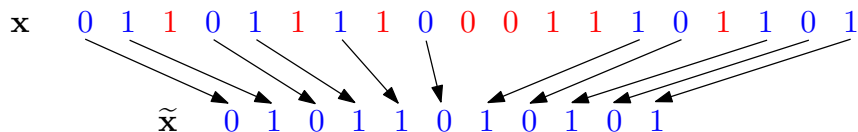


Lower Bounds for Trace Reconstruction

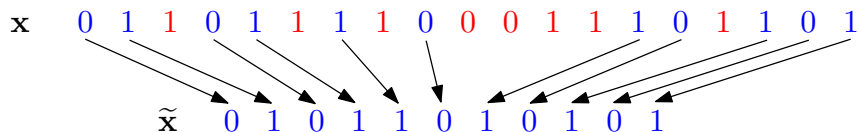
Nina Holden and Russell Lyons

ETH zürich



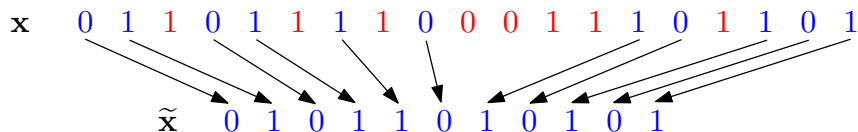


We obtain the **trace** \tilde{x} by deleting (red) or copying (blue) each bit of the finite input binary string x .



We obtain the **trace** $\tilde{\mathbf{x}}$ by deleting (red) or copying (blue) each bit of the finite input binary string \mathbf{x} .

We delete with probability q and copy with probability $p = 1 - q$.



We obtain the **trace** $\tilde{\mathbf{x}}$ by deleting (red) or copying (blue) each bit of the finite input binary string \mathbf{x} .

We delete with probability q and copy with probability $p = 1 - q$.

We get T independent traces $\mathfrak{X} = \{\tilde{\mathbf{x}}^{(1)}, \dots, \tilde{\mathbf{x}}^{(T)}\}$ from the deletion channel. How many traces are needed to reconstruct the input \mathbf{x} with high probability?

Introduced by Batu, Kannan, Khanna, and McGregor [BKKM04] 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.



001001	0111001	001010100	0010
00010101	0010000	000110	1110101100
110	0100101001	01010101101	10110100
01011101	1001101011	10110101	0110010111
0100101010	100101	0101101001	01110010

0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1

001001	0111001	001010100	0010
00010101	0010000	000110	1110101100
110	0100101001	01010101101	10110100
01011101	1001101011	10110101	0110010111
0100101010	100101	0101101001	01110010



	1011	10101101	10010	010011011
	010011	0010	1101001	0110100111
11010001010	01100100110	0011001011	0011	
011101	011000011	01000101	010010101	
0101001	010100110	0010010	01011011	

0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0

1011	10101101	10010	010011011
010011	0010	1101001	0110100111
11010001010	01100100110	0011001011	0011
011101	011000011	01000101	010010101
0101001	010100110	0010010	01011011

Upper bounds: Trivial: $e^{O(n)}$. Holenstein, Mitzenmacher, Panigrahy, and Wieder (2008) proved that every string $\mathbf{x} \in \{0, 1\}^n$ can be reconstructed from $\exp(n^{1/2} \text{polylog}(n))$ traces.

Improved by De, O'Donnell, and Servedio (2016) and Nazarov and Peres (2016) to $\exp(O(n^{1/3}))$ traces.

Upper bounds: Trivial: $e^{O(n)}$. Holenstein, Mitzenmacher, Panigrahy, and Wieder (2008) proved that every string $\mathbf{x} \in \{0, 1\}^n$ can be reconstructed from $\exp(n^{1/2} \text{polylog}(n))$ traces.

Improved by De, O'Donnell, and Servedio (2016) and Nazarov and Peres (2016) to $\exp(O(n^{1/3}))$ traces. (Best possible using single-bit statistics.)

Upper bounds: Trivial: $e^{O(n)}$. Holenstein, Mitzenmacher, Panigrahy, and Wieder (2008) proved that every string $\mathbf{x} \in \{0, 1\}^n$ can be reconstructed from $\exp(n^{1/2} \text{polylog}(n))$ traces.

Improved by De, O'Donnell, and Servedio (2016) and Nazarov and Peres (2016) to $\exp(O(n^{1/3}))$ traces. (Best possible using single-bit statistics.)

Lower bounds: Trivial: $\Omega(\log n)$ (must see each bit at least once). [BKKM04] proved that the number of traces needed for reconstruction is $\Omega(n)$. For example,

Upper bounds: Trivial: $e^{O(n)}$. Holenstein, Mitzenmacher, Panigrahy, and Wieder (2008) proved that every string $\mathbf{x} \in \{0, 1\}^n$ can be reconstructed from $\exp(n^{1/2} \text{polylog}(n))$ traces.

Improved by De, O'Donnell, and Servedio (2016) and Nazarov and Peres (2016) to $\exp(O(n^{1/3}))$ traces. (Best possible using single-bit statistics.)

Lower bounds: Trivial: $\Omega(\log n)$ (must see each bit at least once). [BKKM04] proved that the number of traces needed for reconstruction is $\Omega(n)$. For example, the pair of strings $\mathbf{x}'_n = (0)^{n-1}1(0)^n$ and $\mathbf{y}'_n = (0)^n1(0)^{n-1}$ (where $(b)^m$ means a string of m consecutive bs) requires $\Omega(n)$ traces to be distinguished.

Upper bounds: Trivial: $e^{O(n)}$. Holenstein, Mitzenmacher, Panigrahy, and Wieder (2008) proved that every string $\mathbf{x} \in \{0, 1\}^n$ can be reconstructed from $\exp(n^{1/2} \text{polylog}(n))$ traces.

Improved by De, O'Donnell, and Servedio (2016) and Nazarov and Peres (2016) to $\exp(O(n^{1/3}))$ traces. (Best possible using single-bit statistics.)

Lower bounds: Trivial: $\Omega(\log n)$ (must see each bit at least once). [BKKM04] proved that the number of traces needed for reconstruction is $\Omega(n)$. For example, the pair of strings $\mathbf{x}'_n = (0)^{n-1}1(0)^n$ and $\mathbf{y}'_n = (0)^n1(0)^{n-1}$ (where $(b)^m$ means a string of m consecutive bs) requires $\Omega(n)$ traces to be distinguished.

An enormous gap!

We improve this lower bound. Define the strings $\mathbf{x}_n, \mathbf{y}_n \in \{0, 1\}^{4n}$ to be

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

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.

We improve this lower bound. Define the strings $\mathbf{x}_n, \mathbf{y}_n \in \{0, 1\}^{4n}$ to be

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

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 (Holden-L., 2019)

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 $1 - \varepsilon$.

x_n versus y_n ,

$x_n := (01)^{n-1}10(01)^n$ and $y_n := (01)^n10(01)^{n-1}$



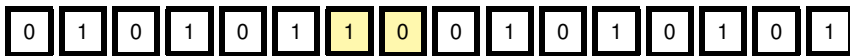
011101001	101101010	01010000	111010
100000	1001001001	0101001000	0101110
010001010	100110101010	0101010101	11000
00101001	0110010110	11101010	1110101
10011011	01001001	0111110101	010101



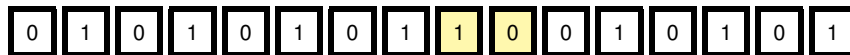
0111101	010101011	10100	1111000
01101001	011110011	01010100100	01110
010110100	0110101010	010111001	01101011
111010010	101001101	0111111	0101111
1000110001	11000	0001000	1110100

x_n versus y_n ,

$x_n := (01)^{n-1}10(01)^n$ and $y_n := (01)^n10(01)^{n-1}$



011101001	101101010	01010000	111010
100000	1001001001	0101001000	0101110
010001010	100110101010	0101010101	11000
00101001	0110010110	11101010	1110101
10011011	01001001	0111110101	010101



0111101	010101011	10100	1111000
01101001	011110011	01010100100	01110
010110100	0110101010	010111001	01101011
111010010	101001101	0111111	0101111
1000110001	11000	0001000	1110100

Proposition (Holden-L., 2019)

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

[▶ Proof](#)

The number of traces required to reconstruct an arbitrary $\mathbf{x} \in \{0, 1\}^n$, 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.

The number of traces required to reconstruct an arbitrary $\mathbf{x} \in \{0, 1\}^n$, 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 $\{0, 1\}^n$, we assume the input string \mathbf{x} is sampled from μ_n . The question now is: what T ensures that the probability is large 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.

The number of traces required to reconstruct an arbitrary $\mathbf{x} \in \{0, 1\}^n$, 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 $\{0, 1\}^n$, we assume the input string \mathbf{x} is sampled from μ_n . The question now is: what T ensures that the probability is large 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 $A_n \subset \{0, 1\}^n$ is a set of large μ_n -measure and $\{0, 1\}^n \setminus A_n$ 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 Vortnikova [MPV14] proved that $\Omega(\log^2 n)$ traces are needed to reconstruct random strings.

Using the lower bound of $\Omega(n)$ for worst-case strings, McGregor, Price, and Vorotnikova [MPV14] proved that $\Omega(\log^2 n)$ traces are needed to reconstruct random strings. Similarly, we transfer lower bounds for worst-case strings to a lower bound for random strings:

Proposition (Holden-L., 2019)

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 $c \log^{9/4} n / \sqrt{\log \log n}$ traces is at most $\exp(-n^{0.15})$.

Using the lower bound of $\Omega(n)$ for worst-case strings, McGregor, Price, and Vorotnikova [MPV14] proved that $\Omega(\log^2 n)$ traces are needed to reconstruct random strings. Similarly, we transfer lower bounds for worst-case strings to a lower bound for random strings:

Proposition (Holden-L., 2019)

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 $c \log^{9/4} n / \sqrt{\log \log n}$ traces is at most $\exp(-n^{0.15})$.

Upper bounds for random strings were studied by [BKKM04], Peres and Zhai (2017), and Holden, Pemantle, and Peres (2018); the last proved that $e^{O(\log^{1/3} n)} = n^{o(1)}$ traces suffice for reconstruction of random strings with any deletion probability $q \in (0, 1)$.

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

$G: (\bigcup_{k=0}^n \{0, 1\}^k)^T \rightarrow \{0, 1\}^n$ such that for all $\mathbf{x} \in \{0, 1\}^n$,

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

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

$G: (\bigcup_{k=0}^n \{0, 1\}^k)^T \rightarrow \{0, 1\}^n$ such that for all $\mathbf{x} \in \{0, 1\}^n$,

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

For $\mathbf{x}, \mathbf{y} \in \{0, 1\}^n$, 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

$G: (\bigcup_{k=0}^n \{0, 1\}^k)^T \rightarrow \{0, 1\}^n$ such that

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

Let μ and ν be probability measures on a finite space, X . The **total variation** distance between μ and ν is defined by

$$d_{\text{TV}}(\mu, \nu) := \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)| = \max_{A \subseteq X} [\mu(A) - \nu(A)] \in [0, 1]$$

Let μ and ν be probability measures on a finite space, X . The **total variation** distance between μ and ν is defined by

$$\begin{aligned}d_{\text{TV}}(\mu, \nu) &:= \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)| = \max_{A \subseteq X} [\mu(A) - \nu(A)] \in [0, 1] \\ &= \frac{1}{2} \sum_{x \in X} (\mu(x) \vee \nu(x) - \mu(x) \wedge \nu(x)).\end{aligned}$$

(Here, \vee is max and \wedge is min.)

Let μ and ν be probability measures on a finite space, X . The **total variation** distance between μ and ν is defined by

$$\begin{aligned} d_{\text{TV}}(\mu, \nu) &:= \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)| = \max_{A \subseteq X} [\mu(A) - \nu(A)] \in [0, 1] \\ &= \frac{1}{2} \sum_{x \in X} (\mu(x) \vee \nu(x) - \mu(x) \wedge \nu(x)). \end{aligned}$$

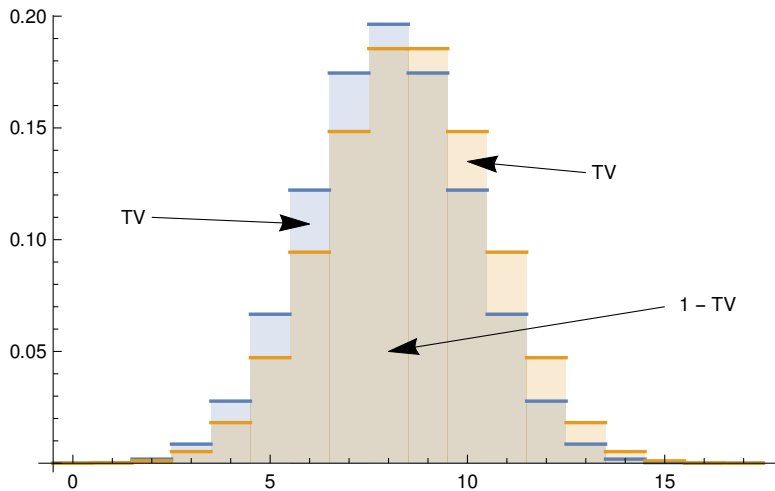
(Here, \vee is max and \wedge is min.) Since

$$1 = \frac{1}{2} \sum_{x \in X} (\mu(x) \vee \nu(x) + \mu(x) \wedge \nu(x)),$$

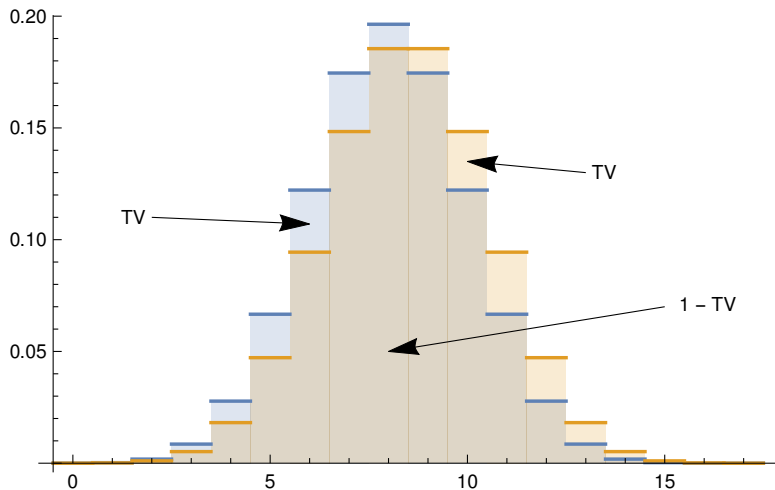
we have

$$1 - d_{\text{TV}}(\mu, \nu) = \sum_{x \in X} \mu(x) \wedge \nu(x).$$

Example: Binomial Distributions $\text{Bin}(16, 1/2)$ versus $\text{Bin}(17, 1/2)$



Example: Binomial Distributions $\text{Bin}(16, 1/2)$ versus $\text{Bin}(17, 1/2)$



For fixed $s \in (0, 1)$,

$$d_{\text{TV}}(\text{Bin}(n, s), \text{Bin}(n + 1, s)) = \Theta(n^{-1/2}).$$

Let $G: X \rightarrow \{\mu, \nu\}$ 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

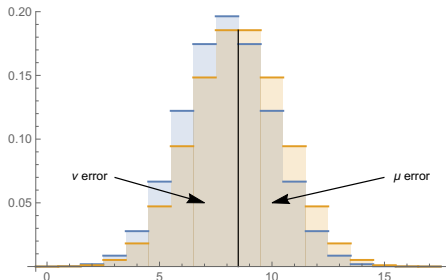
$$\mu[G(x) = \nu] + \nu[G(x) = \mu].$$

TV for Distinguishing Two Measures from Samples

Let $G: X \rightarrow \{\mu, \nu\}$ 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 $\mu[G(x) = \nu] + \nu[G(x) = \mu]$. The error probability sum is minimized by

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

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



Let $G: X \rightarrow \{\mu, \nu\}$ 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 $\mu[G(x) = \nu] + \nu[G(x) = \mu]$. The error probability sum is minimized by

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

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

Replacing μ, ν on X by μ^m, ν^m on X^m here, we get that for general m , the number of independent samples required to distinguish between μ and ν is determined precisely by $d_{\text{TV}}(\mu^m, \nu^m)$.

Let $G: X \rightarrow \{\mu, \nu\}$ 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 $\mu[G(x) = \nu] + \nu[G(x) = \mu]$. The error probability sum is minimized by

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

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

Replacing μ, ν on X by μ^m, ν^m on X^m here, we get that for general m , the number of independent samples required to distinguish between μ and ν is determined precisely by $d_{\text{TV}}(\mu^m, \nu^m)$.

One can then show that the number m of samples needed to distinguish μ from ν satisfies

$$\Omega(1/d_{\text{TV}}(\mu, \nu)) \leq m \leq O(1/d_{\text{TV}}^2(\mu, \nu)).$$

► Proof

The **Hellinger** distance between μ and ν is defined by

$$d_H(\mu, \nu) := \left(\sum_{x \in X} [\sqrt{\mu(x)} - \sqrt{\nu(x)}]^2 \right)^{1/2} \in [0, \sqrt{2}].$$

The **Hellinger** distance between μ and ν is defined by

$$d_{\text{H}}(\mu, \nu) := \left(\sum_{x \in X} [\sqrt{\mu(x)} - \sqrt{\nu(x)}]^2 \right)^{1/2} \in [0, \sqrt{2}].$$

This is useful because the number of samples to distinguish μ from ν is $\Theta(1/d_{\text{H}}^2(\mu, \nu))$.

► Proof

The **Hellinger** distance between μ and ν is defined by

$$d_H(\mu, \nu) := \left(\sum_{x \in X} [\sqrt{\mu(x)} - \sqrt{\nu(x)}]^2 \right)^{1/2} \in [0, \sqrt{2}].$$

This is useful because the number of samples to distinguish μ from ν is $\Theta(1/d_H^2(\mu, \nu))$.

► Proof

For fixed $s \in (0, 1)$, $d_H(\text{Bin}(n, s), \text{Bin}(n+1, s)) = \Theta(n^{-1/2})$.

The **Hellinger** distance between μ and ν is defined by

$$d_{\text{H}}(\mu, \nu) := \left(\sum_{x \in X} [\sqrt{\mu(x)} - \sqrt{\nu(x)}]^2 \right)^{1/2} \in [0, \sqrt{2}].$$

This is useful because the number of samples to distinguish μ from ν is $\Theta(1/d_{\text{H}}^2(\mu, \nu))$.

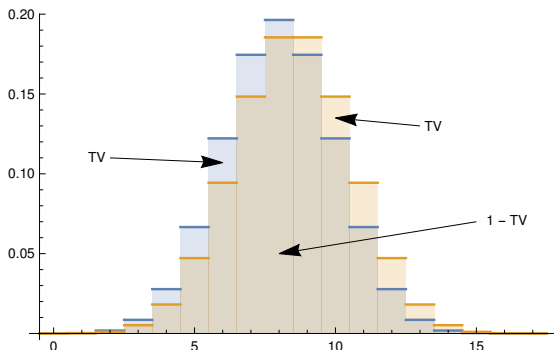
► Proof

For fixed $s \in (0, 1)$, $d_{\text{H}}(\text{Bin}(n, s), \text{Bin}(n+1, s)) = \Theta(n^{-1/2})$.

So it seems we should use $d_{\text{H}}(\mu, \nu)$ to distinguish μ from ν . But this is hard if μ and ν are complicated (such as the law of $\tilde{\mathbf{x}}_n$). By comparison, we have probabilistic techniques to estimate $d_{\text{TV}}(\mu, \nu)$:

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

where the minimum is taken over all couplings of U and V .



So it seems we should use $d_H(\mu, \nu)$ to distinguish μ from ν . But this is hard if μ and ν are complicated (such as the law of $\tilde{\mathbf{x}}_n$). By comparison, we have probabilistic techniques to estimate $d_{TV}(\mu, \nu)$:

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

where the minimum is taken over all couplings of U and V .

We have

$$d_{\text{TV}}(\mu, \nu) \leq d_{\text{H}}(\mu, \nu) \leq \sqrt{2 d_{\text{TV}}(\mu, \nu)}.$$

We have

$$d_{\text{TV}}(\mu, \nu) \leq d_{\text{H}}(\mu, \nu) \leq \sqrt{2 d_{\text{TV}}(\mu, \nu)}.$$

The right-hand inequality can be strengthened if for all $x \in X$, the ratio $\mu(x)/\nu(x)$ is close to 1:

Lemma

$$d_{\text{H}}^2(\mu, \nu) \leq \mu\{x; \nu(x) = 0\} + 2 \cdot \left\| \frac{\mu(x) - \nu(x)}{\nu(x)} \right\|_{\ell^\infty(\nu)} \cdot d_{\text{TV}}(\mu, \nu).$$

Distinguishing Between $\mathbf{x}'_n = (0)^{n-1}1(0)^n$ and $\mathbf{y}'_n = (0)^n1(0)^{n-1}$

Distinguishing Between $\mathbf{x}'_n = (0)^{n-1}1(0)^n$ and $\mathbf{y}'_n = (0)^n1(0)^{n-1}$

We need only track the number of 0s that appear to either side of the 1. (If the 1 does not appear in the trace, there is no information in that trace.)
These are $\text{Bin}(n-1, p) \otimes \text{Bin}(n, p)$ or $\text{Bin}(n, p) \otimes \text{Bin}(n-1, p)$.

We need only track the number of 0s that appear to either side of the 1. (If the 1 does not appear in the trace, there is no information in that trace.) These are $\text{Bin}(n-1, p) \otimes \text{Bin}(n, p)$ or $\text{Bin}(n, p) \otimes \text{Bin}(n-1, p)$.

Since the Hellinger distance between these distributions is $\Theta(n^{-1/2})$, it takes $\Theta(n)$ traces to distinguish them.

We need only track the number of 0s that appear to either side of the 1. (If the 1 does not appear in the trace, there is no information in that trace.) These are $\text{Bin}(n-1, p) \otimes \text{Bin}(n, p)$ or $\text{Bin}(n, p) \otimes \text{Bin}(n-1, p)$.

Since the Hellinger distance between these distributions is $\Theta(n^{-1/2})$, it takes $\Theta(n)$ traces to distinguish them.

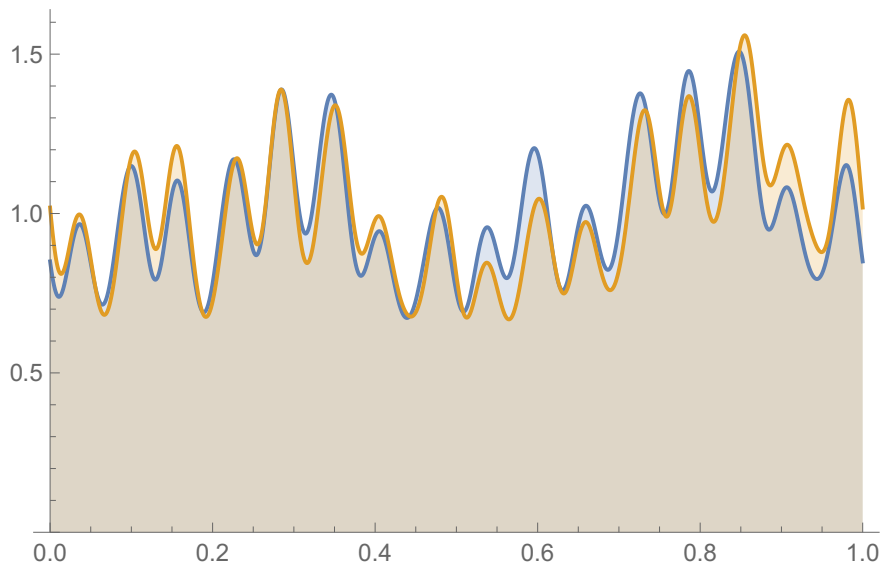
Similarly, if two n -bit strings differ in the number of 1s they contain, then $O(n)$ traces suffice to distinguish them.

Distinguishing Between $\mathbf{x}_n := (01)^{n-1}10(01)^n$ and $\mathbf{y}_n := (01)^n10(01)^{n-1}$

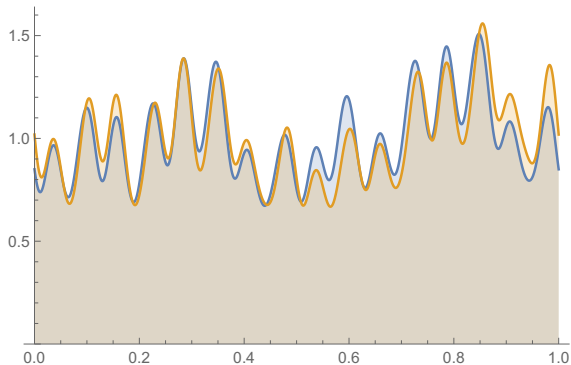
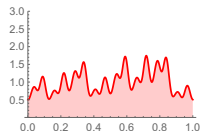
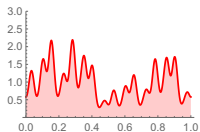
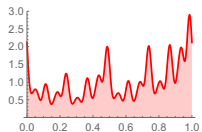
Distinguishing Between $\mathbf{x}_n := (01)^{n-1}10(01)^n$ and $\mathbf{y}_n := (01)^n10(01)^{n-1}$

We will construct a coupling of the traces from \mathbf{x}_n and \mathbf{y}_n in two stages.

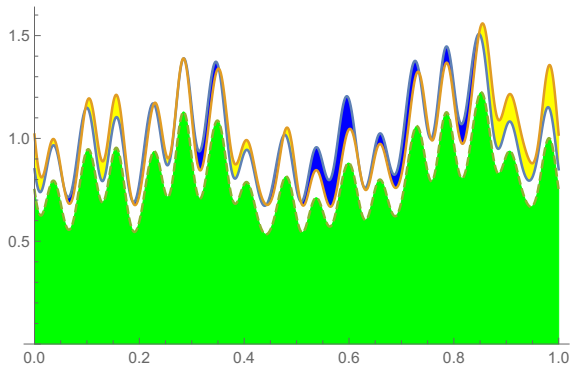
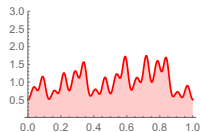
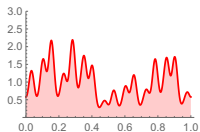
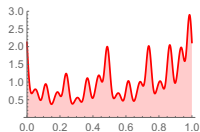
Two-Stage Coupling



Two-Stage Coupling

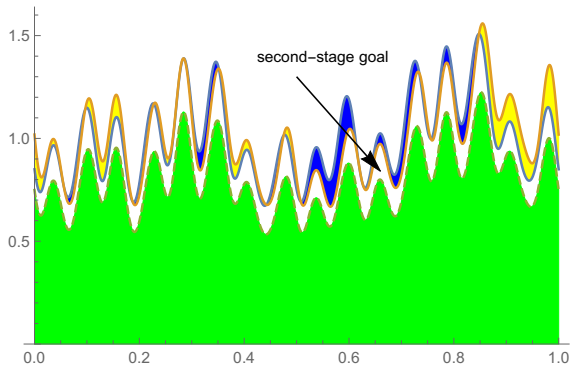
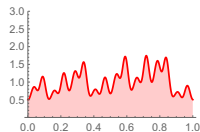
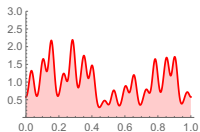
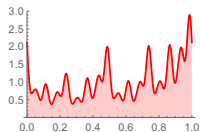


Two-Stage Coupling



- mixture (.2, .3, .5)
- mixture (.3, .4, .3)
- - - mixture (.2, .3, .3)

Two-Stage Coupling



- mixture (.2, .3, .5)
- mixture (.3, .4, .3)
- - - mixture (.2, .3, .3)

Distinguishing Between $\mathbf{x}_n := (01)^{n-1}10(01)^n$ and $\mathbf{y}_n := (01)^n10(01)^{n-1}$

We will construct a coupling of the traces from \mathbf{x}_n and \mathbf{y}_n in two stages.

We will construct a coupling of the traces from \mathbf{x}_n and \mathbf{y}_n in two stages. The first stage of the coupling is similar to what one does for \mathbf{x}'_n and \mathbf{y}'_n : 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 $\Theta(n^{-1/2})$.

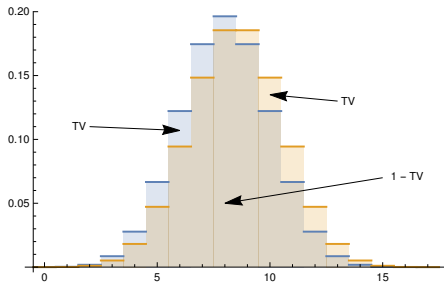
We will construct a coupling of the traces from \mathbf{x}_n and \mathbf{y}_n in two stages.

The first stage of the coupling is similar to what one does for \mathbf{x}'_n and \mathbf{y}'_n : 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 stage, 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.

Call the result of the first stage 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.

In the second stage of the coupling, we lower the bound on total variation by coupling still further on the event that the first coupling did not succeed in making the 2-partial traces the same for \mathbf{x}_n and \mathbf{y}_n .



Call the result of the first stage 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.

In the second stage of the coupling, we lower the bound on total variation by coupling still further on the event that the first coupling did not succeed in making the 2-partial traces the same for \mathbf{x}_n and \mathbf{y}_n .

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.

Call the result of the first stage 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.

In the second stage of the coupling, we lower the bound on total variation by coupling still further on the event that the first coupling did not succeed in making the 2-partial traces the same for \mathbf{x}_n and \mathbf{y}_n .

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. 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 stage.

The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect becomes,

0101, 0101, ..., 0101, 0101, 10 , 0101, ..., 0101, 0101, 0101
 0101, 0101, ..., 0101, 0101, 0101, 0101, ..., 0101, 0101, 0101

The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect becomes, and at the same time, has the remarkable property that what becomes of the other 0101-blocks is unaffected,

0101, 0101, ..., 0101, 0101, 10, 0101, ..., 0101, 0101, 0101
 0101, 0101, ..., 0101, 0101, 0101, 0101, ..., 0101, 0101, 0101

The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect becomes, 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.

0101, 0101, ..., 0101, 0101, 10 , 0101, ..., 0101, 0101, 0101
 0101, 0101, ..., 0101, 0101, 0101, 0101, ..., 0101, 0101, 0101

The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect becomes, 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.

0101, 0101, ..., 0101, 0101, **1**, 0101, ..., 0101, 0101, 0101
 0101, 0101, ..., 0101, 0101, 0101, 0101, ..., **1**, 0101, 0101

The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect becomes, 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.

0101, 0101, ..., 0101, 0101, **1**, 0101, ..., 0101, 0101, 0101
 0101, 0101, ..., 0101, 0101, 0101, 0101, ..., **1**, 0101, 0101

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).

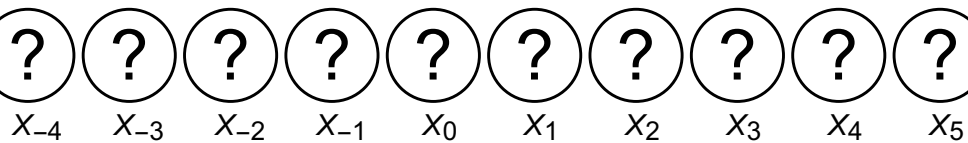
The idea is to find randomly a special 0101-block in the string without defect that becomes the same after deletion as the defect becomes, and at the same time, has the remarkable property that what becomes of the other 0101-blocks is unaffected, so that we can couple the defect to that special 0101-block.

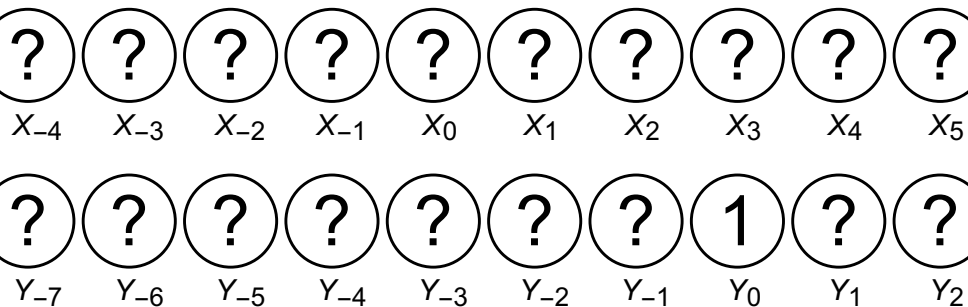
0101, 0101, ..., 0101, 0101, **1**, 0101, ..., 0101, 0101, 0101
 0101, 0101, ..., 0101, 0101, 0101, 0101, ..., **1**, 0101, 0101

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, 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.

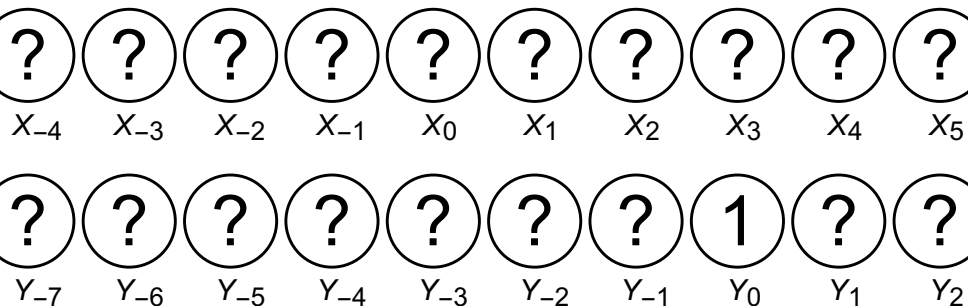
Combining the two coupling stages gives that the total variation distance between the traces is $O(n^{-3/4})$. Knowing the total variation distance is not sufficient to determine the number of traces required for reconstruction (it gives a lower bound of only $\Omega(n^{3/4})$ traces).

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). However, by throwing away a very small set of 2-partial traces in the first stage of the coupling, and using ratios for binomial distributions that are $1 + O(n^{-1/2}\sqrt{\log n})$, 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.



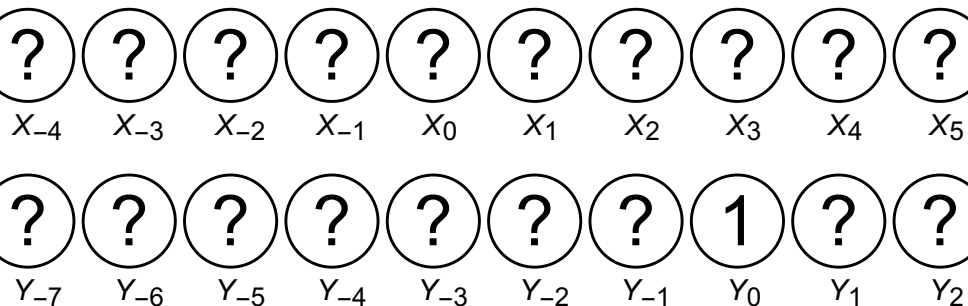


$$Y_k := X_{\tau+k} \quad \text{with} \quad X_\tau = 1$$



$$Y_k := X_{\tau+k} \quad \text{with} \quad X_\tau = 1$$

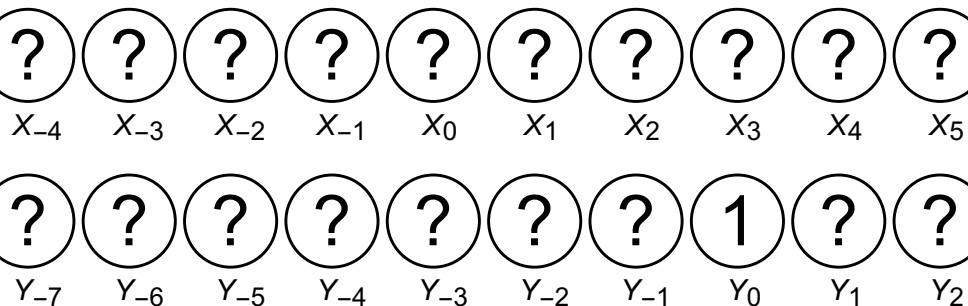
One *cannot* choose τ so that $(X_j)_{j \neq \tau}$ is a Bernoulli process conditioned on τ , because



$$Y_k := X_{\tau+k} \quad \text{with} \quad X_\tau = 1$$

One *cannot* choose τ so that $(X_j)_{j \neq \tau}$ is a Bernoulli process conditioned on τ , because that would lead to the contradiction

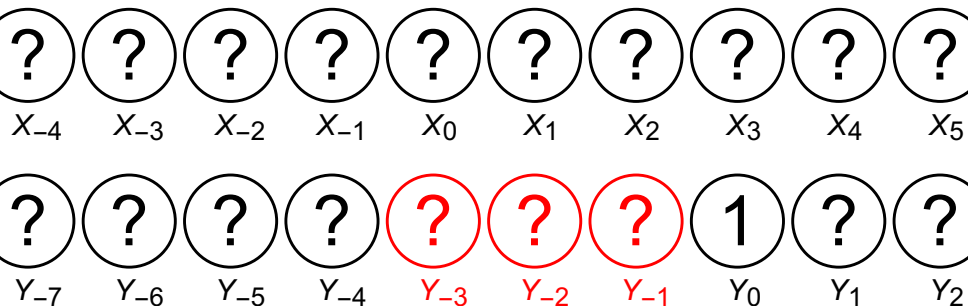
$$\mathbf{E} \left[\sum_{j \geq 0} X_j 2^{-j} \mid \tau \right] > \mathbf{E} \left[\sum_{j \geq 0} X_j 2^{-j} \right].$$



$$Y_k := X_{\tau+k} \quad \text{with} \quad X_\tau = 1$$

One *cannot* choose τ so that $(X_j)_{j \neq \tau}$ is a Bernoulli process conditioned on τ , because that would lead to the contradiction

$$\mathbf{E} \left[\mathbf{E} \left[\sum_{j \geq 0} X_j 2^{-j} \mid \tau \right] \right] > \mathbf{E} \left[\sum_{j \geq 0} X_j 2^{-j} \right].$$



$$Y_k := X_{\tau+k} \quad \text{with} \quad X_\tau = 1$$

One *cannot* choose τ so that $(X_j)_{j \neq \tau}$ is a Bernoulli process conditioned on τ , because that would lead to the contradiction

$$\mathbf{E} \left[\mathbf{E} \left[\sum_{j \geq 0} X_j 2^{-j} \mid \tau \right] \right] > \mathbf{E} \left[\sum_{j \geq 0} X_j 2^{-j} \right].$$

Assume $(X_k)_{k \in \mathbb{Z}}$ is a Bernoulli(1/2) process. Let $\sigma \geq 2$ be the first time that simple random walk $(\sum_{k=0}^t (2X_k - 1))_{t \geq 0}$ on \mathbb{Z} returns to 0. Set

$$\tau := \begin{cases} 0 & \text{if } X_0 = 1, \\ \sigma & \text{if } X_0 = 0. \end{cases}$$

Assume $(X_k)_{k \in \mathbb{Z}}$ is a Bernoulli(1/2) process. Let $\sigma \geq 2$ be the first time that simple random walk $(\sum_{k=0}^t (2X_k - 1))_{t \geq 0}$ on \mathbb{Z} returns to 0. Set

$$\tau := \begin{cases} 0 & \text{if } X_0 = 1, \\ \sigma & \text{if } X_0 = 0. \end{cases}$$

One checks that

- $X_0 = 0$ iff $X_\sigma = 1$;

Assume $(X_k)_{k \in \mathbb{Z}}$ is a Bernoulli(1/2) process. Let $\sigma \geq 2$ be the first time that simple random walk $(\sum_{k=0}^t (2X_k - 1))_{t \geq 0}$ on \mathbb{Z} returns to 0. Set

$$\tau := \begin{cases} 0 & \text{if } X_0 = 1, \\ \sigma & \text{if } X_0 = 0. \end{cases}$$

One checks that

- $X_0 = 0$ iff $X_\sigma = 1$;
- $(X_k)_{k \in \mathbb{Z}}$ has the same law as $(X_{\sigma-k})_{k \in \mathbb{Z}}$;

Assume $(X_k)_{k \in \mathbb{Z}}$ is a Bernoulli(1/2) process. Let $\sigma \geq 2$ be the first time that simple random walk $(\sum_{k=0}^t (2X_k - 1))_{t \geq 0}$ on \mathbb{Z} returns to 0. Set

$$\tau := \begin{cases} 0 & \text{if } X_0 = 1, \\ \sigma & \text{if } X_0 = 0. \end{cases}$$

One checks that

- $X_0 = 0$ iff $X_\sigma = 1$;
- $(X_k)_{k \in \mathbb{Z}}$ has the same law as $(X_{\sigma-k})_{k \in \mathbb{Z}}$;
- $(Y_k)_{k \in \mathbb{Z} \setminus \{0\}}$ is a Bernoulli(1/2) process, even conditioned on X_0 .

Supplementary Material

Write $\Delta_{\mathbf{x}}$ for the law of the trace we obtain when applying the deletion channel with deletion probability q to \mathbf{x} . It suffices to prove the lower bound of

Proposition (Holden-L., 2019)

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

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

Write $\Delta_{\mathbf{x}}$ for the law of the trace we obtain when applying the deletion channel with deletion probability q to \mathbf{x} . It suffices to prove the lower bound of

Proposition (Holden-L., 2019)

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

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

It suffices to define $Z(\tilde{\mathbf{x}})$ such that

$$d_{\text{TV}}(\Delta_{\mathbf{x}_n}, \Delta_{\mathbf{y}_n}) \geq d_{\text{TV}}(Z(\tilde{\mathbf{x}}_n), Z(\tilde{\mathbf{y}}_n))$$

Write $\Delta_{\mathbf{x}}$ for the law of the trace we obtain when applying the deletion channel with deletion probability q to \mathbf{x} . It suffices to prove the lower bound of

Proposition (Holden-L., 2019)

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

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

It suffices to define $Z(\tilde{\mathbf{x}})$ such that

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

For $n \in \mathbb{N}$, $\mathbf{x} \in \{0, 1\}^{4n}$, and $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_\ell)$ the trace of \mathbf{x} , define $Z(\tilde{\mathbf{x}})$ as $Z(\tilde{\mathbf{x}}) := \#\{k; 2np + 1 \leq k \leq (2np + \sqrt{npq}) \wedge (\ell - 1), \tilde{x}_k = \tilde{x}_{k+1} = 1\}$.

Lemma

We have

$$\mathbf{E}[Z(\tilde{\mathbf{y}}_n)] - \mathbf{E}[Z(\tilde{\mathbf{x}}_n)] = \Theta(n^{-1/2}) \quad \text{and} \quad \mathbf{E}[Z(\tilde{\mathbf{y}}_n)] > \mathbf{E}[Z(\tilde{\mathbf{x}}_n)]$$

for all sufficiently large n .

Lemma

There is a constant $c > 0$ such that for all $r > 0$ and $n \in \mathbb{N}$,

$$\mathbf{P}\left[|Z(\tilde{\mathbf{y}}_n) - \mathbf{E}[Z(\tilde{\mathbf{y}}_n)]| > rn^{1/4}\right] \leq 2e^{-cr^2}$$

and

$$\mathbf{P}\left[|Z(\tilde{\mathbf{x}}_n) - \mathbf{E}[Z(\tilde{\mathbf{x}}_n)]| > rn^{1/4}\right] \leq 2e^{-cr^2}.$$

Lemma

Let X and Y be discrete, real-valued random variables such that

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

Then

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

◀ Back to Upper Bound

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

where the minimum is taken over all couplings of U and V . By using independent couplings, it follows that

$$1 - d_{\text{TV}}(\mu^m, \nu^m) \geq (1 - d_{\text{TV}}(\mu, \nu))^m.$$

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

where the minimum is taken over all couplings of U and V . By using independent couplings, it follows that

$$1 - d_{\text{TV}}(\mu^m, \nu^m) \geq (1 - d_{\text{TV}}(\mu, \nu))^m.$$

Thus, we need at least $m = \Omega(1/d_{\text{TV}}(\mu, \nu))$ to distinguish μ from ν .

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

where the minimum is taken over all couplings of U and V . By using independent couplings, it follows that

$$1 - d_{\text{TV}}(\mu^m, \nu^m) \geq (1 - d_{\text{TV}}(\mu, \nu))^m.$$

Thus, we need at least $m = \Omega(1/d_{\text{TV}}(\mu, \nu))$ to distinguish μ from ν .

Let $A \subset X$ be such that $\mu(A) - \nu(A) = d_{\text{TV}}(\mu, \nu)$. Given m independent samples $\mathfrak{X}_m = (x_1, \dots, x_m)$ from one of the measures, let

$\bar{u} := m^{-1} \sum_{j=1}^m \mathbf{1}_{[x_j \in A]}$ be the number of times that A occurs. Define

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

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

where the minimum is taken over all couplings of U and V . By using independent couplings, it follows that

$$1 - d_{\text{TV}}(\mu^m, \nu^m) \geq (1 - d_{\text{TV}}(\mu, \nu))^m.$$

Thus, we need at least $m = \Omega(1/d_{\text{TV}}(\mu, \nu))$ to distinguish μ from ν .

Let $A \subset X$ be such that $\mu(A) - \nu(A) = d_{\text{TV}}(\mu, \nu)$. Given m independent samples $\mathfrak{X}_m = (x_1, \dots, x_m)$ from one of the measures, let

$\bar{u} := m^{-1} \sum_{j=1}^m \mathbf{1}_{[x_j \in A]}$ be the number of times that A occurs. Define

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

Apply Hoeffding–Azuma:

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

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

where the minimum is taken over all couplings of U and V . By using independent couplings, it follows that

$$1 - d_{\text{TV}}(\mu^m, \nu^m) \geq (1 - d_{\text{TV}}(\mu, \nu))^m.$$

Thus, we need at least $m = \Omega(1/d_{\text{TV}}(\mu, \nu))$ to distinguish μ from ν .

Let $A \subset X$ be such that $\mu(A) - \nu(A) = d_{\text{TV}}(\mu, \nu)$. Given m independent samples $\mathfrak{X}_m = (x_1, \dots, x_m)$ from one of the measures, let

$\bar{u} := m^{-1} \sum_{j=1}^m \mathbf{1}_{[x_j \in A]}$ be the number of times that A occurs. Define

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

Apply Hoeffding–Azuma:

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

Thus, $m = O(1/d_{\text{TV}}(\mu, \nu)^2)$ samples suffice.

Theorem

If μ and ν are probability measures with $d_{\text{H}}(\mu, \nu) \leq 1/2$, then for $m \geq 1/(4 d_{\text{H}}^2(\mu, \nu))$,

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

(The right-hand inequality holds for all m, μ, ν .)

Theorem

If μ and ν are probability measures with $d_{\text{H}}(\mu, \nu) \leq 1/2$, then for $m \geq 1/(4 d_{\text{H}}^2(\mu, \nu))$,

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

(The right-hand inequality holds for all m, μ, ν .)

Proof of upper bound.

$$d_{\text{H}}^2(\mu, \nu) = 2\left(1 - \sum_x \sqrt{\mu(x)\nu(x)}\right),$$

Theorem

If μ and ν are probability measures with $d_{\text{H}}(\mu, \nu) \leq 1/2$, then for $m \geq 1/(4 d_{\text{H}}^2(\mu, \nu))$,

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

(The right-hand inequality holds for all m, μ, ν .)

Proof of upper bound.

$$d_{\text{H}}^2(\mu, \nu) = 2\left(1 - \sum_x \sqrt{\mu(x)\nu(x)}\right),$$

whence

$$1 - \frac{1}{2} d_{\text{H}}^2(\mu^m, \nu^m) = \left(1 - \frac{1}{2} d_{\text{H}}^2(\mu, \nu)\right)^m.$$

Theorem

If μ and ν are probability measures with $d_{\text{H}}(\mu, \nu) \leq 1/2$, then for $m \geq 1/(4 d_{\text{H}}^2(\mu, \nu))$,

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

(The right-hand inequality holds for all m, μ, ν .)

Proof of upper bound.

$$d_{\text{H}}^2(\mu, \nu) = 2\left(1 - \sum_x \sqrt{\mu(x)\nu(x)}\right),$$

whence

$$1 - \frac{1}{2} d_{\text{H}}^2(\mu^m, \nu^m) = \left(1 - \frac{1}{2} d_{\text{H}}^2(\mu, \nu)\right)^m.$$

Combine with $d_{\text{TV}}(\mu^m, \nu^m) \geq \frac{1}{2} d_{\text{H}}^2(\mu^m, \nu^m)$ and $1 - x \leq e^{-x}$. □

Proof of lower bound.

Define $r := \lfloor 1/(4 d_{\text{H}}^2(\mu, \nu)) \rfloor \geq 1$. We have

$$d_{\text{TV}}^2(\mu^r, \nu^r) \leq d_{\text{H}}^2(\mu^r, \nu^r) \leq r d_{\text{H}}^2(\mu, \nu) \leq 1/4.$$

Proof of lower bound.

Define $r := \lfloor 1/(4 d_{\text{H}}^2(\mu, \nu)) \rfloor \geq 1$. We have

$$d_{\text{TV}}^2(\mu^r, \nu^r) \leq d_{\text{H}}^2(\mu^r, \nu^r) \leq r d_{\text{H}}^2(\mu, \nu) \leq 1/4.$$

Also, $1 - d_{\text{TV}}(\mu^{\lceil m/r \rceil}, \nu^{\lceil m/r \rceil}) \geq [1 - d_{\text{TV}}(\mu, \nu)]^{\lceil m/r \rceil}$.

Proof of lower bound.

Define $r := \lfloor 1/(4 d_{\text{H}}^2(\mu, \nu)) \rfloor \geq 1$. We have

$$d_{\text{TV}}^2(\mu^r, \nu^r) \leq d_{\text{H}}^2(\mu^r, \nu^r) \leq r d_{\text{H}}^2(\mu, \nu) \leq 1/4.$$

Also, $1 - d_{\text{TV}}(\mu^{\lceil m/r \rceil}, \nu^{\lceil m/r \rceil}) \geq [1 - d_{\text{TV}}(\mu, \nu)]^{\lceil m/r \rceil}$. Therefore,

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

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