (2.11) we may couple $\mathbf{w}$ and $\mathbf{w}''$ so that $R(\mathbf{w}_1) = R(\mathbf{w}''_1)$ and $R(\mathbf{w}_3) = R(\mathbf{w}''_3)$.

By (i) and (iv) of this coupling, we see that on the event $B \cap B'$,

$$R\big((w_{-j_\ell}, \dots, w_{j_r})\big) = R(\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3) = R\big(\mathbf{w}''_1\mathbf{w}''_2\mathbf{w}''_3\big) = R\big((w''_{-j_\ell}, \dots, w''_{j_r})\big).$$

To conclude the proof, it is therefore sufficient to show that $\mathbf{P}[B \cap B'] \geq 1 - C_1 n^{-1/4}$ for some constant C_1.

By (2.7), (A.8) (with $n = 2$ there), (iii) of the coupling, and the fact that $Y_\ell \sim \mathrm{Bin}(j_\ell, 1 - p_\alpha)$ and $Y_r \sim \mathrm{Bin}(j_r - \lfloor\sqrt{n}\rfloor + 1, 1 - p_\alpha)$, the total variation distance between (Y_ℓ, Y_r) and (Y''_ℓ, Y''_r) conditional on X and on the event B is at most $C_3 X/\sqrt{n}$ for some constant C_3 that depends on C_0 and p_α. Summing over the possible values of X, we get the following for

some constant $C_4 > 0$:

$$\mathbf{P}[B \cap (B')^{\mathrm{c}}] \le C_3 \mathbf{E}[\mathbf{1}_B X/\sqrt{n}] = C_3 n^{-1/2} \sum_{k=0}^{\lfloor n^{1/2} \rfloor - 1} \mathbf{P}[X > k]$$
$$\le C_4 n^{-1/2} \cdot \left(n^{1/2}\right)^{1/2} = C_4 n^{-1/4}.$$

Combining this with $\mathbf{P}[B^{\mathrm{c}}] \le C_2 n^{-1/4}$, we obtain $\mathbf{P}[B \cap B'] \ge 1 - C_1 n^{-1/4}$ for some constant C_1, which concludes the proof. □

PROOF OF LEMMA 2.1. We will always couple the deletions made to the defects so that they are the same. If both defects are wholly deleted, then the remaining strings obviously can be coupled to have the exact same traces; this occurs with probability q^4 and forms part of the measure μ_1 that we need to define. It will be most convenient from now on to *condition on the event that neither defect is wholly deleted*.

A trace may be constructed in three steps (see Figure 3 for an illustration):

(I) First, we construct the 2-*partial trace*. A 01-*block* (resp., 10-*block*) is the string of length two given by $(0, 1)$ (resp., $(1, 0)$). The input string $\mathbf{x}_n$ may be viewed as the concatenation of $n - 1$ 01-blocks, followed by a single 10-block and then n 01-blocks. We sample the 2-partial trace of $\mathbf{x}_n$ by setting $s := 1 - q^2/2$, letting $Y_\ell \sim \mathrm{Bin}(n-1, s)$ and $Y_r \sim \mathrm{Bin}(n, s)$ be independent binomial random variables, and defining the partial trace to be the concatenation of Y_ℓ 01-blocks, followed by a single 10-block and then Y_r 01-blocks. The partial trace of $\mathbf{y}_n$ is defined in the exact same way, except that $Y_\ell \sim \mathrm{Bin}(n, s)$ and $Y_r \sim \mathrm{Bin}(n-1, s)$.

(II) Given a 2-partial trace, we define the 4-*partial trace* by the following deterministic procedure. Defining a 0101-*block* to be the length-4 string $(0, 1, 0, 1)$, the 4-partial trace associated with the 2-partial trace in (I) is the concatenation of the following blocks in the listed order:

- if Y_ℓ is odd, a 01-block,
- $\lfloor Y_\ell/2 \rfloor$ 0101-blocks,
- one 10-block (the defect),
- $\lfloor Y_r/2 \rfloor$ 0101-blocks,
- if Y_r is odd, a 01-block.

(III) From the 4-partial trace, we construct the final traces $\widetilde{\mathbf{x}}_n$ and $\widetilde{\mathbf{y}}_n$ as follows, where we treat each block independently and obtain a string in $\mathcal{S}$ by concatenating the bits of the various blocks in the same order as they appear in the 4-partial trace.

- A 01-block is replaced by 01, 1, 0, $\varnothing$ with probability p^2/s, pq/s, pq/s, $q^2/(2s)$, respectively, where $\varnothing$ denotes the trivial (length zero) string.

$$\begin{aligned}
\mathbf{x_6} &= (0,1,0,1,0,1,0,1,0,1,\mathbf{1},\mathbf{0},0,1,0,1,0,1,0,1,0,1,0,1) \in \mathcal{S}_{24} \\
&\quad (01,01,01,01,01,\mathbf{10},01,01,01,01,01,01) \in \mathcal{S}^{\mathrm{p}} \\
&\quad (01,01,01,01,\mathbf{10},01,01,01) \in \mathcal{S}^{\mathrm{p}} \quad \text{(2-partial trace)} \\
&\quad (0101,0101,\mathbf{10},0101,01) \in \mathcal{S}^{\mathrm{p}} \quad \text{(4-partial trace)} \\
&\quad (001,01,\mathbf{0},00,1) \in \mathcal{S}^{\mathrm{p}} \\
\widetilde{\mathbf{x}}_6 &= (0,0,1,0,1,\mathbf{0},0,0,1) \in \mathcal{S}
\end{aligned}$$

FIG. 3. *The figure illustrates the steps* (I)–(III) *for constructing a trace as described in the proof of Lemma* 2.1. *The steps from the second line to the third line and from the fourth line to the fifth line are random, while the other steps are deterministic. In the displayed example, we have* $Y_\ell = 4$ *and* $Y_r = 3$. *The defect is colored in red to simplify the reading of the figure but, of course, is not part of the information in the actual trace.*

- A 0101-block is first replaced by two 01-blocks, and then each 01-block is treated independently as in the preceding bullet point. The result is a block in the set

$$\mathcal{B}_{0101} := \{0101, 101, 011, 001, 010, 01, 10, 11, 00, 0, 1, \varnothing\}.$$

- The 10-block representing the defect is replaced by 10,1,0 with probability $p^2/(1-q^2)$, $pq/(1-q^2)$, $pq/(1-q^2)$, respectively.

Let $\mathcal{S}^{\mathrm{p}}$ denote the set of strings $\mathbf{w} = (\mathbf{u}_1, \dots, \mathbf{u}_\ell)$ for $\ell \in \mathbb{N}_0$, where each $\mathbf{u}_j$ is a 01-block, a 0101-block, or a 10-block; the case $\ell = 0$ corresponds to the empty string, $\varnothing$. In particular, both the 2-partial trace and the 4-partial trace considered above are contained in $\mathcal{S}^{\mathrm{p}}$. Notice that (III) provides a general procedure for obtaining a random string in $\mathcal{S}$ from a string $\mathbf{w} \in \mathcal{S}^{\mathrm{p}}$; let $\Delta^{\mathrm{p}}(\mathbf{w})$ denote the resulting law on strings in $\mathcal{S}$.

Let ν and ν' denote the laws of the 4-partial traces associated with $\mathbf{x}_n$ and $\mathbf{y}_n$, respectively. Then ν and ν' are probability measures on $\mathcal{S}^{\mathrm{p}}$. Notice that sampling the 4-partial traces above is equivalent to sampling the random variables Y_ℓ and Y_r describing the numbers of 01-blocks on either side of the defect in the associated 2-partial trace. We decompose

$$\nu = \nu_1 + \nu_2 + \nu_3 \quad \text{and} \quad \nu' = \nu_1 + \nu_2' + \nu_3', \tag{2.12}$$

where the measures ν_1, ν_2, ν_2', ν_3, ν_3' are defined as follows. The measures ν_3 and ν_3' correspond to the events that unusually many or few 01-blocks were deleted on at least one side of the defect. More precisely, for an appropriate constant C_0 to be defined later, let the event A be given by

$$\begin{aligned} A &:= \bigl\{|Y_\ell - a_n| \le C_0\sqrt{n\log n}, |Y_r - a_n| \le C_0\sqrt{n\log n}\bigr\} \\ &\quad \text{where } a_n := n(1-q^2/2). \end{aligned} \tag{2.13}$$

For an arbitrary measure $\widehat{\nu}$ on a measurable space $\widehat{\mathcal{S}}$ and with $\widehat{A} \subset \widehat{\mathcal{S}}$, let $\mathbf{1}_{\widehat{A}}\widehat{\nu}$ denote the measure that assigns mass $\widehat{\nu}(U \cap \widehat{A})$ to any measurable set $U \subset \widehat{\mathcal{S}}$. Define

$$\nu_3 := \mathbf{1}_{A^{\mathrm{c}}}\nu \quad \text{and} \quad \nu_3' := \mathbf{1}_{A^{\mathrm{c}}}\nu'.$$

Now choose the measures ν_1, ν_2 and ν_2' so that (2.12) is satisfied and $\nu_1(\mathcal{S}^{\mathrm{p}})$ is maximized. In particular, the measures ν_2 and ν_2' have disjoint support and $\nu_2(\mathcal{S}^{\mathrm{p}}) = \nu_2'(\mathcal{S}^{\mathrm{p}}) = d_{\mathrm{TV}}(\nu - \nu_3, \nu' - \nu_3')$. Note that the distribution of the total number of 01-blocks is the same under ν_2 as it is under ν_2'.

By choosing C_0 sufficiently large and applying equation (2.6), we obtain

$$\nu_3(\mathcal{S}^{\mathrm{p}}) = \nu_3'(\mathcal{S}^{\mathrm{p}}) \le n^{-10}. \tag{2.14}$$

By equation (2.5), for an appropriate constant C_1, for all $n \ge 2$, and for all $x \in \mathcal{S}^{\mathrm{p}}$ such that $(\nu_1 + \nu_2)(x) \ne 0$,

$$\begin{aligned} &\frac{|\nu_2(x) - \nu_2'(x)|}{(\nu_1+\nu_2)(x)} \\ &\quad \le \max_{k_1,k_2 \in \{0,\dots,\lfloor C_0\sqrt{n\log n}\rfloor\}} \left|1 - \frac{\mathbf{P}_{\mathbf{y}_n}[|Y_\ell - a_n| = k_1, |Y_r - a_n| = k_2]}{\mathbf{P}_{\mathbf{x}_n}[|Y_\ell - a_n| = k_1, |Y_r - a_n| = k_2]}\right| \\ &\quad \le C_1 n^{-1/2}\sqrt{\log n}. \end{aligned} \tag{2.15}$$

Furthermore,

$$\nu_1(x) + \nu_2(x) = 0 \iff \nu_1(x) + \nu_2'(x) = 0 \quad \forall x \in \mathcal{S}^{\mathrm{p}} \tag{2.16}$$

since each of these conditions holds if and only if the binomial random variables Y_ℓ and Y_r associated with x satisfy the condition of (2.13). Finally, (2.7) and (2.14) give that upon increasing C_1 if necessary

$$\nu_2(\mathcal{S}^{\mathrm{p}}) = \nu_2'(\mathcal{S}^{\mathrm{p}}) = d_{\mathrm{TV}}(\nu - \nu_3, \nu' - \nu_3') \le C_1 n^{-1/2}. \tag{2.17}$$

The bounds in the preceding paragraph are expressed in term of measures on 4-partial traces. We now transfer these bounds to the final traces. Let $\widetilde{\mu}_1 := \nu_1(\mathcal{S}^{\mathrm{p}}) \cdot \Delta^{\mathrm{p}}(\mathbf{w})$ for $\mathbf{w} \sim \nu_1(\mathcal{S}^{\mathrm{p}})^{-1}\nu_1$. Define $\widetilde{\mu}_2$, $\widetilde{\mu}_3$, $\widetilde{\mu}_2'$ and $\widetilde{\mu}_3'$ similarly from the measures ν_2, ν_3, ν_2' and ν_3'. Then

$$\Delta_{\mathbf{x}_n} = \widetilde{\mu}_1 + \widetilde{\mu}_2 + \widetilde{\mu}_3 \quad \text{and} \quad \Delta_{\mathbf{y}_n} = \widetilde{\mu}_1 + \widetilde{\mu}_2' + \widetilde{\mu}_3'. \tag{2.18}$$

By (2.14) and (2.17), we have

$$\widetilde{\mu}_3(\mathcal{S}) = \widetilde{\mu}_3'(\mathcal{S}) \le n^{-10} \quad \text{and} \quad \widetilde{\mu}_2(\mathcal{S}) = \widetilde{\mu}_2'(\mathcal{S}) \le C_1 n^{-1/2}. \tag{2.19}$$

Furthermore, by (2.16),

$$\widetilde{\mu}_1(x) + \widetilde{\mu}_2(x) = 0 \iff \widetilde{\mu}_1(x) + \widetilde{\mu}_2'(x) = 0 \quad \forall x \in \mathcal{S}. \tag{2.20}$$

Finally, by Lemma A.2 and (2.15), with $\nu_A := \mathbf{1}_A \nu = \nu_1 + \nu_2$ and $\nu_A' := \mathbf{1}_A \nu' = \nu_1' + \nu_2'$,

$$\left\| \frac{\widetilde{\mu}_2 - \widetilde{\mu}_2'}{\widetilde{\mu}_1 + \widetilde{\mu}_2} \right\|_{\ell^\infty(\widetilde{\mu}_1 + \widetilde{\mu}_2)} \le \left\| \frac{\nu_A - \nu_A'}{\nu_A} \right\|_{\ell^\infty(\nu_A)} \le C_1 n^{-1/2} \sqrt{\log n}. \tag{2.21}$$

The preceding paragraph provides a coupling of $\Delta_{\mathbf{x}_n}$ and $\Delta_{\mathbf{y}_n}$ such that the traces are identical with probability $\widetilde{\mu}_1(\mathcal{S}) > 1 - C_1 n^{-1/2}\sqrt{\log n} - n^{-10}$. To obtain (2.4), we construct a better coupling by making a second attempt to couple the traces on the event that the first coupling fails.

Let

$$\sigma := \widetilde{\mu}_2(\mathcal{S})^{-1}\widetilde{\mu}_2 \quad \text{and} \quad \sigma' := \widetilde{\mu}_2'(\mathcal{S})^{-1}\widetilde{\mu}_2'$$

denote the laws of the traces on the event that the first coupling attempt failed and that the event A occurs. We will argue that for an appropriate constant C_2,

$$d_{\mathrm{TV}}(\sigma, \sigma') \le 2C_2 n^{-1/4}. \tag{2.22}$$

Before proving (2.22), we explain how (2.22) implies the lemma.

Assuming (2.22) holds, by (2.19) we can write $\widetilde{\mu}_2 = \overline{\mu}_2 + \mu_2$ and $\widetilde{\mu}_2' = \overline{\mu}_2 + \mu_2'$, where

$$\mu_2(\mathcal{S}) = \mu_2'(\mathcal{S}) \le 2C_2 n^{-1/4} \cdot C_1 n^{-1/2} = 2C_1 C_2 n^{-3/4}. \tag{2.23}$$

With $\mu_1 := \widetilde{\mu}_1 + \overline{\mu}_2$, $\mu_3 := \widetilde{\mu}_3$ and $\mu_3' := \widetilde{\mu}_3'$, all the requirements of Lemma 2.1 are satisfied because of (2.18), (2.19), (2.20), (2.21) and (2.23). Note in particular that (2.2) is satisfied because for any $x \in \mathcal{S}$ for which $\mu_1(x) + \mu_2(x) \ne 0$,

$$\begin{aligned} \frac{|\mu_2(x) - \mu_2'(x)|}{\mu_1(x) + \mu_2(x)} &= \frac{|(\widetilde{\mu}_2(x) - \overline{\mu}_2(x)) - (\widetilde{\mu}_2'(x) - \overline{\mu}_2(x))|}{(\widetilde{\mu}_1(x) + \overline{\mu}_2(x)) + (\widetilde{\mu}_2(x) - \overline{\mu}_2(x))} \\ &= \frac{|\widetilde{\mu}_2(x) - \widetilde{\mu}_2'(x)|}{\widetilde{\mu}_1(x) + \widetilde{\mu}_2(x)} \le C_1 n^{-1/2}\sqrt{\log n}. \end{aligned}$$

We will now prove (2.22). Let $\mathbf{w}$ denote the 4-partial trace associated with $\mathbf{x}_n$. Let $\mathbf{w}'' \in \mathcal{S}^{\mathrm{p}}$ be identical in law to $\mathbf{w}$ except that the 10-block (i.e., the defect) is replaced by a 0101-block, and denote by σ'' the law of $\Delta^{\mathrm{p}}(\mathbf{w}'')$. Because the lengths of the 4-partial traces associated to $\mathbf{x}_n$ or to $\mathbf{y}_n$ have the same law, we have symmetry in $\mathbf{x}_n$ and $\mathbf{y}_n$ and may apply the triangle inequality for d_{TV}. That is, it suffices to show the following in order to prove (2.22):

$$d_{\mathrm{TV}}(\sigma, \sigma'') \le C_2 n^{-1/4}. \tag{2.24}$$

When proving (2.24), we condition on Y_ℓ and Y_r, so these random variables are viewed as constants. In particular, we will take the lengths of $\mathbf{w}$ and $\mathbf{w}''$ to be the same, that is, the number of blocks in the two 4-partial traces is the same. We will construct a coupling of $\Delta^{\mathrm{p}}(\mathbf{w})$ and $\Delta^{\mathrm{p}}(\mathbf{w}'')$ by sampling these random variables stepwise.

Assume first that Y_ℓ and Y_r are both even, namely, $2j_\ell$ and $2j_r$, respectively, for $j_\ell, j_r \in \mathbb{N}_0$. For each block $\mathbf{u} \in \mathcal{B}_{0101}$, let $p_{\mathbf{u}}$ denote the probability that a 0101-block reduces to $\mathbf{u}$ in the definition of Δ^{p}. The trace $\Delta^{\mathrm{p}}(\mathbf{w}'')$ may be sampled in the following four steps:

- Sample the block $\mathbf{u}_{\mathrm{d}} \in \{10, 1, 0\}$ that replaces the defect in $\mathbf{w}$, using probabilities as in the third bullet point of (III) above.
- Let $(a_j)_{j\in\{-j_\ell,\dots,j_r\}}$ be an i.i.d. sequence such that $a_j \in \{\alpha, \beta, \gamma\}$ for each j and such that

$$\mathbf{P}[a_j = \alpha] = p_{\varnothing}, \qquad \mathbf{P}[a_j = \beta] = p_{\mathbf{u}_{\mathrm{d}}}, \qquad \mathbf{P}[a_j = \gamma] = 1 - p_{\varnothing} - p_{\mathbf{u}_{\mathrm{d}}}. \tag{2.25}$$

- Let the jth block in $\mathbf{w}''$ reduce to $\varnothing$ (resp., $\mathbf{u}_{\mathrm{d}}$) if $a_j = \alpha$ (resp., $a_j = \beta$).
- If $a_j = \gamma$, then the jth block in $\mathbf{w}''$ reduces to any given block $\mathbf{u} \in \mathcal{B}_{0101} \setminus \{\varnothing, \mathbf{u}_{\mathrm{d}}\}$ with probability $p_{\mathbf{u}}/\mathbf{P}[a_j = \gamma]$, independently of what the other blocks reduce to.

The trace $\Delta^{\mathrm{p}}(\mathbf{w})$ may be sampled in the exact same way, except that we condition on the event that $a_0 = \beta$. Recall the function R defined preceding the statement of Lemma 2.4. Let ρ'' denote the law of $R((a_j)_{j\in\{-j_\ell,\dots,j_r\}})$, where the a_js are i.i.d. given by (2.25), and let ρ denote ρ'' conditioned on $a_0 = \beta$. By equation (A.6),

$$d_{\mathrm{TV}}(\sigma, \sigma'') \le d_{\mathrm{TV}}(\rho, \rho'').$$

By Lemma 2.4, we have $d_{\mathrm{TV}}(\rho, \rho'') \le C_2 n^{-1/4}$ for some constant C_2, which gives (2.22).

To conclude the proof, we briefly explain which modifications are needed to the above proof in the case where Y_ℓ or Y_r is odd. In this case, the 4-partial traces $\mathbf{w}$ and $\mathbf{w}''$ will contain one or two 01-blocks. The total variation distance between $\Delta^{\mathrm{p}}(\mathbf{w})$ and $\Delta^{\mathrm{p}}(\mathbf{w}'')$ will be the same in this case as before since we simply couple the 01-blocks of $\mathbf{w}$ and $\mathbf{w}''$ together so that they always reduce to the same block in $\{01, 0, 1, \varnothing\}$. □

3. Upper bound: Proof of Proposition 1.2. In this section, we prove the lower bound of Proposition 1.4, which immediately implies Proposition 1.2.

The idea in the proof of Proposition 1.4 is to define an integer-valued random variable $Z(\widetilde{\mathbf{x}})$ that is a function of the trace $\widetilde{\mathbf{x}}$ and such that $d_{\mathrm{TV}}(Z(\widetilde{\mathbf{x}}), Z(\widetilde{\mathbf{y}}))$ can be bounded from below. For $n \in \mathbb{N}$, $\mathbf{x} \in \mathcal{S}_{4n}$ and $\widetilde{\mathbf{x}} = (\widetilde{x}_1, \dots, \widetilde{x}_\ell)$ the trace of $\mathbf{x}$, define $Z(\widetilde{\mathbf{x}})$ as

$$Z(\widetilde{\mathbf{x}}) := \#\{k; 2np + 1 \le k \le (2np + \sqrt{npq}) \wedge (\ell - 1), \widetilde{x}_k = \widetilde{x}_{k+1} = 1\}. \tag{3.1}$$

We will use several lemmas in the proof of Proposition 1.4.

LEMMA 3.1. *We have* $\mathbf{E}_{\mathbf{y}_n}[Z(\widetilde{\mathbf{y}}_n)] - \mathbf{E}_{\mathbf{x}_n}[Z(\widetilde{\mathbf{x}}_n)] = \Theta(n^{-1/2})$ *and* $\mathbf{E}_{\mathbf{y}_n}[Z(\widetilde{\mathbf{y}}_n)] > \mathbf{E}_{\mathbf{x}_n}[Z(\widetilde{\mathbf{x}}_n)]$ *for all sufficiently large* n.

PROOF. Let $E(j,k)$ be the event that bit j in the input string is copied to position k in the trace. If one or both of the positions are not well defined (i.e., if $j \notin \{1, \dots, 4n\}$ or if k is smaller than 1 or larger than the length of the trace), then let $E(j,k)$ be the empty event. If $j, k \in \{1, \dots, 4n\}$, then

$$\mathbf{P}_{\mathbf{x}_n}[E(j,k)] = \mathbf{P}_{\mathbf{y}_n}[E(j,k)] = \binom{j-1}{k-1} p^{k-1} q^{j-k} \cdot p. \tag{3.2}$$

Let $\widetilde{x}_k$ (resp., $\widetilde{y}_k$) denote bit number k of $\widetilde{\mathbf{x}}_n$ (resp., $\widetilde{\mathbf{y}}_n$). Assume we send the strings $\mathbf{x}_n$ and $\mathbf{y}_n$ through the deletion channel, and that the indices of the deleted bits are exactly the same

for the two strings. Then the events $[\widetilde{x}_k = \widetilde{x}_{k+1} = 1]$ and $[\widetilde{y}_k = \widetilde{y}_{k+1} = 1]$ may differ only due to occurrence of the events $E(2n+1,k)$ or $E(2n-1,k)$ (which give $\widetilde{y}_k = 1$ and $\widetilde{x}_k = 1$, resp.), or due to occurrence of the events $E(2n+1,k+1)$ or $E(2n-1,k+1)$ (which give $\widetilde{y}_{k+1} = 1$ and $\widetilde{x}_{k+1} = 1$, resp.). Therefore,

$$
\begin{aligned}
&\mathbf{E}_{\mathbf{y}_n}\big[Z(\widetilde{\mathbf{y}}_n)\big] - \mathbf{E}_{\mathbf{x}_n}\big[Z(\widetilde{\mathbf{x}}_n)\big] \\
&\quad= \sum_{2np+1 \le k \le 2np+\sqrt{npq}} \big(\mathbf{P}_{\mathbf{y}_n}\big[E(2n+1,k) \cap \{\widetilde{y}_{k+1} = 1\}\big] \\
&\qquad - \mathbf{P}_{\mathbf{x}_n}\big[E(2n-1,k) \cap \{\widetilde{x}_{k+1} = 1\}\big] \\
&\qquad + \mathbf{P}_{\mathbf{y}_n}\big[\{\widetilde{y}_k = 1\} \cap E(2n+1,k+1)\big] \\
&\qquad - \mathbf{P}_{\mathbf{x}_n}\big[\{\widetilde{x}_k = 1\} \cap E(2n-1,k+1)\big]\big).
\end{aligned} \tag{3.3}
$$

First, we estimate the sum in (3.3) restricted to only the first two terms in each summand. Notice that $\mathbf{x}_n$ restricted to bits $\{2n-1, 2n, \dots, 4n-2\}$ is identical to $\mathbf{y}_n$ restricted to bits $\{2n+1, 2n+2, \dots, 4n\}$. On the event $E(2n+1,k)$, the value of $\widetilde{y}_{k+1}$ can be obtained by sending bits $\{2n+2, 2n+3, \dots, 4n\}$ of $\mathbf{y}_n$ through the deletion channel and recording the first bit, and the analogous statement holds for $E(2n-1,k)$, $\widetilde{x}_{k+1}$, $\mathbf{x}_n$ and $\{2n, 2n+1, \dots, 4n\}$. Therefore, if $\mathbf{x}_n^\dagger$ is the string identical to $\mathbf{x}_n$ but with the last two bits removed, we have $\mathbf{P}_{\mathbf{y}_n}[\widetilde{y}_{k+1} = 1 | E(2n+1,k)] = \mathbf{P}_{\mathbf{x}_n^\dagger}[\widetilde{x}_{k+1}^\dagger = 1 | E(2n-1,k)]$. The probability that the bits $\{4n-1, 4n\}$ of $\mathbf{x}_n$ affect the value of $\widetilde{y}_{k+1}$ conditioned on the event $E(2n-1,k)$ is $O(q^{2n})$. Using these observations and that the considered probabilities are of order 1, we get

$$
\frac{\mathbf{P}_{\mathbf{y}_n}[\widetilde{y}_{k+1} = 1 | E(2n+1,k)]}{\mathbf{P}_{\mathbf{x}_n}[\widetilde{x}_{k+1} = 1 | E(2n-1,k)]} = 1 + O\big(q^{2n}\big).
$$

This and (3.2) give, with $\xi := k - 2np \in [1, \sqrt{npq}\,]$,

$$
\begin{aligned}
&\frac{\mathbf{P}_{\mathbf{y}_n}[E(2n+1,k) \cap \{\widetilde{y}_{k+1} = 1\}]}{\mathbf{P}_{\mathbf{x}_n}[E(2n-1,k) \cap \{\widetilde{x}_{k+1} = 1\}]} \\
&\quad= \frac{\mathbf{P}_{\mathbf{y}_n}[E(2n+1,k)]}{\mathbf{P}_{\mathbf{x}_n}[E(2n-1,k)]} \cdot \frac{\mathbf{P}_{\mathbf{y}_n}[\widetilde{y}_{k+1} = 1 | E(2n+1,k)]}{\mathbf{P}_{\mathbf{x}_n}[\widetilde{x}_{k+1} = 1 | E(2n-1,k)]} \\
&\quad= \frac{2n(2n-1)q^2}{(2n-k)(2n-k+1)}\big(1 + O\big(q^{2n}\big)\big) \\
&\quad= \frac{(1 - 1/(2n))}{(1 - \xi/(2nq))(1 - (\xi-1)/(2nq))}\big(1 + O\big(q^{2n}\big)\big) \\
&\quad= 1 + \Theta(\xi/n),
\end{aligned}
$$

and that the ratio on the left-hand side is greater than 1 for sufficiently large n. Using this and that

$$
\begin{aligned}
&\mathbf{P}_{\mathbf{x}_n}\big[E(2n-1,k) \cap \{\widetilde{x}_{k+1} = 1\}\big] \\
&\quad= \mathbf{P}_{\mathbf{x}_n}\big[E(2n-1,k)\big] \cdot \mathbf{P}_{\mathbf{x}_n}\big[\widetilde{x}_{k+1} = 1 | E(2n-1,k)\big] \\
&\quad= \Theta\bigg(\frac{1}{\sqrt{n}}\bigg) \quad \text{when } 2np+1 \le k \le 2np + \sqrt{npq},
\end{aligned}
$$

we get

$$\begin{aligned}
&\sum_{2np+1\le k\le 2np+\sqrt{npq}} \big(\mathbf{P}_{\mathbf{y}_n}\big[E(2n+1,k)\cap\{\widetilde{y}_{k+1}=1\}\big] \\
&\qquad -\mathbf{P}_{\mathbf{x}_n}\big[E(2n-1,k)\cap\{\widetilde{x}_{k+1}=1\}\big]\big) \\
&\qquad =\Theta\Big(\sum_{2np+1\le k\le 2np+\sqrt{npq}} \mathbf{P}_{\mathbf{x}_n}\big[E(2n-1,k)\cap\{\widetilde{x}_{k+1}=1\}\big]\cdot\frac{\xi}{n}\Big) \\
&\qquad =\Theta\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned} \tag{3.4}$$

and that the left-hand side of (3.4) is positive for all large n.

Now we bound the sum in (3.3) restricted to only the third and the fourth term in each summand, that is, we bound the sum

$$\begin{aligned}
&\sum_{2np+1\le k\le 2np+\sqrt{npq}} \big(\mathbf{P}_{\mathbf{y}_n}\big[\{\widetilde{y}_k=1\}\cap E(2n+1,k+1)\big] \\
&\qquad -\mathbf{P}_{\mathbf{x}_n}\big[\{\widetilde{x}_k=1\}\cap E(2n-1,k+1)\big]\big) \\
&\qquad =\sum_{2np+1\le k\le 2np+\sqrt{npq}}\ \sum_{\substack{j\le 2n-1,\\ j\text{ odd}}} \big(\mathbf{P}_{\mathbf{y}_n}\big[E(j+1,k)\cap E(2n+1,k+1)\big] \\
&\qquad -\mathbf{P}_{\mathbf{x}_n}\big[E(j-1,k)\cap E(2n-1,k+1)\big]\big).
\end{aligned} \tag{3.5}$$

First, we notice that the contribution in (3.5) from the terms for which $|j-2n|>\sqrt{npq}$ is $q^{O(\sqrt{n})}$, since all the bits whose position is in $\{j+2,\dots,2n-2\}$ are deleted on this event, whence

$$\mathbf{P}_{\mathbf{x}_n}\big[E(j-1,k)\cap E(2n-1,k+1)\big]\le q^{2n-j-3},$$

and a similar bound holds for $\mathbf{y}_n$. Therefore, in the remainder of the proof, we will consider only the terms of (3.5) for which $|j-2n|\le\sqrt{npq}$. Notice that this condition implies $|jp-k|\le|jp-2np|+|2np-k|\le 2\sqrt{npq}$. By the definition of the events $E(\cdot,\cdot)$, we have for $j<2n$,

$$\begin{aligned}
\mathbf{P}_{\mathbf{y}_n}\big[E(2n+1,k+1)|E(j+1,k)\big]&=q^{2n-j-1}p \\
&=\mathbf{P}_{\mathbf{x}_n}\big[E(2n-1,k+1)|E(j-1,k)\big].
\end{aligned}$$

Using this and that $|jp-k|\le 2\sqrt{npq}$, writing $X\sim\mathrm{Bin}(j-2,p)$ and $Y\sim\mathrm{Bin}(j,p)$, we have

$$\begin{aligned}
&\frac{\mathbf{P}_{\mathbf{y}_n}[E(j+1,k)\cap E(2n+1,k+1)]}{\mathbf{P}_{\mathbf{x}_n}[E(j-1,k)\cap E(2n-1,k+1)]} \\
&\qquad =\frac{\mathbf{P}_{\mathbf{y}_n}[E(j+1,k)]}{\mathbf{P}_{\mathbf{x}_n}[E(j-1,k)]}=\frac{\mathbf{P}[Y=k-1]}{\mathbf{P}[X=k-1]} \\
&\qquad =1+O\Big(\frac{|jp-k|+1}{n}\Big)\le 1+O\Big(\frac{1}{\sqrt{n}}\Big),
\end{aligned} \tag{3.6}$$

where we apply (2.5) in the second-to-last step. Furthermore, by (2.8) and the fact that $\mathbf{E}[X]<\mathbf{E}[Y]=jp\le(2n-1)p<2np\le k-1$, we get $\mathbf{P}[Y=k-1]>\mathbf{P}[X=k-1]$,

so the left-hand side of (3.6) is greater than 1. Now we get that the right-hand side of (3.5) is positive for large n and bounded above by

$$\sum_{2np+1\le k\le 2np+\sqrt{npq}} \sum_{\substack{j\le 2n-1,\\ j \text{ odd}}} \mathbf{P}_{\mathbf{x}_n}\big[E(j-1,k)\cap E(2n-1,k+1)\big]\cdot O\Big(\frac{1}{\sqrt{n}}\Big)$$

$$= O\Big(\frac{1}{\sqrt{n}}\Big).$$

Combining this with (3.4) gives the lemma. □

LEMMA 3.2. *There is a constant $c>0$ depending only on q such that for all $r>0$ and $n\in\mathbb{N}$,*

$$\mathbf{P}\big[\big|Z(\widetilde{\mathbf{y}}_n)-\mathbf{E}\big[Z(\widetilde{\mathbf{y}}_n)\big]\big|>rn^{1/4}\big]\le 2e^{-cr^2} \quad \textit{and}$$

$$\mathbf{P}\big[\big|Z(\widetilde{\mathbf{x}}_n)-\mathbf{E}\big[Z(\widetilde{\mathbf{x}}_n)\big]\big|>rn^{1/4}\big]\le 2e^{-cr^2}.$$

PROOF. We will prove the result only for $\mathbf{x}_n$ since the proof for $\mathbf{y}_n$ is identical, and we write Z instead of $Z(\mathbf{x}_n)$ to simplify notation. Recall from Notation 1.6 that all constants $c_1, c_2, \dots$ may depend on q but on no other parameters.

First, we prove a concentration result for a random variable V that is closely related to Z. Let $\mathbf{w} := (w_1, w_2, \dots) = (01)^{\mathbb{N}}$ be a half-infinite bit string with period 01, and let $\widetilde{\mathbf{w}} := (\widetilde{w}_1, \widetilde{w}_2, \dots)$ denote the trace obtained by sending $\mathbf{w}$ through the deletion channel with deletion probability q. Then set

$$V := \#\big\{k\in[1,\sqrt{npq}];\, \widetilde{w}_k=\widetilde{w}_{k+1}=1\big\}.$$

For $j\in\mathbb{N}$, let $u_j\sim$ Bernoulli(p) be the indicator that bit j of $\mathbf{w}$ is not deleted. Let E be the event that at least $\sqrt{npq}$ bits are *not* deleted among the first $m := \lceil 2\sqrt{npq}/p\rceil$ bits of the trace, that is,

$$E := \Big[\sum_{j=1}^{m} u_j \ge \sqrt{npq}\Big].$$

Then $\mathbf{P}[E^{\mathrm{c}}]\le \exp(-c_1\sqrt{n})$ for some constant $c_1>0$ by a large-deviations bound.

Notice that $V\mathbf{1}_E$ can be written as a function of $u_1,\dots,u_m$. Furthermore, changing one u_j changes $V\mathbf{1}_E$ by at most 2 if both $u_1,\dots,u_m$ and the modified sequence lie in the event E. By [4], which is a variant of McDiarmid's inequality when differences are bounded with high probability, there is a constant $c_1'>0$ such that

$$\mathbf{P}\big[\big|V\mathbf{1}_E-\mathbf{E}[V\mid E]\big|>rn^{1/4}\big]\le 2\exp\big(-c_1'r^2\big)\quad \forall r>0.$$

Because $\mathbf{E}[V\mathbf{1}_E]\le\mathbf{E}[V\mid E]\le\mathbf{E}[V\mathbf{1}_E]+\sqrt{npq}\exp(-c_1\sqrt{n})/(1-e^{-c_1})$, it follows that there is a constant $c_2>0$ such that

$$\mathbf{P}\big[\big|V\mathbf{1}_E-\mathbf{E}[V\mathbf{1}_E]\big|>rn^{1/4}\big]\le 2\exp\big(-c_2r^2\big)\quad \forall r>0. \tag{3.7}$$

Now we return to the string $\mathbf{x}_n$. Let u_j' be the indicator that the bit in position j of $\mathbf{x}_n$ is not deleted. Let J be the random variable describing the position of the bit copied to position $\lfloor 2np\rfloor$ of $\widetilde{\mathbf{x}}_n$, that is,

$$J := \inf\Big\{j\in\mathbb{N};\, \sum_{i=1}^{j} u_i' = \lfloor 2np\rfloor\Big\}.$$

Extend u'_j to be a Bernoulli(p) process also for $j > 4n$, so that J is a.s. well defined as a natural number. Let E' be the event that at least $\sqrt{npq}$ bits are *not* deleted among the bits in position $\{J+1, J+2, \dots, J+m\}$, that is,

$$E' := \left[\sum_{j=J+1}^{J+m} u'_j \geq \sqrt{npq} \right].$$

Then $\mathbf{P}[E'] = \mathbf{P}[E]$. If the event $E'' := [J + m < 4n]$ occurs, then $V\mathbf{1}_E$ and $Z\mathbf{1}_{E'}$ can be coupled so they differ by at most 2, for example, by taking $u_j = u'_{J+j}$ for all j. For some constant c_3, $\mathbf{P}[(E'')^{\mathrm{c}}] \leq \exp(-c_3 n)$. Combining these observations with the fact that V and Z are bounded by $\sqrt{npq}$, we obtain that for all sufficiently large n,

$$\left|\mathbf{E}[V\mathbf{1}_E] - \mathbf{E}[Z]\right| \leq \left|\mathbf{E}[V\mathbf{1}_E] - \mathbf{E}[Z\mathbf{1}_{E'\cap E''}]\right| + \sqrt{npq}\left(\mathbf{P}\left[(E')^{\mathrm{c}}\right] + \mathbf{P}\left[(E'')^{\mathrm{c}}\right]\right) \leq 3.$$

Assembling the above bounds, we obtain that for all sufficiently large n,

$$\begin{aligned}
&\mathbf{P}\left[\left|Z - \mathbf{E}[Z]\right| > rn^{1/4} + 5\right] \\
&\quad\leq \mathbf{P}\left[E' \cap E'' \cap \left[\left|Z - \mathbf{E}[Z]\right| > rn^{1/4} + 5\right]\right] + \mathbf{P}\left[(E')^{\mathrm{c}}\right] + \mathbf{P}\left[(E'')^{\mathrm{c}}\right] \\
&\quad\leq \mathbf{P}\left[\left|V\mathbf{1}_E - \mathbf{E}[V\mathbf{1}_E]\right| > rn^{1/4}\right] + \exp\left(-c_1 n^{1/2}\right) + \exp(-c_3 n) \\
&\quad\leq 2\exp\left(-c_2 r^2\right) + \exp\left(-c_1 n^{1/2}\right) + \exp(-c_3 n).
\end{aligned} \tag{3.8}$$

The first term on the right-hand side dominates for $r = o(n^{1/4})$. Since Z is bounded by $\sqrt{npq}$, the left-hand side of (3.8) is zero for $r > \sqrt{pq}n^{1/4}$. Combining these two observations yields the lemma. □

LEMMA 3.3. *Let X and Y be discrete, real-valued random variables such that*

$$\forall r > 0 \quad \mathbf{P}\left[|X| > r\right] \vee \mathbf{P}\left[|Y| > r\right] \leq 2\exp\left(-r^2\right).$$

Then

$$\left|\mathbf{E}[X] - \mathbf{E}[Y]\right| \leq 4d_{\mathrm{TV}}(X, Y)\sqrt{\log \frac{2}{d_{\mathrm{TV}}(X, Y)}}.$$

PROOF. We let $\delta := d_{\mathrm{TV}}(X, Y)$ to simplify notation. Letting μ_X and μ_Y denote the law of X and Y, respectively, write

$$\mu_X = \mu + \mu_X^- + \mu_X^+ \quad \text{and} \quad \mu_Y = \mu + \mu_Y^- + \mu_Y^+,$$

where $\mu(\mathbb{R}) = 1 - \delta$, μ_X^- and μ_Y^- are supported on $(-\infty, 0)$, and μ_X^+ and μ_Y^+ are supported on $[0, \infty)$. Then

$$\begin{aligned}
\left|\mathbf{E}[X] - \mathbf{E}[Y]\right| \leq{}& \left|\sum x\mu_X^-(x)\right| + \left|\sum x\mu_X^+(x)\right| + \left|\sum x\mu_Y^-(x)\right| \\
&+ \left|\sum x\mu_Y^+(x)\right|.
\end{aligned} \tag{3.9}$$

We have

$$\begin{aligned}
&\left|\sum x\mu_X^-(x)\right| + \left|\sum x\mu_X^+(x)\right| \\
&\quad= \int_{\mathbb{R}_+} \mu_X^-\left((-\infty, -r]\right) + \mu_X^+\left([r, \infty)\right) dr \\
&\quad\leq \int_{\mathbb{R}_+} \min\left\{2e^{-r^2}, \delta\right\} dr \leq \delta\sqrt{\log \frac{2}{\delta}} + \frac{\delta}{2\sqrt{\log \frac{2}{\delta}}} \leq 2\delta\sqrt{\log \frac{2}{\delta}}.
\end{aligned}$$

Inserting this estimate and analogous estimates for μ_Y^- and μ_Y^+ into (3.9), we obtain the lemma,

$$\bigl|\mathbf{E}[X]-\mathbf{E}[Y]\bigr| \le 4\delta\sqrt{\log\frac{2}{\delta}}. \qquad \square$$

PROOF OF PROPOSITION 1.4, LOWER BOUND. By (A.6),

$$d_{\mathrm{TV}}\bigl(Z(\tilde{\mathbf{x}}_n), Z(\tilde{\mathbf{y}}_n)\bigr) \le d_{\mathrm{TV}}(\Delta_{\mathbf{x}_n}, \Delta_{\mathbf{y}_n}). \tag{3.10}$$

Letting c denote the constant in Lemma 3.2 and

$$a := \frac{1}{2}\bigl(\mathbf{E}\bigl[Z(\tilde{\mathbf{x}}_n)\bigr] + \mathbf{E}\bigl[Z(\tilde{\mathbf{y}}_n)\bigr]\bigr) = \Theta\bigl(n^{1/2}\bigr),$$

define

$$X := \frac{(Z(\tilde{\mathbf{x}}_n)-a)c}{2n^{1/4}} \quad \text{and} \quad Y := \frac{(Z(\tilde{\mathbf{y}}_n)-a)c}{2n^{1/4}}.$$

Notice that

$$d_{\mathrm{TV}}(X,Y) = d_{\mathrm{TV}}\bigl(Z(\tilde{\mathbf{x}}_n), Z(\tilde{\mathbf{y}}_n)\bigr). \tag{3.11}$$

By Lemma 3.1,

$$\bigl|\mathbf{E}\bigl[Z(\tilde{\mathbf{x}}_n)\bigr]-a\bigr| \vee \bigl|\mathbf{E}\bigl[Z(\tilde{\mathbf{y}}_n)\bigr]-a\bigr| = \Theta\bigl(n^{-1/2}\bigr).$$

Combining this with Lemma 3.2, we see that X and Y satisfy the condition of Lemma 3.3 for all sufficiently large n. The proposition now follows from Lemma 3.3, (3.10), (3.11) and Lemma 3.1. $\square$

4. Lower bound for random strings: Proof of Proposition 1.5. Recall the reconstruction problem for random strings described in Section 1.2. Proposition 4.1 below transfers lower bounds for deterministic strings to lower bounds for random strings, yielding almost exponentially small success probability. Proposition 4.1 is proved by adapting the method of [13], Theorem 1.

Proposition 1.5 follows from Theorem 1.1 and Proposition 4.1 applied with the function $f(n) = \lfloor cn^{5/4}/\sqrt{\log n}\rfloor$. The lower bound of $\Omega(\log^2 n)$ from [13], Theorem 1, may be obtained from the proposition with $f(n) = \lfloor cn\rfloor$.

In order to state the proposition, we need to describe the trace reconstruction problem with *random* G. We say that all n-bit strings can be *reconstructed with probability at least* $1-\varepsilon$ *from* T *traces with additional randomness* if there is a Borel function $G'\colon \mathcal{S}^T \times [0,1] \to \{0,1\}^n$ such that for all $\mathbf{x} \in \mathcal{S}_n$,

$$\int_0^1 \mathbf{P}_{\mathbf{x}}\bigl[G'(\mathfrak{X},t) = \mathbf{x}\bigr]\,dt \ge 1-\varepsilon. \tag{4.1}$$

For the purpose of distinguishing between two input strings, reconstruction with extra randomness is equivalent to reconstruction without extra randomness, at least if we are willing to change ε by a factor of 2: Let $\mathbf{x} \ne \mathbf{y}$. As noted in Appendix A.2,

$$\min_G\bigl(\mathbf{P}_{\mathbf{x}}\bigl[G(\mathfrak{X}) = \mathbf{y}\bigr] + \mathbf{P}_{\mathbf{y}}\bigl[G(\mathfrak{X}) = \mathbf{x}\bigr]\bigr) = 1 - d_{\mathrm{TV}}(\Delta_{\mathbf{x}}, \Delta_{\mathbf{y}}).$$

Therefore, for any $G'\colon \mathcal{S}^T \times [0,1] \to \{\mathbf{x},\mathbf{y}\}$,

$$\int_0^1 \bigl(\mathbf{P}_{\mathbf{x}}\bigl[G'(\mathfrak{X},t) = \mathbf{y}\bigr] + \mathbf{P}_{\mathbf{y}}\bigl[G'(\mathfrak{X},t) = \mathbf{x}\bigr]\bigr)\,dt \ge 1 - d_{\mathrm{TV}}(\Delta_{\mathbf{x}}, \Delta_{\mathbf{y}}).$$

Since the maximum error probability is at most this sum of error probabilities and also is at least half the same sum, our claim follows. In particular, the lower bound $\Omega(\log(1/\varepsilon)n^{5/4}/\sqrt{\log n})$ in Theorem 1.1 also holds if we consider reconstruction with extra randomness. A similar definition holds for reconstructing random strings with extra randomness when the random string is chosen according to a probability measure, ρ: one simply takes the expectation of the left-hand side of (4.1) with $\mathbf{x} \sim \rho$.

PROPOSITION 4.1. *Suppose that for all $n \in \mathbb{N}$, the probability that all n-bit strings can be reconstructed with $f(n) \cdot n$ traces is at most $1 - e^{-n}$, even with extra randomness. Then for all large $n \in \mathbb{N}$, the probability of reconstructing random n-bit strings with $\lfloor \frac{1}{2} f(\frac{1}{2}\log n) \cdot \log n \rfloor$ traces is at most $\exp(-n^{0.15})$, even with extra randomness.*

PROOF. Let $r := \lfloor \frac{1}{2}\log n \rfloor$ and $T := f(r)r$. It was observed by Yao [18] that von Neumann's minimax theorem yields

$$\min_{G'} \max_{\mathbf{x}\in\mathcal{S}_r} \int_0^1 \mathbf{P}_{\mathbf{x}}\big[G'(\mathfrak{X}, t) \neq \mathbf{x}\big]\, dt = \max_{\rho} \min_{G} \sum_{\mathbf{x}\in\mathcal{S}_r} \mathbf{P}_{\mathbf{x}}\big[G(\mathfrak{X}) \neq \mathbf{x}\big] \cdot \rho(\mathbf{x}),$$

where we take the minima over functions $G' \colon \mathcal{S}^T \times [0,1] \to \{0,1\}^r$ and $G \colon \mathcal{S}^T \to \{0,1\}^r$, and the second maximum is over probability measures ρ on $\mathcal{S}_r$. By assumption, the left-hand side is at least equal to e^{-r}. Therefore, there is some probability measure ρ on r-bit strings such that $\sum_{\mathbf{x}\in\mathcal{S}_r} \mathbf{P}_{\mathbf{x}}[G(\mathfrak{X}) \neq \mathbf{x}] \cdot \rho(\mathbf{x}) \geq e^{-r}$ for all G, that is, the probability of reconstructing an r-bit string chosen according to ρ with T traces is at most $1 - e^{-r}$. Furthermore, this result for r-bit strings sampled from ρ holds also for reconstruction with additional randomness, since for any $\widehat{G} \colon \mathcal{S}^T \times [0,1] \to \{0,1\}^r$ and $t \in [0,1]$,

$$\min_{G} \sum_{\mathbf{x}\in\mathcal{S}_r} \mathbf{P}_{\mathbf{x}}\big[G(\mathfrak{X}) \neq \mathbf{x}\big] \cdot \rho(\mathbf{x}) \leq \sum_{\mathbf{x}\in\mathcal{S}_r} \mathbf{P}_{\mathbf{x}}\big[\widehat{G}(\mathfrak{X}, t) \neq \mathbf{x}\big] \cdot \rho(\mathbf{x}),$$

which implies

$$\min_{G} \sum_{\mathbf{x}\in\mathcal{S}_r} \mathbf{P}_{\mathbf{x}}\big[G(\mathfrak{X}) \neq \mathbf{x}\big] \cdot \rho(\mathbf{x}) \leq \min_{\widehat{G}} \sum_{\mathbf{x}\in\mathcal{S}_r} \int_0^1 \mathbf{P}_{\mathbf{x}}\big[\widehat{G}(\mathfrak{X}, t) \neq \mathbf{x}\big]\, dt \cdot \rho(\mathbf{x}).$$

Sample the random uniform string $\mathbf{x} \in \mathcal{S}_n$ in the following manner. Denote

$$\mathbf{z}_j := (x_{(j-1)r+1}, x_{(j-1)r+2}, \ldots, x_{jr}) \quad \text{for } 1 \leq j \leq n/r$$

and $\mathbf{w} := (x_{\lfloor n/r \rfloor r+1}, x_{\lfloor n/r \rfloor r+2}, \ldots, x_n)$. Write λ for the uniform distribution on strings of length r and define $\sigma := (\lambda - 2^{-r}\rho)/(1 - 2^{-r})$, which is a probability measure. Let $(Q_j)_{j\geq 1}$ be a Bernoulli(2^{-r}) process. For each j, choose $\mathbf{z}_j$ from σ if $Q_j = 0$ and from ρ if $Q_j = 1$, independently for different j. Let $\mathbf{w}$ be uniform (independent of the preceding). Let $\mathfrak{X}$ be the T traces obtained from $\mathbf{x}$; it is the trace-wise concatenation of the traces $\mathfrak{Z}_j \in \mathcal{S}^T$ obtained from $\mathbf{z}_j$ and $\mathfrak{W} \in \mathcal{S}^T$ obtained from $\mathbf{w}$. The probability of reconstructing $\mathbf{x}$ from $\mathfrak{X}$ is at most the probability of reconstructing $\mathbf{x}$ from[2] $\mathfrak{Z}_1, \ldots, \mathfrak{Z}_{\lfloor n/r \rfloor r}, \mathfrak{W}$ (because we could simply ignore the additional information in the separate traces $\mathfrak{Z}_i$ and $\mathfrak{W}$ that is not inherent in $\mathfrak{X}$). Conditional on $Q_j = 1$, the probability of reconstructing $\mathbf{z}_j$ with T traces is at most $1 - e^{-r}$ by assumption. Therefore, the unconditional probability of reconstructing $\mathbf{z}_j$ from $\mathfrak{Z}_j$ is at most $1 - 2^{-r}e^{-r}$. Since these events are independent in j, we obtain that the probability of reconstructing $\mathbf{x}$ from $\mathfrak{X}$ is at most $(1 - 2^{-r}e^{-r})^{\lfloor n/r \rfloor} \leq \exp(-0.9 \cdot 2^{-r}e^{-r}n/r)$ for $n/r \geq 10$. Inserting the definition of r gives the result. □

[2]We did not define this reconstruction problem, but its meaning should be obvious.

APPENDIX: INEQUALITIES FOR DISTANCES BETWEEN MEASURES

Throughout this appendix, μ and ν are positive measures on a countable set, X. Most of the material in this Appendix is standard and elementary, although we have not found a good reference presenting all the material needed for the body of our paper. One possible novelty, however, is Lemma A.1, which we have not seen elsewhere. This lemma, though completely elementary, is a key input to the proof of Theorem 1.1, and we believe it is also useful to bound the squared Hellinger distance between two measures in many other contexts.

A.1. Inequalities for Hellinger distance and total variation distance. The *total variation* distance between μ and ν is defined by

$$d_{\mathrm{TV}}(\mu,\nu) := \frac{1}{2}\sum_{x\in X}\bigl|\mu(x)-\nu(x)\bigr|.$$

Thus, in order to maximize $\sigma(X)$ over all decompositions $\mu=\sigma+\mu'$, $\nu=\sigma+\nu'$, where σ, μ' and ν' are positive measures on X, one takes $\sigma := \mu\wedge\nu$, yielding $\mu'(X)+\nu'(X) = 2d_{\mathrm{TV}}(\mu,\nu)$. If μ and ν are probability measures, then

$$d_{\mathrm{TV}}(\mu,\nu) = \max_{A\subseteq X}\bigl[\mu(A)-\nu(A)\bigr]. \tag{A.1}$$

The *Hellinger* distance between μ and ν is defined by[3]

$$d_{\mathrm{H}}(\mu,\nu) := \Biggl(\sum_{x\in X}\bigl[\sqrt{\mu(x)}-\sqrt{\nu(x)}\bigr]^2\Biggr)^{1/2}.$$

It is well known (e.g., [17], Lemma 2.3) that for probability measures μ and ν, we have

$$d_{\mathrm{TV}}(\mu,\nu) \le d_{\mathrm{H}}(\mu,\nu) \le \sqrt{2d_{\mathrm{TV}}(\mu,\nu)}. \tag{A.2}$$

The next lemma shows that the right-hand inequality can be strengthened if for all $x\in X$, the ratio $\mu(x)/\nu(x)$ is close to 1. Before stating it, we introduce the notation $\|f\|_{\ell^\infty(\nu)}$ for a function $f\colon X\to\mathbb{R}$,

$$\|f\|_{\ell^\infty(\nu)} := \sup\bigl\{\bigl|f(x)\bigr|;\, x\in X,\, \nu(x)\neq 0\bigr\}. \tag{A.3}$$

LEMMA A.1. *For all positive measures μ and ν, we have*

$$d_{\mathrm{H}}^2(\mu,\nu) \le \mu\bigl\{x;\,\nu(x)=0\bigr\} + 2\cdot\biggl\|\frac{\mu-\nu}{\nu}\biggr\|_{\ell^\infty(\nu)}\cdot d_{\mathrm{TV}}(\mu,\nu). \tag{A.4}$$

PROOF. Since $|a-1|\le|a^2-1|$ for all $a\ge 0$, we have

$$\begin{aligned} &d_{\mathrm{H}}^2(\mu,\nu) - \mu\bigl\{x;\,\nu(x)=0\bigr\} \\ &\quad= \sum_{x\in X;\,\nu(x)\neq 0}\biggl(\sqrt{\frac{\mu(x)}{\nu(x)}}-1\biggr)^2\nu(x) \le \sum_{x\in X;\,\nu(x)\neq 0}\biggl(\frac{\mu(x)}{\nu(x)}-1\biggr)^2\nu(x) \\ &\quad\le \biggl\|\frac{\mu}{\nu}-1\biggr\|_{\ell^\infty(\nu)}\cdot\biggl\|\frac{\mu}{\nu}-1\biggr\|_{\ell^1(\nu)}, \end{aligned}$$

which is equation (A.4). □

One way to bound this ℓ^∞-norm is to use the following observation.

[3]Some authors use another normalization, for example, with a factor of $1/\sqrt{2}$ on the right-hand side.

LEMMA A.2. *Let ρ and σ be positive measures on a countable space Y, let λ be a probability measure on a measurable space Z, and let $\phi\colon Y \times Z \to X$ be a function. Defining $\mu := \phi_*(\rho \times \lambda)$ and $\nu := \phi_*(\sigma \times \lambda)$ to be the push-forward measures, we have*

$$\left\| \frac{\mu}{\nu} - 1 \right\|_{\ell^\infty(\nu)} \le \left\| \frac{\rho}{\sigma} - 1 \right\|_{\ell^\infty(\sigma)}. \tag{A.5}$$

PROOF. If $\nu(x) > 0$, then there must exist $y \in Y$ and $z \in Z$ such that $x = \phi(y, z)$ and $\sigma(y) > 0$. Therefore, the following holds for any $x \in X$ for which $\nu(x) > 0$, with δ denoting the right-hand side of (A.5) and $U \sim \lambda$:

$$\begin{aligned} \bigl|\mu(x) - \nu(x)\bigr| &= \Bigl| \sum_{y \in Y; \sigma(y) > 0} \bigl(\rho(y) - \sigma(y)\bigr) \mathbf{P}\bigl[\phi(y, U) = x\bigr] \Bigr| \\ &\le \delta \sum_{y \in Y; \sigma(y) > 0} \sigma(y) \mathbf{P}\bigl[\phi(y, U) = x\bigr] = \delta \cdot \nu(x). \end{aligned}$$

□

By equation (A.1), pushing forward two probability[4] measures by the same map cannot increase the total variation distance:

$$d_{\mathrm{TV}}(\phi_*\rho, \phi_*\sigma) \le d_{\mathrm{TV}}(\rho, \sigma). \tag{A.6}$$

The following is immediate from the definition:

$$d_{\mathrm{H}}^2(\mu, \nu) \le \mu(X) + \nu(X). \tag{A.7}$$

LEMMA A.3. *For any positive measures μ_1, μ_2, ν_1 and ν_2 on X, we have*

$$d_{\mathrm{H}}^2(\mu_1 + \mu_2, \nu_1 + \nu_2) \le d_{\mathrm{H}}^2(\mu_1, \nu_1) + d_{\mathrm{H}}^2(\mu_2, \nu_2).$$

PROOF. This is immediate from the inequality

$$(\sqrt{a+b} - \sqrt{c+d})^2 \le (\sqrt{a} - \sqrt{c})^2 + (\sqrt{b} - \sqrt{d})^2, \quad a, b, c, d \ge 0.$$

□

The following is well known (see, e.g., [16], p. 100).

LEMMA A.4. *For any probability measures μ_1, μ_2, ν_1, ν_2 on X, we have*

$$d_{\mathrm{H}}^2(\mu_1 \times \mu_2, \nu_1 \times \nu_2) \le d_{\mathrm{H}}^2(\mu_1, \nu_1) + d_{\mathrm{H}}^2(\mu_2, \nu_2).$$

A.2. Distinguishing between measures by independent sampling. In this section, we consider two distinct probability measures μ and ν, and for $m \in \mathbb{N}$, we consider m independent samples from one of the measures. We are interested in how large we need to choose m in order to determine whether our samples are from μ or ν. Our bounds are expressed in terms of the Hellinger distance and the total variation distance between the measures.

Consider first the case where $m = 1$. Let $G\colon X \to \{\mu, \nu\}$ be a function that (roughly speaking) says whether some element $x \in X$ is more likely to be sampled from μ or ν. We are interested in the sum of the error probabilities $\mu[G(x) = \nu] + \nu[G(x) = \mu]$. By equation (A.1), the error probability sum is minimized by taking

$$G(x) := \begin{cases} \mu & \text{if } \mu(x) \ge \nu(x), \\ \nu & \text{otherwise,} \end{cases}$$

in which case we get that the error probability sum equals $1 - d_{\mathrm{TV}}(\mu, \nu)$.

[4] We remark that equation (A.6) also holds when ρ and σ are not probability measures.

Replacing μ, ν by μ^m, ν^m in this discussion, we get that for general m, the number of samples required to distinguish between μ and ν is determined precisely by $d_{\mathrm{TV}}(\mu^m, \nu^m)$.

Now we derive a lower bound for the number of required samples, expressed in terms of $d_{\mathrm{TV}}(\mu, \nu)$. It is well known that total variation distance can be expressed via coupling:

$$d_{\mathrm{TV}}(\mu, \nu) = \min\{\mathbf{P}[U \neq V]; U \sim \mu, V \sim \nu\},$$

where the minimum is taken over all couplings of U and V. By using couplings of the pairs (μ_i, ν_i) that are independent in i, it follows that for probability measures $\mu_1, \ldots, \mu_n, \nu_1, \ldots, \nu_n$,

$$1 - d_{\mathrm{TV}}(\mu_1 \times \cdots \times \mu_n, \nu_1 \times \cdots \times \nu_n) \geq \prod_{i=1}^{n} [1 - d_{\mathrm{TV}}(\mu_i, \nu_i)]. \tag{A.8}$$

In particular,

$$1 - d_{\mathrm{TV}}(\mu^m, \nu^m) \geq e^{-\alpha(\mu,\nu)\cdot m \cdot d_{\mathrm{TV}}(\mu,\nu)}, \tag{A.9}$$

where

$$\alpha(\mu, \nu) := -\frac{\log[1 - d_{\mathrm{TV}}(\mu, \nu)]}{d_{\mathrm{TV}}(\mu, \nu)}.$$

Note that $\alpha(\mu, \nu)$ approaches 1 as $d_{\mathrm{TV}}(\mu, \nu) \to 0$, and that, for example, $\alpha(\mu, \nu)$ is at most 3/2 when $d_{\mathrm{TV}}(\mu, \nu) \leq 1/2$. We can interpret equation (A.9) as saying that in order to distinguish μ from ν when given m i.i.d. samples from an unknown choice from $\{\mu, \nu\}$, we need at least

$$m = \Omega\big(1/d_{\mathrm{TV}}(\mu, \nu)\big) \tag{A.10}$$

samples. Alternatively, we can say that if r samples yield an error probability at least $1/e$, then $r\lfloor \log(1/\varepsilon) \rfloor$ samples yield an error probability at least ε.

Next, we derive an upper bound for the number of required samples, also expressed in terms of $d_{\mathrm{TV}}(\mu, \nu)$. Namely, we will prove the well-known result that we need at most $m = O(1/d^2_{\mathrm{TV}}(\mu, \nu))$ samples. By equation (A.1), we can find an event $A \subset X$ such that $\mu(A) - \nu(A) = d_{\mathrm{TV}}(\mu, \nu)$. Given m independent samples $\mathfrak{X}_m = (x_1, \ldots, x_m)$ from one of the measures, let $\overline{u} := m^{-1} \sum_{j=1}^{m} \mathbf{1}_{[x_j \in A]}$ be the fraction of times that A occurs. Define

$$G(\mathfrak{X}_m) := \begin{cases} \mu & \text{if } \overline{u} > \nu(A) + \dfrac{1}{2} d_{\mathrm{TV}}(\mu, \nu) = \dfrac{1}{2}\big(\mu(A) + \nu(A)\big), \\ \nu & \text{otherwise.} \end{cases}$$

An application of the inequality of Hoeffding–Azuma gives the following bound for the sum of the error probabilities:

$$\mu\big[G(\mathfrak{X}_m) = \nu\big] + \nu\big[G(\mathfrak{X}_m) = \mu\big] \leq 2\exp\big(-m \cdot d^2_{\mathrm{TV}}(\mu, \nu)/2\big).$$

In particular,

$$m \geq \frac{2}{d^2_{\mathrm{TV}}(\mu, \nu)} \log \frac{2}{\varepsilon} \tag{A.11}$$

samples are sufficient to distinguish between the measures with error probability at most ε.

Both the lower and upper bounds for the number of samples required in terms of total variation distance are sharp, as illustrated by the following examples, where we use Bernoulli(s) to denote the law of a Bernoulli random variable with parameter $s \in [0, 1]$: (1) $\mu :=$ Bernoulli(0) and $\nu :=$ Bernoulli(δ), where $d_{\mathrm{TV}}(\mu, \nu) = \delta$ and $\Theta(\delta^{-1})$ samples are necessary and sufficient, and (2) $\mu :=$ Bernoulli(1/2) and $\nu :=$ Bernoulli($1/2 + \delta$),

where $d_{\mathrm{TV}}(\mu,\nu)=\delta$ and $\Theta(\delta^{-2})$ samples are necessary and sufficient. More generally, for $\alpha\in[0,1]$, if $\mu:=\mathrm{Bernoulli}(\delta^{1-\alpha}/2)$ and $\nu:=\mathrm{Bernoulli}(\delta^{1-\alpha}/2+\delta)$, then $d_{\mathrm{TV}}(\mu,\nu)=\delta$ and $\Theta(\delta^{-1-\alpha})$ samples are necessary and sufficient.

If $d_{\mathrm{H}}^2(\mu,\nu)$ is much smaller than $d_{\mathrm{TV}}(\mu,\nu)$, then the lower bound (A.10) can be improved. The following type of result seems to be folklore; we saw a version of it in [13], Corollary 1. It says that $m=\Omega(1/d_{\mathrm{H}}^2(\mu,\nu))$ samples are necessary to distinguish between μ and ν.

LEMMA A.5. *If μ and ν are probability measures with $d_{\mathrm{H}}(\mu,\nu)\le 1/2$, then for $m\ge 1/(4d_{\mathrm{H}}^2(\mu,\nu))$,*

$$1-d_{\mathrm{TV}}(\mu^m,\nu^m)\ge\exp\{-9m\cdot d_{\mathrm{H}}^2(\mu,\nu)\}.$$

In particular, $1-d_{\mathrm{TV}}(\mu^m,\nu^m)\ge\varepsilon$ if

$$m\le\frac{1}{9d_{\mathrm{H}}^2(\mu,\nu)}\log\frac{1}{\varepsilon}.$$

PROOF. Define $r:=\lfloor 1/(4d_{\mathrm{H}}^2(\mu,\nu))\rfloor\ge 1$. By equation (A.2) and Lemma A.4,

$$d_{\mathrm{TV}}^2(\mu^r,\nu^r)\le d_{\mathrm{H}}^2(\mu^r,\nu^r)\le rd_{\mathrm{H}}^2(\mu,\nu)\le 1/4.$$

This allows us to apply equation (A.9) as follows:

$$\begin{aligned}1-d_{\mathrm{TV}}(\mu^m,\nu^m)&\ge\exp\left\{-\frac{3}{2}\cdot\left\lceil\frac{m}{r}\right\rceil d_{\mathrm{TV}}(\mu^r,\nu^r)\right\}\ge\exp\left\{-3\cdot\frac{m}{r}d_{\mathrm{TV}}(\mu^r,\nu^r)\right\}\\&\ge\exp\left\{-3\cdot\frac{m}{r}\sqrt{r}d_{\mathrm{H}}(\mu,\nu)\right\}\ge\exp\{-3\sqrt{8}\cdot m\cdot d_{\mathrm{H}}^2(\mu,\nu)\},\end{aligned}$$

where in the last step, we used $r\ge 1/(8d_{\mathrm{H}}^2(\mu,\nu))$. □

Acknowledgments. We thank the anonymous referees for useful comments and careful reading of our paper. Most of this paper was written while visiting Microsoft Research Redmond, and we thank Microsoft for the hospitality.

The first author was supported in part by an internship at Microsoft Research and by a fellowship from the Norwegian Research Council.

The second author was supported in part by NSF Grant DMS-1612363.

REFERENCES

[1] BATU, T., KANNAN, S., KHANNA, S. and MCGREGOR, A. (2004). Reconstructing strings from random traces. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 910–918. ACM, New York. MR2290981

[2] BENJAMINI, I. and KESTEN, H. (1996). Distinguishing sceneries by observing the scenery along a random walk path. *J. Anal. Math.* **69** 97–135. MR1428097 https://doi.org/10.1007/BF02787104

[3] CHASE, Z. (2019). New lower bounds for trace reconstruction. Available at arXiv:1905.03031.

[4] COMBES, R. (2015). An extension of McDiarmid's inequality. Available at arXiv:1511.05240.

[5] DE, A., O'DONNELL, R. and SERVEDIO, R. A. (2017). Optimal mean-based algorithms for trace reconstruction. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* 1047–1056. ACM, New York. MR3678250

[6] HART, A., MACHADO, F. and MATZINGER, H. (2015). Information recovery from observations by a random walk having jump distribution with exponential tails. *Markov Process. Related Fields* **21** 939–970. MR3496231

[7] HOLDEN, N., PEMANTLE, R. and PERES, Y. (2018). Subpolynomial trace reconstruction for random strings and arbitrary deletion probability. In *Proceedings of the 31st Conference on Learning Theory* (S. S. Bubeck, V. V. Perchet and P. P. Rigollet, eds.). *Proceedings of Machine Learning Research* **75** 1799–1840.

[8] HOLENSTEIN, T., MITZENMACHER, M., PANIGRAHY, R. and WIEDER, U. (2008). Trace reconstruction with constant deletion probability and related results. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 389–398. ACM, New York. MR2487606

[9] LIGGETT, T. M. (2002). Tagged particle distributions or how to choose a head at random. In *In and Out of Equilibrium* (*Mambucaba*, 2000). *Progress in Probability* **51** 133–162. Birkhäuser, Boston, MA. MR1901951

[10] MATZINGER, H. and PINZON, A. P. (2011). DNA approach to scenery reconstruction. *Stochastic Process. Appl.* **121** 2455–2473. MR2832409 https://doi.org/10.1016/j.spa.2011.04.010

[11] MATZINGER, H. and ROLLES, S. W. W. (2003). Reconstructing a piece of scenery with polynomially many observations. *Stochastic Process. Appl.* **107** 289–300. MR1999792 https://doi.org/10.1016/S0304-4149(03)00085-1

[12] MATZINGER, H. and ROLLES, S. W. W. (2006). Finding blocks and other patterns in a random coloring of $\mathbb{Z}$. *Random Structures Algorithms* **28** 37–75. MR2187482 https://doi.org/10.1002/rsa.20110

[13] MCGREGOR, A., PRICE, E. and VOROTNIKOVA, S. (2014). Trace reconstruction revisited. In *Algorithms—ESA* 2014. *Lecture Notes in Computer Science* **8737** 689–700. Springer, Heidelberg. MR3253172 https://doi.org/10.1007/978-3-662-44777-2_57

[14] NAZAROV, F. and PERES, Y. (2017). Trace reconstruction with $\exp(O(n^{1/3}))$ samples. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* 1042–1046. ACM, New York. MR3678249

[15] PERES, Y. and ZHAI, A. (2017). Average-case reconstruction for the deletion channel: Subpolynomially many traces suffice. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 228–239. IEEE Computer Soc., Los Alamitos, CA. MR3734232

[16] REISS, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics. Springer Series in Statistics*. Springer, New York. MR0988164 https://doi.org/10.1007/978-1-4613-9620-8

[17] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[18] YAO, A. C. C. (1977). Probabilistic computations: Toward a unified measure of complexity (extended abstract). In *18th Annual Symposium on Foundations of Computer Science* (*Providence, RI*, 1977) 222–227. IEEE Comput. Sci., Long Beach, CA. MR0489016