Project House

Abstract—

Contents

1 Preliminary Concepts

A metric on a set is a function that satisfies the minimal properties we might expect of a distance. Formally, we have:

Definition 1 (Metric Space)**.** *A metric d on a set X is a function $d : X \times X \to [0, \infty)$ such that for all $x, y, z \in X$:*

1) $d(x, y) \geqq 0$ *and* $d(x, y) = 0$ *if and only if* $x = y$
2) $d(x, y) = d(y, x)$ *(symmetry)*
3) $d(x, y) \leqq d(x, z) + d(z, x)$ *(triangle inequality)*

A metric space (X, d) is a set X with a metric d defined on X.

Definition 2 (Quasi-Metric Space)**.** *A quasi-metric space is a set Z with a function $\rho : Z \times Z \to [0, \infty)$ which satisfies the conditions:*

1) $\rho(z, z') \geqq 0$ *for every* $z, z' \in Z$ *and* $\rho(z, z') = 0$ *if and only if* $z = z'$;
2) $\rho(z, z') = \rho(z', z)$ *for every* $z, z' \in Z$;
3) $\rho(z, z'') \leqq K \max\{\rho(z, z'), \rho(z', z'')\}$ *for every* $z, z', z'' \in Z$ *and some fixed* $K \geqq 1$ *independent of* z, z', z''.

The function ρ is known as a quasi-metric, or more specifically, a K-quasi-metric. The property (3) represents a weakening of the triangular inequality.

A quasi-metric may be effectively used to induce a metric [2], albeit under some restrictions [3].

Definition 3 (Jaccard Distance)**.** *The Jaccard distance, which measures dissimilarity between sample sets is defined as the ratio of the difference of the sizes of the union and the intersection of two sets to the cardinality of the union:*

$$d_J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \tag{1}$$

Note, that d_J is the ratio of the cardinality of the symmetric difference

$$A \triangle B = (A \cup B) - (A \cap B) \tag{2}$$

to the union. This distance is provably a metric on the collection of all finite sets.

The Jaccard distance may be naturally extended to a distance on a space of sequences, via considering the subword set.

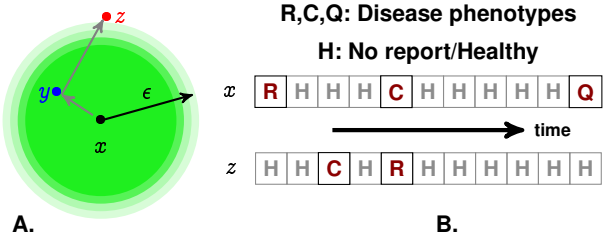

Fig. 1. Plate A illustrates the need of a distance metric in defining neighborhoods meaningfully. If z is just outside an ϵ-neighborhood of x, then every possible path to z from x should be larger than ϵ; one should not be able to find a shorter path by first jumping to some point y in the neighborhood. This, along with the fact, that only coincident points have zero distance, sums up our physical notion of a "distance". Here we want to define such a function in the space of medical histories. Plate (B) illustrates two such history fragments. We want the distance between them to be reflective of our intuitive understanding of the problem. Namely, rare events present in one history and not the other should lead to a larger distance, while shared rare events should reduce it. Also, very common event sequences should not increase the distance too much.

Definition 4 (Subword Set)**.** *Given an ordered sequence $s = \{s_i\}$, the subword set of s, denoted as $\Omega(s)$, is defined as:*

$$\Omega(s) = \{\omega : |\omega| \geqq 1, \text{ and } \exists \omega_1, \omega_2, \text{ s.t. } s = \omega_1 \omega \omega_2\} \tag{3}$$

Note that the subword set is much smaller than the power set:

$$|\Omega(s)| = |s| + (|s| - 1) + \cdots + 1 = \frac{1}{2} |s| (|s| + 1) \tag{4}$$

where $|s|$ denotes the length of the sequence s, and $|\Omega(s)|$ denotes the cardinality of the subword set.

Definition 5 (Sequential Jaccard Distance)**.** *The sequential Jaccard distance between two sequences s_1, s_2, denoted as $\mathcal{J}(s_1, s_2)$, is defined as:*

$$\mathcal{J}(s_1, s_2) = \frac{|\Omega(s_1) \cup \Omega(s_2)| - |\Omega(s_1) \cap \Omega(s_2)|}{|\Omega(s_1) \cup \Omega(s_2)|} \tag{5}$$

Definition 6 (Weighted Jaccard Sequential Distance)**.** *Given two sequences s_1, s_2 over some finite akphabet Σ, and an integer-valued weighting function $w : \Sigma \to \mathbb{N}$ on sequence entries, the weighted Jaccard sequential distance is defined as:*

$$\mathcal{J}_w(s_1, s_2) = \frac{\sum_{r \in \Omega(s_1) \cup \Omega(s_2)} w(r) \quad - \quad \sum_{r \in \Omega(s_1) \cap \Omega(s_2)} w(r)}{\sum_{r \in \Omega(s_1) \cup \Omega(s_2)} w(r)} \tag{6}$$

It is trivial to show that weighted Jaccard sequential distance is indeed a metric.

Notation 1.
1) *The set of disease phenotypes referenced in our database of consideration is denoted as $\mathbb{D}$. We augment $\mathbb{D}$ with an additional symbol $\{d_0\}$ which indicates a "healthy" status.*
2) *The time interval of interest is indexed by the set of positive integers $\mathcal{T}$, where:*

$$\mathcal{T} = \{1, 2, \cdots\} \tag{7}$$

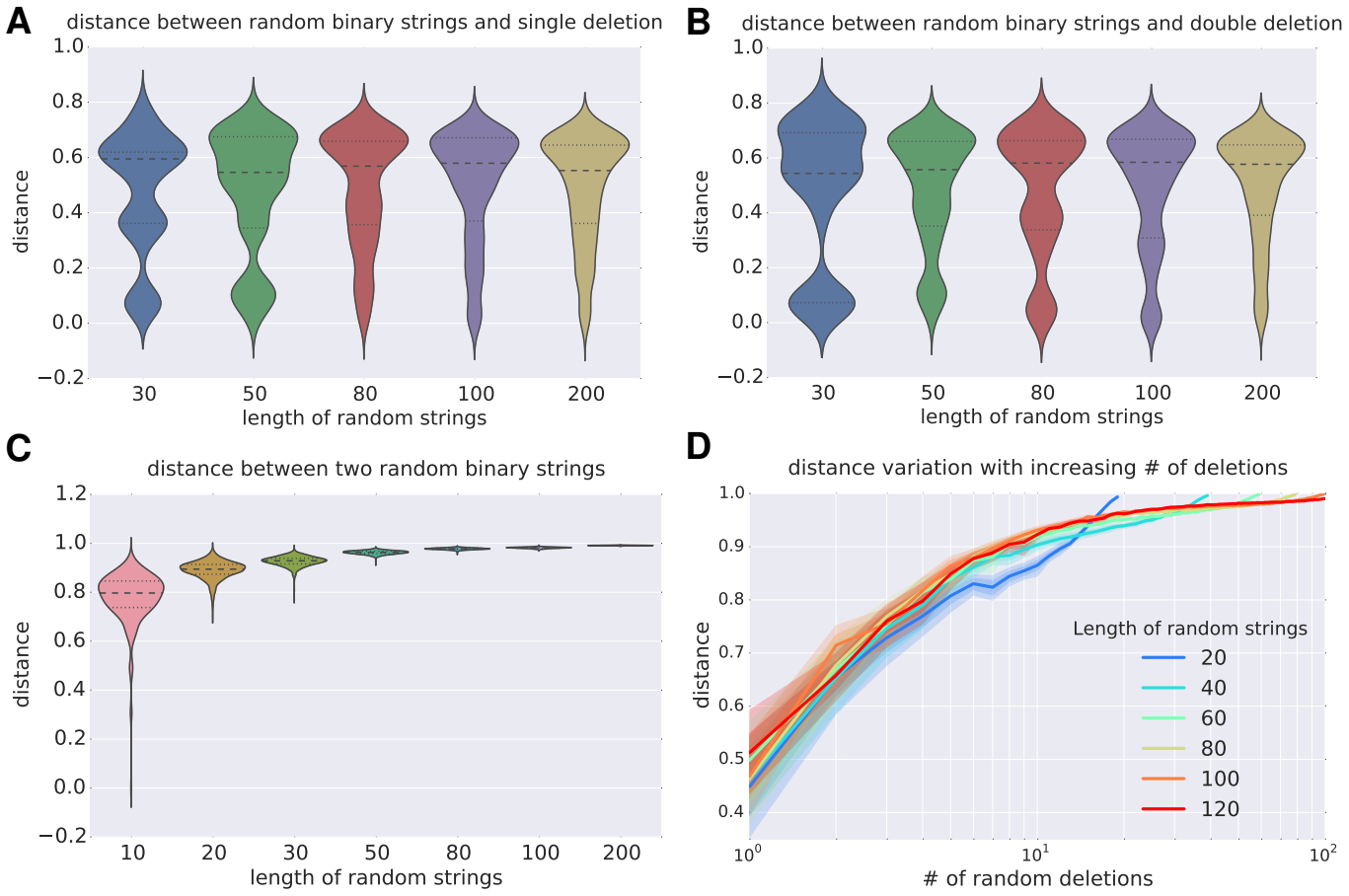

Fig. 2. Effect of deletions/variations on the sequential Jaccard distance between random binary strings. Plate A shows the effect of single deletions on random binary strings of increasing length. Plate B illustrates the same effect for two random deletions. Plate C shows how the distance between two random strings converge to 1 as the length of the strings is increased. Plate D illustrates the effect of increasing the number of deletions on random binary strings of increasing length, with 99.9999999% confidence bounds. Note that the sequential *weighted* Jaccard distance is more complicated.

For example, if the unit of time is chosen to be [years], then a disease reported at $t = 1$ refers to the report occurring before the first birthday of the subject.

3) *A indexed phenotype is a tuple:*

$$(t, d), t \in \mathcal{T}, d \in \mathbb{D} \tag{8}$$

where the particular reported phenotype is indexed at the age it is reported.

Definition 7 (History)**.** *A history h_t is a sequence of indexed phenotypes representing the reported medical history of a subject, ending at time index t:*

$$h_{t_n} = \{(t_1, d_1), \cdots, (t_i, d_i), \cdots, (t_n, d_n)\}, \quad \text{where } t_i \in \mathcal{T}, d_i \in \mathbb{D}, t_1 \leqq \cdots \leqq t_n \tag{9}$$

Notation 2.

- *For history h_t, we denote:*

$$h_t(t_1) = \begin{cases} d \in \mathbb{D}, & \text{if } (t_1, d) \in h_t \\ d_0, & \text{otherwise} \end{cases} \tag{10}$$

- *Sub-histories are denoted as follows:*

$$h_t|_\tau = \{(t', d') \in h_t : t' \leqq \tau\} \tag{11}$$

We also simply write h_τ for $h_t|_\tau$ where no confusion arises.

- *Also, we use $\mathcal{T}(h_t)$ to denote the set of timestamps in h_t.*

$$\mathcal{T}(h_t) \subseteqq \mathcal{T}, \text{ such that } t' \in \mathcal{T}(h_t) \Rightarrow \exists d \in \mathbb{D}(\exists (t', d) \in h_t)$$

Notation 3. *If two appropriately defined empirical distributions f, g pass the Kolmogorov-Smirnov two sample equality test at significance level α, then we write:*

$$f \sim_{\mathcal{KS}(\alpha)} g \tag{12}$$

2 Conceptual Framework: The House Function Ψ

Notation 4 (ϵ-Neighborhood of History)**.** *We would define a quasi-metric on the space of histories in the sequel. Open ϵ-neighborhoods in this space, centered at h_t, is denoted by $\left[h_t\right]_\epsilon$*

For the purpose of the next definition, we interpret $\left[h_t\right]_\epsilon$ with respect to an arbitrary quasi-metric on the space of histories. Later, we would make precise the specific metric we use.

Definition 8 (The House Function Ψ)**.** *The House function specifies the probability of reporting a specific phenotype at a specific time, conditioned on an ϵ-neighborhood of a specific history:*

$$\forall t \in \mathcal{T}, d \in \mathbb{D}, \ \Psi_\epsilon(t, d, h_{t-1}) = Pr\left((t, d) \middle| \left[h_{t-1}\right]_\epsilon\right) \tag{13}$$

where, denoting the chosen quasi-metric as $\Theta(\cdot, \cdot)$, we have:

$$h'_{t'} \in \left[h_{t-1}\right]_\epsilon \text{ if } \Theta(h'_{t'}, h_t) < \epsilon \tag{14}$$

The central objective in this work is to efficiently compute the House function, along with estimates of statistical significance.

The unconditional probability of observing a specific history can now be written in terms of the House function as follows:

$$Pr(h_t) = \prod_{i=1}^{t} \Psi_\epsilon(i, h_t(i), h_{i-1}) \tag{15}$$

Importantly, there are no Markovian assumptions injected in Eq. (15); we do not yet have any notion of state, and the probability of future observations is allowed to depend on the entire observed past.

The House function thus represents a model of the emergent dynamical relationships in the database, that is maximally agnostic of prior knowledge. The need for conditioning on ϵ-neighborhoods should be clear: even with a substantially large database of patient histories, there are very few exact repeats; implying that probabilities cannot be reliably estimated unless we put togther a large enough set from "similar" histories. Also, note that such a notion of similarity must be specified a priori; we cannot for example define histories to be similar if future evolution is close in some sense, since then we would need to quantify what we mean by similar futures (thus we cannot invoke the notion of "causal states" [1] directly). The absence of a Markovian assumption at this level makes the framework significantly more general compared to n-gram models (we do not a priori decide on a bound on n). Furthermore, while the formulation definitely is remniscent of a *language model*, note that the presence of substantial repeats of words and phrases in any natural language makes the explicit use of neighborhoods in probability calculations somewhat unnecessary, and rare sequences are generally handled via some kind of pre-defined smoothing function. In contrast, in our case, almost all histories are rare.

It is clear that we need to explicitly specify the quasi-metric to compute the House function. However, any intutively satisfactory notion of distance that we come up with, must in turn depend on the probabilistic structure emergent in the space of observed histories, and therefore must functionally depend on the House function itself. This leads to a recursive relationship between the two.

Before proceeding, we note that the average hazard rate of any phenotype may be expressed in terms of the House function as follows:

$$\forall d \in \mathbb{D}, \zeta(d) = \frac{1}{\mathcal{T}} \sum_{t \in \mathcal{T}} \lim_{\epsilon \to \infty} \Psi(t, d, [h_t]_\epsilon) \tag{16}$$

Notation 5. *Let $\Psi_\epsilon(t, \cdot, h_{t-1})$ denote a probability distribution over $\mathbb{D}$, such that:*

$$\forall j \in \mathbb{D}, \Psi_\epsilon(t, \cdot, h_{t-1})\big|_j = \Psi_\epsilon(t, j, h_{t-1}) \tag{17}$$

3 The Quasi-Metric Θ

The quasi-metric is what injects our physical intuition into the problem. Defining a distance between tuple-sequences is not terribly difficult; however identifying a distance function that comports with our physical intuition on the properties that such a function should have, is somewhat more non-trivial.

3.1 Intuitive Properties Of the Quasi-metric

1) Histories which have high unconditional probability of occurrence should be close. In other words, a sequence of common occurrences should not contribute to a large deviation in the computed distance.
2) Rare events should induce large deviations; if two histories are nearly identical except that one carries the record of an uncommon disease phenotype, then the distance between the two should be large.
3) If two histories are substantially different, but share the record of a rare phenotype, then they should be close under the proposed metric.
4) Permutations of the records in time should matter. If one history is simply a permutation of another, and such a transformation leads to a reduction in its unconditional probability of occurrence, *i.e.*, makes it more rare, then the distance between them should increase. Even if the unconditional probability is identical, the distance should not be zero.
5) Any distance that we propose must be informed by the database of histories; it must not be computable from the sequences alone in the absence of a sufficiently large set of observed histories. In other words, classical string edit metrics will not suffice.

Additionally, we should also strive to keep our proposed function as simple as possible.

Definition 9 (Quasi-metric)**.** *Let $h_t, h'_{t'}$ be two histories. Let $\zeta^{-1} : \mathbb{D} \to \mathbb{N}$ be:*

$$\forall d \in \mathbb{D}, t \in \mathcal{T}, \zeta^{-1}((t,d)) = \left\lceil \frac{1}{\zeta(d)} \right\rceil^\gamma, \gamma > 1, \gamma \in \mathbb{N} \tag{18}$$

and where:

$$\forall d \in \mathbb{D}, \zeta(d) = \frac{1}{\mathcal{T}} \sum_{t \in \mathcal{T}} \lim_{\epsilon \to \infty} \Psi(t, d, [h_t]_\epsilon) \tag{19}$$

Then, we specify the quasi-metric, as a function of γ, as:

$$\Theta_\gamma(h_t, h'_{t'}) = \frac{1}{2}\left(\frac{1}{Pr(h'_{t'})} + \frac{1}{Pr(h_t)} - 2\right) \mathcal{J}_{\zeta^{-1}}(h_t, h'_{t'}) \tag{20}$$

Unless otherwise specified, we drop the subscript in Θ_γ, and assume $\gamma = 2$ in the sequel. The interpretation of this parameter will be discussed later.

It is immediate from Definition 9 that $\Theta_\gamma(h_t, h'_{t'})$ is symmetric, always non-negative, and zero if and only if the histories $h_t, h'_{t'}$ are identical. In particular, we have the following result.

Lemma 1 (Quasi-metric)**.** *If the unconditional probability of the histories is bounded between $(\theta, 1-\theta), \theta > 0$, then Θ_γ is a $\left(\frac{1}{\theta^2} + O(\theta)\right)$-quasi-metric on the space of histories, for all $\gamma \in \mathbb{N}$.*

Proof: To be given in SI. □

We first explore the salient properties of this quasi-metric, to check that it does indeed satisfy the intuitive constraints laid out before.

3.2 Θ Properties

Let $h_t, h'_{t'}$ be two histories. Then:

1) If $Pr(h_t) \to 1, Pr(h'_{t'}) \to 1$, then clearly they do not contain any rare phenotype, implying that $\mathcal{J}_{\zeta^{-1}}(h_t, h'_{t'})$ remains bounded. It follows that
$$\frac{1}{Pr(h'_{t'})} + \frac{1}{Pr(h_t)} \to 2^+ \Rightarrow \Theta_\gamma(h_t, h'_{t'}) \to 0^+ \tag{21}$$
2) On the other hand, if $Pr(h_t) \to 0, Pr(h'_{t'}) \to 0$, but there are no common rare phenotypes, then again $\mathcal{J}_{\zeta^{-1}}(h_t, h'_{t'})$ remains bounded. It then follows that $\Theta_\gamma(h_t, h'_{t'}) \to \infty$.
3) But, if $Pr(h_t) \to 0, Pr(h'_{t'}) \to 0$, and the histories share the record of a rare phenotype d' (possibly at different time points), such that $\zeta(d') \to 0$, then for $\gamma > 1$, the Jaccard term goes to zero faster, and we have $\Theta_\gamma(h_t, h'_{t'}) \to 0$.
4) Clearly, permutations lead to a positive deviation, since we use the Jaccard distance on the weighted subword set of the histories (See Definition 6)
5) It follows from definition that the proposed quasi-metric is informed by the observations in the database, Indeed, as observed before, $\Theta_\gamma(h_t, h'_{t'})$ may be written exclusively in terms of the House function.

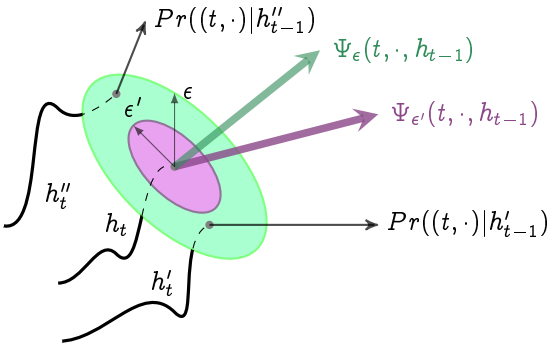

Fig. 3. Notion of stable neighborhoods: There should exist some $\epsilon' < \epsilon$, such that $\Psi_\epsilon(t,d,h_{t-1})$ and $\Psi_{\epsilon'}(t,d,h_{t-1})$ is not too different.

4 Computation of the House Function

4.1 Recursive Computation

The House function is approximated via the empirical estimate:

$$\forall d \in \mathbb{D}, \Psi_\epsilon(t,d,h_{t-1}) \approx \left|[h_{t-1}]_\epsilon\right|^{-1} \sum_{h'_t : h'_{t-1} \in [h_{t-1}]_\epsilon} \mathbb{I}((t,d) \in h'_t) \tag{22}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Definitions 8 and 9 might seem to have a circular dependence. To resolve this, one must adopt a recursive approach for computing the House function. Note that at age of 1 week (assuming our time unit is *weeks*) the past history is empty, and the House function simply specifies the unconditional probabilities of the different phenotypes occurring at age 1 week :

$$\forall d \in \mathbb{D}, \Psi_\epsilon(1,d,h_0) = Pr((1,d)|\varnothing) = Pr((1,d)) \tag{23}$$

For computing Ψ_ϵ at later times, we must first compute ϵ-neighborhoods of histories for the preceding time unit. Since we can compute the House function at age 1 week, we can compute the distance between histories upto age 1 week, which then allows us to compute the House functions for age 2 weeks, which in turn allows us to compute distances between histories upto age 2, and so on.

4.2 Calculation of Statistical Significance

In the development so far, ϵ appears as a free parameter. However, a choice of too large an ϵ makes the neighborhood include histories that do not have comparable futures. When we choose a small enough neighborhood, the probability distribution of the next events should be invariant over sub-neighborhoods. This leads us to the notion of stable neighborhoods, and allows us to formulate a notion of optimal choice for the parameter ϵ.

Definition 10 (α-stable Neighborhoods)**.** *For $\epsilon' < \epsilon$, let:*

$$\forall d \in \mathbb{D}, p(d) \triangleq \left|[h_{t-1}]_\epsilon \setminus [h_{t-1}]_{\epsilon'}\right|^{-1} \sum_{h'_t : h'_{t-1} \in [h_{t-1}]_\epsilon \setminus [h_{t-1}]_{\epsilon'}} \mathbb{I}((t,d) \in h'_t) \tag{24}$$

Then, $[h_{t-1}]_\epsilon$ is α-stable if:

$$\exists \epsilon' \in (0,\epsilon), \text{ such that } p \sim_{\mathcal{KS}(\alpha)} \Psi_{\epsilon'}(t,\cdot,h_{t-1}) \tag{25}$$

Definition 11 (Optimal Significance Pair)**.** *Let us denote:*

$$\epsilon_\alpha \triangleq \sup_{\epsilon \in (0,\infty)} (\{\epsilon : [h_t]_\epsilon \text{ is } \alpha\text{-stable } \} \cup \{0\}) \tag{26}$$

Then, the optimal significance pair $(\alpha^\star, \epsilon^\star)$ w.r.t h_t is defined as:

$$\epsilon^\star = \sup_{\alpha \in [0,1]} \epsilon_\alpha \tag{27}$$

$$\alpha^\star = \inf_{\alpha \in [0,1]} \{\alpha^2 : \epsilon_\alpha \leqq \epsilon^\star\} \tag{28}$$

The optimal significance pair is unique by definition, and is clearly a function of the chosen metric/quasi-metric. Thus, it can be used to test how well our quasi-metric performs in identifying functionally similar histories.

Remark 1 (Ranking metrics via average optimal significance)**.** *If we have the option of choosing between different metrics or quasi-metrics, the average optimal significance achieved over the database may be used to rank the options, and choose the one achieving the higher average optimal value.*

The optimal significance pair dictates the confidence bounds on the House function estimate.

Lemma 2 (Optimal Confidence Bounds)**.** *If $(\alpha^\star, \epsilon^\star)$ is the optimal signifcance pair w.r.t. h_{t-1}, then the confidence bounds at significance level $\alpha^\star$ on $\Psi_{\epsilon^\star}(t,\cdot,h_{t-1})$ is given by:*

$$\forall d \in \mathbb{D}, \Psi_{\epsilon^\star}(t,d,h_{t-1}) \pm \frac{2k'}{\sqrt{|[h_{t-1}]_{\epsilon^\star}|}} \tag{29}$$

where we have:

$$\mathcal{K}(k') = 1 - \sqrt{\alpha^\star} \tag{30}$$

and $\mathcal{K}$ is the Kolmogorov distribution function:

$$\forall x \in \mathbb{R}, \mathcal{K}(x) = \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{\infty} e^{-\frac{1}{8}\left((2k-1)\frac{\pi}{x}\right)^2} \tag{31}$$

Proof: Since $(\alpha^\star, \epsilon^\star)$ is the optimal significance pair at h_{t-1}:

$$\exists \epsilon' < \epsilon, p \sim_{\mathcal{KS}(\sqrt{\alpha^\star})} \Psi_{\epsilon'}(t,\cdot,h_{t-1}) \tag{32}$$

Fix an admissible $\epsilon' < \epsilon$. Let us denote:

$$\frac{|[h_{t-1}]_{\epsilon'}|}{|[h_{t-1}]_\epsilon|} = r \leqq 1 \tag{33}$$

Now, from the formulation of the standard KS two sample test:

$$||p - \Psi_{\epsilon'}(t,\cdot,h_{t-1})||_\infty > \frac{k'}{\sqrt{r(1-r)\,|[h_{t-1}]_\epsilon|}} \text{ w.p. } \sqrt{\alpha^\star} \tag{34}$$

Let the true distribution we are attempting to infer is f. Then:

$$\begin{aligned} &||p - f||_\infty + ||f - \Psi_{\epsilon'}(t,\cdot,h_{t-1})||_\infty \\ &\geqq ||p - \Psi_{\epsilon'}(t,\cdot,h_{t-1})||_\infty > \frac{k'}{\sqrt{r(1-r)\,|[h_{t-1}]_\epsilon|}} \text{ w.p. } \sqrt{\alpha^\star} \end{aligned} \tag{35}$$

Now, it is obvious, that we have the mixture:

$$\Psi_\epsilon(t,\cdot,h_{t-1}) = r\Psi_{\epsilon'}(t,\cdot,h_{t-1}) + (1-r)p \tag{36}$$

Using Eq. (36), we get two bounds:

$$\begin{aligned} ||\Psi_\epsilon(t,\cdot,h_{t-1}) - f||_\infty \leqq &\frac{rk'}{\sqrt{r(1-r)\,|[h_{t-1}]_\epsilon|}} \\ &+ ||p - f||_\infty \text{ w.p. } 1 - \sqrt{\alpha^\star} \end{aligned} \tag{37a}$$

$$\begin{aligned} ||\Psi_\epsilon(t,\cdot,h_{t-1}) - f||_\infty \leqq &\frac{(1-r)k'}{\sqrt{r(1-r)\,|[h_{t-1}]_\epsilon|}} \\ &+ ||\Psi_{\epsilon'}(t,\cdot,h_{t-1}) - f||_\infty \text{ w.p. } 1 - \sqrt{\alpha^\star} \end{aligned} \tag{37b}$$

We can rewrite the above bounds as:

$$\begin{aligned} ||\Psi_\epsilon(t,\cdot,h_{t-1}) - f||_\infty > &\frac{rk'}{\sqrt{r(1-r)\,|[h_{t-1}]_\epsilon|}} \\ &+ ||p - f||_\infty \text{ w.p. } \sqrt{\alpha^\star} \end{aligned} \tag{38a}$$

$$\begin{aligned} ||\Psi_\epsilon(t,\cdot,h_{t-1}) - f||_\infty > &\frac{(1-r)k'}{\sqrt{r(1-r)\,|[h_{t-1}]_\epsilon|}} \\ &+ ||\Psi_{\epsilon'}(t,\cdot,h_{t-1}) - f||_\infty \text{ w.p. } \sqrt{\alpha^\star} \end{aligned} \tag{38b}$$

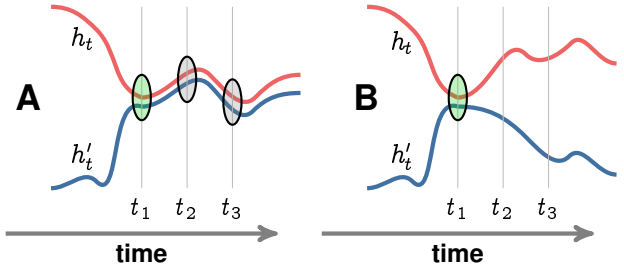

Fig. 4. Notion of causal states. Plate A illustrates the situation $h_t \sim h'_t$. Unlike plate B, here the histories remain close in future, inducing the notion that h_t, h'_t lead to a common causal state at time t.

Summing the last two inequalities, we get:

$$2\left\|\Psi_\epsilon(t,\cdot,h_{t-1})-f\right\|_\infty > \frac{k'}{\sqrt{r(1-r)\left|[h_{t-1}]_\epsilon\right|}} + \left\|p-f\right\|_\infty + \left\|f-\Psi_{\epsilon'}(t,\cdot,h_{t-1})\right\|_\infty \text{ w.p. } \sqrt{\alpha^\star} \tag{39}$$

Using Eq. (35), we get:

$$2\left\|\Psi_\epsilon(t,\cdot,h_{t-1})-f\right\|_\infty > \frac{2k'}{\sqrt{r(1-r)\left|[h_{t-1}]_\epsilon\right|}} \text{ w.p. } \alpha^\star \tag{40}$$

The proof is then completed by noting that the maximum value of $r(1-r)$ is $1/4$. □

Remark 2 (The Matching Problem)**.** *Matching refers to finding patient histories which are similar. Computation of the House function at least partially solves the matching problem, since, every history in $[h_t]_{\epsilon^\star}$ is matched to h_t. There is however the issue of demographics stratification, which we have ignored so far. However, this is a trivial issue, easily solved by partitioning the database a priori into demographic strata.*

5 Inducing The Notion of Causal States

We define a right invariant equivalence relation on the space of histories to induce the notion of states.

Definition 12 (Causal States)**.** *Two histories h_t, h'_t are equivalent (denoted as $h_t \sim h'_t$) if,*

$$\forall z = \{(t+1,d_1),\cdots,(t+i,d_i),\cdots,(t+m,d_m)\}, d_i \in D,$$
$$\forall d \in D, \Psi_\epsilon(t+m+1,d,h_t \cup z) = \Psi_\epsilon(t+m+1,d,h'_t \cup z) \tag{41}$$

It follows from Defn. 12 that:

$$h_t \sim h'_t \Rightarrow h_t \cup z \sim h'_t \cup z \tag{42}$$

for any right extension z into future. The property of right invariant equivalence is what formalizes the notion of a dynamical state: once we arrive at a state, the future behavior no longer distinguishes between different paths that brought us there.

References

[1] I. Chattopadhyay and H. Lipson, *Abductive learning of quantized stochastic processes with probabilistic finite automata*, Philos Trans A, 371 (2013), p. 20110543.
[2] J. Gustavsson, *Metrization of quasi-metric spaces.*, Mathematica Scandinavica, 35 (1974), pp. 56–60.
[3] V. Schroeder, *Quasi-metric and Metric Spaces*, Conformal Geometry and Dynamics, 10 (2006), pp. 355–360.