A Non-Asymptotic Analysis of Mismatched Guesswork

Alexander Mariona*, Homa Esfahanizadeh*, Rafael G. L. D'Oliveira†, and Muriel Médard*
* Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139
† School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634
Emails: amariona@mit.edu, homaesf@mit.edu, rdolive@clemson.edu, medard@mit.edu

***Abstract*—The problem of mismatched guesswork considers the additional cost incurred by using a guessing function which is optimal for a distribution q when the random variable to be guessed is actually distributed according to a different distribution p. This problem has been well-studied from an asymptotic perspective, but there has been little work on quantifying the difference in guesswork between optimal and suboptimal strategies for a finite number of symbols. In this non-asymptotic regime, we consider a definition for mismatched guesswork which we show is equivalent to a variant of the Kendall tau permutation distance applied to optimal guessing functions for the mismatched distributions. We use this formulation to bound the cost of guesswork under mismatch given a bound on the total variation distance between the two distributions.**

I. Introduction

Let X be a random variable taking values in $\{1, \ldots, n\}$ according to the distribution p. Suppose that we would like to try and guess the realization of X by sequentially guessing possible values according to a guessing function G, which is some permutation on $\{1, \ldots, n\}$, stopping when we correctly guess the value of X. The guesswork associated with G is $\mathbb{E}[G(X)]$. This is the general setting of the theory of guesswork [1], [2].

The goal of this work is to characterize the difference in guesswork between the guessing function which is optimal for the generating distribution p and that which is optimal for a different distribution q. This problem is generally referred to as *guessing under mismatch*. We define the expected cost of guesswork under mismatch between p and q to be

$$\delta(p, q) = \min_{G_q} \mathop{\mathbb{E}}_{X \sim p}[G_q(X) - G_p(X)],$$

where G_p and G_q are optimal guessing functions for p and q respectively. Our approach is non-asymptotic and our methods are combinatorial: we study the cost of mismatch for a fixed, finite value of n and we analyze guessing functions as permutations. This viewpoint emphasizes the structure of the space of probability distributions in the context of guesswork. The main contributions of this work are (1) an equivalence between the cost of mismatch and a variant of the Kendall tau permutation distance [3], [4] and (2) a bound on the cost of mismatch given a bound on the total variation distance between the two distributions.

This work was supported by DARPA Grant HR00112120008.

A. Related Work

In line with the majority of the literature on guesswork, the problem of guessing under mismatch has generally been approached from an asymptotic perspective. The problems of mismatched guesswork, universal guessing, guessing subject to distortion, and other settings have been well-studied in that asymptotic framework. We describe here only a selection of this work, describing its goals and highlighting how our problem setting and results differ.

Arikan and Merhav considered the problem of guessing to within some distortion measure and propose asymptotically optimal guessing schemes for such a measure [5]. Sundaresan considered the problem of guessing over a distribution which is a member of a known family. To this end, they defined a notion of redundancy which quantifies the increase in the moments of guesswork when using a suboptimal guessing function [6]. Both of these works propose universal strategies for certain families of sources, where "universal" means that some quantity asymptotically approaches its natural limit, e.g., the guesswork growth exponent for a particular moment approaches its minimum over the family, or the normalized redundancy approaches zero. Salamatian et al. consider the asymptotic behavior of mismatched guesswork by means of large deviation principles [7], a theory which has found considerable applications in the context of guesswork [8]–[10]. They treat independent, identically distributed sources using the framework of tilted distributions developed in [10], [11] and derive an expression for limit of the average growth rate of the moments of mismatched guesswork.

Our work distinguishes itself from the above in two key ways. First and foremost, we are not interested in asymptotic equivalences or limiting behavior. We view guessing functions as permutations on a finite alphabet through a framework which highlights applications of combinatorial techniques to guesswork analysis. Second, the bounds and limits which are given in, e.g., [5], [6] and [7] involve the difference in the exponents of guesswork, whereas we consider the difference in guesswork directly. Their bounds are not generally applicable to our problem setting and cannot be directly compared.

The remainder of this paper is organized as follows. Section II establishes our definitions. In Section III we express the expected cost of guesswork under mismatch in terms of a variation of the Kendall tau permutation distance. We then use this formulation of guesswork in Section IV to bound the cost of mismatch in terms of the total variation distance between the distributions in question. Section V offers concluding remarks and notes potential directions for future work.

II. Preliminaries

In this section, we review and establish definitions for guesswork, permutations, and statistical metrics. All probability distributions are over a finite space of cardinality $n \in \mathbb{N}$, which we denote by $[n] = \{1, 2, \ldots, n\}$. We refer to bijections from $[n]$ to $[n]$ as either permutations or guessing functions. Explicitly, if $G : [n] \to [n]$ is a guessing function, then $G(i)$ is equal to the number of guesses used to guess element i. We refer to $\mathbb{E}_p[G] = \mathbb{E}_{X \sim p}[G(X)]$ as the guesswork of G under p. The operator $\circ$ denotes function composition (we limit its use to permutations).

A natural set of guessing functions to consider is that consisting of those which minimize the guesswork.

Definition 1. The set of *optimal guessing functions* for the distribution p is defined to be $\mathcal{G}_p = \arg\min_G \mathbb{E}_p[G]$. □

We denote by G_p an optimal guessing functions for p. If $|\mathcal{G}_p| = 1$, then we may unambiguously refer to *the* optimal guessing function for p. A simple observation first made by Massey [1] is that optimal guessing functions proceed in order of decreasing probability, i.e., $G_p(i) < G_p(j)$ if and only if $p_i \geq p_j$. It follows that there are multiple optimal guessing functions for p if and only if p assigns the same probability to more than one point.

One contribution of this paper is a connection between guesswork and permutations via a metric. A classical metric on permutations is the Kendall tau rank distance [3], which is equal to the minimum number of adjacent transpositions required to transform one permutation to another. An adjacent transposition τ is a permutation which exchanges two consecutive elements, i.e., $\tau(j) = j + 1$ and $\tau(j + 1) = j$, where $1 \leq j < n$, and $\tau(i) = i$ for all $i \notin \{j, j + 1\}$.

Definition 2. The *Kendall tau distance* between two permutations σ_1 and σ_2 is defined to be

$$K(\sigma_1, \sigma_2) = \sum_{\substack{(i,j): \\ \sigma_1(i) < \sigma_1(j)}} \mathbb{1}\{\sigma_2(i) > \sigma_2(j)\}. \qquad \square$$

We say that two permutations σ_1 and σ_2 differ by an adjacent transposition if there exists an adjacent transposition τ such that $\sigma_1 = \tau \circ \sigma_2$. The Kendall tau distance has been generalized in many different ways to account for different notions of distance based on the particular applications [4]. In Section III we introduce a version which we show describes the difference in guesswork between two guessing functions.

Our second main result connects the statistical distance between two distributions and the difference in optimal guesswork for those distributions. The statistical metric we consider is the total variation distance, which is proportional to the L^1 norm for distributions over finite spaces.

Definition 3. The *total variation distance* between two distributions p and q over $[n]$ is defined to be

$$\mathrm{d}_{\mathrm{TV}}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i|. \qquad \square$$

III. Guesswork and a Permutation Divergence

We begin by defining our notion of difference in guesswork.

Definition 4. Let G_1 and G_2 be guessing functions. The *expected cost* of G_2 over G_1 with respect to the distribution p is defined to be $\Delta_p(G_1, G_2) = \mathbb{E}_p[G_2 - G_1]$. □

Definition 5. The *expected cost of guesswork under mismatch* between the distributions p and q is defined to be

$$\delta(p, q) = \min_{G_q \in \mathcal{G}_q} \Delta_p(G_p, G_q),$$

where G_p is any optimal guessing function for p. □

The expected cost of guesswork under mismatch is defined by a minimum over $\mathcal{G}_q$ because it is not necessarily true that $\mathbb{E}_p[G_q] = \mathbb{E}_p[G'_q]$ for all $G_q, G'_q \in \mathcal{G}_q$ (see Appendix A). However, the minimization in Definition 5 need not be taken over $\mathcal{G}_p$ because, by definition, $\mathbb{E}_p[G_p] = \mathbb{E}_p[G'_p]$ for all $G_p, G'_p \in \mathcal{G}_p$. By taking this minimum, Definition 5 gives the "best-case" cost, but we emphasize that we do not formulate the expected cost of guesswork under mismatch in terms of particular guessing functions, but rather distributions. With respect to expected guesswork, we are indifferent to the choice among multiple optimal guessing functions.

We show that $\Delta_p(G_1, G_2)$ is equivalent to a version of the Kendall tau distance which weighs pairs of transposed elements by the difference in their probabilities under p.

Definition 6. The *probability-weighted Kendall tau signed divergence* between two permutations σ_1 and σ_2 with respect to a distribution p is defined to be

$$K_p(\sigma_1, \sigma_2) = \sum_{\substack{(i,j): \\ \sigma_1(i) < \sigma_1(j)}} (p_i - p_j) \cdot \mathbb{1}\{\sigma_2(i) > \sigma_2(j)\}. \qquad \square$$

Under this divergence, permutations which differ on elements of similar probability are considered close. Informally, the sign of the divergence is positive if, in total, the transposition from σ_1 to σ_2 moves elements with higher probability to later positions and elements with lower probability to earlier positions. This notion is formalized in the equivalence shown in Theorem 1. This quantity is not a metric, as it is signed and asymmetric, but these properties are desirable in this setting.

A useful result on this divergence is that it satisfies a kind of "triangle equality." This holds for any choice of three permutations, but we limit the setting of the following lemma for simplicity. Only this weaker version is necessary to prove Theorem 1, which itself implies the general result.

Lemma 1. *Let σ_1, σ_2, and σ_3 be permutations such that σ_2 and σ_3 differ by an adjacent transposition. Then, for all distributions p,*

$$K_p(\sigma_1, \sigma_3) = K_p(\sigma_1, \sigma_2) + K_p(\sigma_2, \sigma_3). \qquad \square$$

PROOF: Let τ be the transposition such that $\sigma_3 = \tau \circ \sigma_2$ and let x and y be the elements whose positions τ inverts, such that $\sigma_2(x) = \sigma_2(y) - 1$ and $\sigma_3(y) = \sigma_3(x) - 1$. Then, since $\sigma_2(z) = \sigma_3(z)$ for all $z \notin \{x, y\}$, we can write

$$\begin{aligned}
K_p(\sigma_1,\sigma_3) &= \sum_{\substack{(i,j):\\ \sigma_1(i)<\sigma_1(j)}} (p_i - p_j) \cdot \mathbb{1}\{\sigma_3(i) > \sigma_3(j)\} \\
&= \sum_{\substack{(i,j)\notin\{(x,y),(y,x)\}:\\ \sigma_1(i)<\sigma_1(j)}} (p_i - p_j) \cdot \mathbb{1}\{\sigma_2(i) > \sigma_2(j)\} \\
&\quad + (p_x - p_y) \cdot \mathbb{1}\{\sigma_1(x) < \sigma_1(y) \wedge \sigma_3(x) > \sigma_3(y)\} \\
&\quad + (p_y - p_x) \cdot \mathbb{1}\{\sigma_1(y) < \sigma_1(x) \wedge \sigma_3(y) > \sigma_3(x)\},
\end{aligned}$$

where $\wedge$ denotes the intersection of events. By assumption, $\sigma_3(x) > \sigma_3(y)$, so the term with a factor of $(p_y - p_x)$ is always zero, and we can write the term with a factor of $(p_x - p_y)$ as

$$\begin{aligned}
&(p_x - p_y) \cdot \mathbb{1}\{\sigma_1(x) < \sigma_1(y) \wedge \sigma_3(x) > \sigma_3(y)\} = \\
&(p_x - p_y) \cdot \mathbb{1}\{\sigma_1(x) < \sigma_1(y)\} = \\
&(p_x - p_y) + (p_x - p_y) \cdot \mathbb{1}\{\sigma_1(x) < \sigma_1(y) \wedge \sigma_2(x) > \sigma_2(y)\} \\
&\qquad + (p_y - p_x) \cdot \mathbb{1}\{\sigma_1(y) < \sigma_1(x) \wedge \sigma_2(y) > \sigma_2(x)\}.
\end{aligned}$$

The second term of the final expression is always zero since $\sigma_2(x) < \sigma_2(y)$ by assumption. The third term is zero if $\sigma_1(x) < \sigma_1(y)$ but cancels the constant $(p_x - p_y)$ otherwise. Finally, it is easy to verify that $K_p(\sigma_2, \sigma_3) = p_x - p_y$. Substituting,

$$\begin{aligned}
K_p(\sigma_1, \sigma_3) &= \sum_{\substack{(i,j):\\ \sigma_1(i)<\sigma_1(j)}} (p_i - p_j) \cdot \mathbb{1}\{\sigma_2(i) > \sigma_2(j)\} \\
&\qquad + K_p(\sigma_2, \sigma_3) \\
&= K_p(\sigma_1, \sigma_2) + K_p(\sigma_2, \sigma_3).
\end{aligned}$$ ■

The main result of this section is that the expected cost of guesswork under mismatch can be formulated in terms of the probability-weighted Kendall tau signed divergence.

Theorem 1. *If G_1 and G_2 are guessing functions and p is a distribution, then the expected cost of G_2 over G_1 with respect to p is equal to the probability-weighted Kendall tau signed divergence between G_1 and G_2 with respect to p, i.e., $\Delta_p(G_1, G_2) = K_p(G_1, G_2)$.* □

PROOF: Let σ be the permutation which takes G_1 to G_2, i.e., $G_2 = \sigma \circ G_1$. Since any permutation can be written as the composition of a finite sequence of adjacent transpositions, for some $M \in \mathbb{N}$ we can write $\sigma = \tau_M \circ \tau_{M-1} \circ \cdots \circ \tau_1$, where τ_i is an adjacent transposition for all $i \in [M]$. This gives a finite sequence of permutations $G_1 = \sigma_0, \sigma_1, \ldots, \sigma_M = G_2$ such that $\sigma_i = \tau_i \circ \sigma_{i-1}$. Let y_i and x_i be the *original* elements which τ_i moves one position up and down respectively. In particular, x_i and y_i are defined such that

$$\begin{aligned}
\sigma_i(x_i) &= \sigma_{i-1}(y_i) = \sigma_{i-1}(x_i) + 1, \\
\sigma_i(y_i) &= \sigma_{i-1}(x_i) = \sigma_{i-1}(y_i) - 1, \\
\sigma_i(z) &= \sigma_{i-1}(z), \quad \forall z \notin \{x_i, y_i\}.
\end{aligned}$$

We proceed by induction on M. If $M = 1$, then we can write $G_2 = \tau_1 \circ G_1$. Since τ_1 is an adjacent transposition which inverts the positions of x_1 and y_1 as defined above,

$$\begin{aligned}
\Delta_p(G_1, G_2) &= \sum_{k=1}^{n} p_k \cdot [G_2(k) - G_1(k)] \\
&= p_{x_1} \cdot [G_2(x_1) - G_1(x_1)] \\
&\quad + p_{y_1} \cdot [G_2(y_1) - G_1(y_1)] \\
&= p_{x_1} - p_{y_1} \\
&= \sum_{\substack{(i,j):\\ G_1(i)<G_1(j)}} (p_i - p_j) \cdot \mathbb{1}\{G_2(i) > G_2(j)\} \\
&= K_p(G_1, G_2).
\end{aligned}$$

Suppose then that $\Delta_p(G_1, G_2) = K_p(G_1, G_2)$ for all guessing functions which differ by at most M adjacent transpositions. Let G_1 and G_2 be two guessing functions which differ by $M + 1$ adjacent transpositions, i.e., $G_2 = \tau_{M+1} \circ \tau_M \circ \cdots \circ \tau_1 \circ G_1$. Let $\sigma = \tau_M \circ \cdots \circ \tau_1$. By definition,

$$\begin{aligned}
\Delta_p(G_1, G_2) &= \mathbb{E}_p[G_2 - G_1] \\
&= \mathbb{E}_p[G_2 - \sigma \circ G_1] + \mathbb{E}_p[\sigma \circ G_1 - G_1] \\
&= \Delta_p(\sigma \circ G_1, G_2) + \Delta_p(G_1, \sigma \circ G_1).
\end{aligned}$$

The permutations G_2 and $\sigma \circ G_1$ differ by a single adjacent transposition and $\sigma \circ G_1$ and G_1 differ by M adjacent transpositions. Thus, by strong induction,

$$\Delta_p(G_1, G_2) = K_p(\sigma \circ G_1, G_2) + K_p(G_1, \sigma \circ G_1).$$

Finally, Lemma 1 yields $\Delta_p(G_1, G_2) = K_p(G_1, G_2)$. ■

Note that this result requires that $K_p(G_1, G_2)$ be signed and asymmetric. We also have a stronger version of Lemma 1.

Corollary 1. *$K_p(\sigma_1, \sigma_3) = K_p(\sigma_1, \sigma_2) + K_p(\sigma_2, \sigma_3)$ for any permutations σ_1, σ_2, and σ_3 and for any distribution p.* □

PROOF: The desired result follows from Theorem 1 and linearity of expectation. ■

When considering distributions rather than particular guessing functions, the following corollary is more applicable.

Corollary 2. *For all distributions p and q, the expected cost of guesswork under mismatch between p and q is given by*

$$\delta(p, q) = \min_{G_q \in \mathcal{G}_q} K_p(G_p, G_q) = \sum_{\substack{(i,j):\\ p_i > p_j}} (p_i - p_j) \cdot \mathbb{1}\{q_i < q_j\},$$

with G_p being any optimal guessing function for p. □

PROOF: The first equality follows directly from Theorem 1. The second equality follows from that fact that if G_p is an optimal guessing function for p, then $G_p(i) < G_p(j)$ implies that $p_i \geq p_j$. This inequality can be made strict, as $p_i - p_j = 0$ if $p_i = p_j$. Similarly, $G_q(i) > G_q(j)$ implies that $q_i \leq q_j$. This inequality can also be made strict, since if G_q^* is the guessing function which achieves the minimum expected cost and if $q_i = q_j$, then $G_p(i) < G_p(j)$ implies that $G_q^*(i) < G_q^*(j)$. ■

IV. Guesswork and a Statistical Distance

The goal of this section is to bound $\delta(p,q)$ given a bound on $\mathrm{d}_{\mathrm{TV}}(p,q)$. This connection between guesswork under mismatch and statistical distance is achieved through a combinatorial argument concerning the summands of the form $(p_i - p_j)$ in the probability-weighted Kendall tau signed divergence. Representing each of these terms by the corresponding pair (i,j), we show how these pairs can be grouped into sets and how to bound the sum over each set. To simplify notation, we write $k \in (i,j)$ to denote that either $k = i$ or $k = j$.

Definition 7. A set $M \subset [n] \times [n]$ is a *set of disjoint pairs* if $i \neq j$ for all $(i,j) \in M$ and if, for all $k \in [n]$, there is at most one element $(i,j) \in M$ such that $k \in (i,j)$. □

Example 1. These are five sets of disjoint pairs for $n = 5$.

M_1	M_2	M_3	M_4	M_5
$(2,5)$	$(1,3)$	$(2,4)$	$(3,5)$	$(1,4)$
$(3,4)$	$(4,5)$	$(1,5)$	$(1,2)$	$(2,3)$

□

The following lemma formalizes how sets of disjoint pairs can be used to bound part of the expected cost of guesswork under mismatch given a bound on the total variation.

Lemma 2. *Let M be a set of disjoint pairs and let p and q be distributions such that $p_i \geq p_j$ and $q_i \leq q_j$ for all $(i,j) \in M$. If $\mathrm{d}_{\mathrm{TV}}(p,q) \leq \epsilon$, then $\sum_{(i,j)\in M}(p_i - p_j) \leq 2\epsilon$.* □

Proof: Suppose that $\sum_{(i,j)\in M}(p_i - p_j) > 2\epsilon$. Then,

$$\begin{aligned}
\mathrm{d}_{\mathrm{TV}}(p,q) &= \frac{1}{2}\sum_{k=1}^{n}|p_k - q_k| \\
&\geq \frac{1}{2}\sum_{(i,j)\in M}(|p_i - q_i| + |q_j - p_j|) \qquad (1)\\
&\geq \frac{1}{2}\sum_{(i,j)\in M}|p_i - p_j + q_j - q_i| \qquad (2)\\
&\geq \frac{1}{2}\sum_{(i,j)\in M}(p_i - p_j) \qquad (3)\\
&> \epsilon,
\end{aligned}$$

where (1) follows by dropping non-negative terms, (2) follows from the triangle inequality, and (3) follows from the assumption that $p_i \geq p_j$ and $q_i \leq q_j$ for all $(i,j) \in M$. ■

Informally, Lemma 2 says that if G_q permutes a set of disjoint pairs relative to G_p, then the sum over the terms $(p_i - p_j)$ corresponding to each permuted pair (i,j) can be bounded in terms of the total variation between p and q.

Lemma 2 and some established results in combinatorics directly lead to Theorem 2. For simplicity, the proof given here only considers the case when the number of symbols n is even. The proof for odd n involves some additional technical details which are deferred to Appendix B.

Theorem 2. *If p and q are distributions on $[n]$ such that $\mathrm{d}_{\mathrm{TV}}(p,q) \leq \epsilon$, then $\delta(p,q) \leq 2(n-1)\epsilon$.* □

Proof: By Corollary 2,

$$\delta(p,q) = \sum_{\substack{(i,j):\\ p_i > p_j}} (p_i - p_j)\cdot \mathbb{1}\{q_i < q_j\}. \qquad (4)$$

Without loss of generality, we may assume that the elements of $[n]$ are labeled such that $p_i \geq p_j$ for all $i < j$. Each term in (4) can be associated with the pair (i,j), yielding $\binom{n}{2}$ pairs.

If n is even, then we can construct a tournament design on n, i.e., all $\binom{n}{2}$ possible pairs of distinct elements of $[n]$ can be arranged into a $\frac{n}{2} \times (n-1)$ array such that every element is contained in precisely one cell of each column [12, §51.1]. This is equivalent to constructing $n-1$ sets of disjoint pairs $M_1, \ldots, M_{n-1}$ whose union covers all $\binom{n}{2}$ possible pairs. Given such a collection of sets, we can write

$$\begin{aligned}
\delta(p,q) &\leq \sum_{i<j}(p_i - p_j) \qquad (5)\\
&= \sum_{r=1}^{n-1}\sum_{(i,j)\in M_r}(p_i - p_j). \qquad (6)
\end{aligned}$$

Equality is achieved in (5) when $q_i < q_j$ for all $i < j$. Combining this with the assumption that $\mathrm{d}_{\mathrm{TV}}(p,q) \leq \epsilon$, we can bound (6) using Lemma 2, which yields

$$\delta(p,q) \leq \sum_{r=1}^{n-1} 2\epsilon = 2(n-1)\epsilon. \qquad ■$$

The bound in Theorem 2 is not tight for all values of ϵ. In particular, since $\delta(p,q) \leq n-1$ trivially, the bound is not meaningful for $\epsilon > 1/2$. However, the following example shows that there exist distributions p and q for which $\delta(p,q)$ is arbitrarily close to $2(n-1)\epsilon$ for $\epsilon \leq 1/n$.

Example 2. Let p be the distribution given by

$$p_i = \begin{cases} 1/n + \gamma\epsilon & i = 1,\\ 1/n & 1 < i < n,\\ 1/n - \gamma\epsilon & i = n,\end{cases}$$

with $0 < \gamma < 1$ and $\epsilon \leq 1/n$. Let q be a distribution such that $q_i < q_j$ for all $i < j$ and $\mathrm{d}_{\mathrm{TV}}(p,q) \leq \epsilon$. Such a distribution q always exists since $\gamma < 1$; it will be very close to uniform distribution, but without any two elements having exactly the same probabilities. By Corollary 2,

$$\begin{aligned}
\delta(p,q) &= \sum_{\substack{(i,j):\\ p_i > p_j}} (p_i - p_j)\cdot \mathbb{1}\{q_i < q_j\}\\
&= \sum_{i=2}^{n}(p_1 - p_i) + \sum_{j=2}^{n}(p_j - p_n)\\
&= (n-1)(p_1 - p_n)\\
&= \gamma \cdot 2(n-1)\epsilon.
\end{aligned}$$

Thus, this choice of p, q comes within a factor of γ (which can be arbitrarily close to 1) of the bound in Theorem 2. □

This example suggests that $2(n-1)\epsilon$ is the supremum, rather than the maximum, of $\delta(p,q)$ over distributions on n

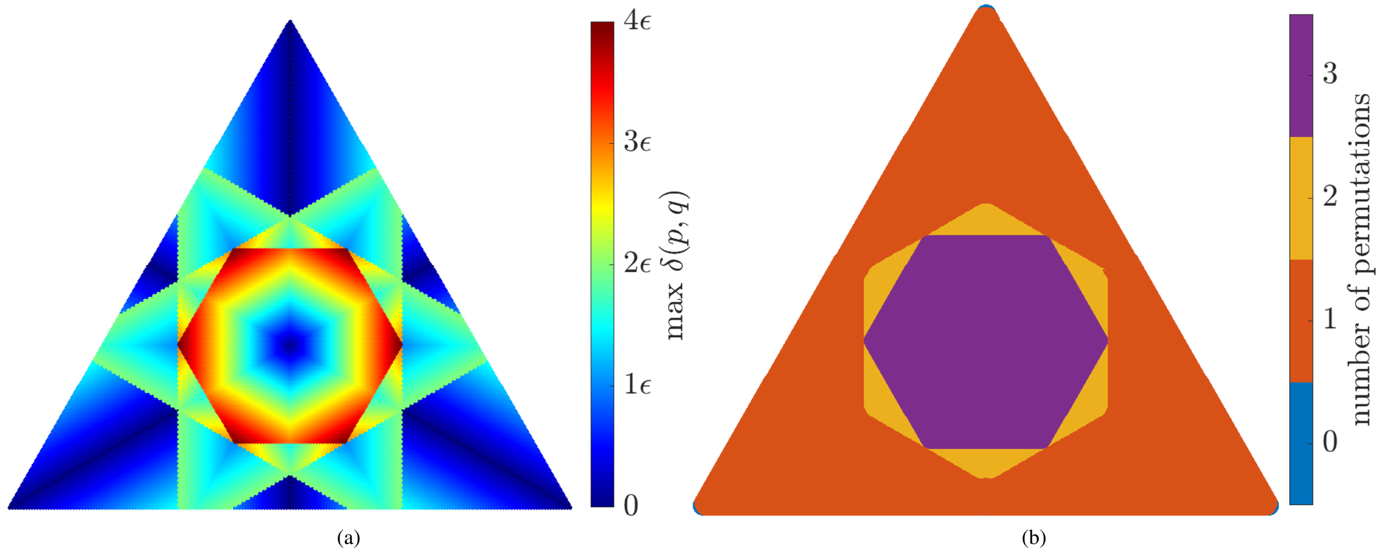

Fig. 1. A visualization of the expected cost of guesswork under mismatch on the 3-dimensional probability simplex. Each point on the simplex corresponds to a different distribution over 3 symbols. The corners correspond to distributions which assign all probability to a single symbol. The centroid corresponds to the uniform distribution. (a) The maximum expected cost of guesswork under mismatch given that the mismatched distribution lies within a total variation radius of $\epsilon = 0.2$. (b) The maximum number of adjacent transpositions between the optimal guessing function for the distribution at each point and the optimal guessing functions for points within a total variation radius of $\epsilon = 0.2$.

symbols such that $\mathrm{d_{TV}}(p,q) \leq \epsilon$ in the regime $\epsilon \leq 1/n$. The technique used for Theorem 2 does not give a strict inequality, however, since it does not take into account that there are multiple optimal guessing functions for distributions which assign the same probability to multiple points. In particular, if we consider the distributions given in Example 2 and let $\gamma = 1$, then the uniform distribution satisfies the total variation bound. However, since any guessing function is optimal for the uniform distribution, it follows that $\delta(p,q) = 0$ if q is uniform. Furthermore, if $\gamma = 1$, then there does not exist a q within ϵ total variation of p such that $q_i < q_j$ for all $i < j$, and hence, there does not exist any q within ϵ total variation of p such that $\delta(p,q) = 2(n-1)\epsilon$.

To visualize this phenomenon and conceptualize how the cost of guesswork under mismatch within a fixed total variation radius varies, consider Figure 1, which illustrates the behavior for $n = 3$. In Figure 1a, we plot the maximum achievable expected cost under mismatch within a radius of $\epsilon = 0.2$ around each point on the simplex. Every such point corresponds to a unique distribution on $[n]$, and that point is taken to be the center around which we consider distributions for mismatch within the radius of ϵ. In Figure 1b, we plot the maximum number of adjacent transpositions achievable between optimal guessing functions for the distributions within the same radius. In three dimensions, the projection of the L^1 ball onto the simplex is a regular hexagon, which gives rise to the geometric patterns visible in Fig. 1.

The distribution p given in Example 2 corresponds to the most red points in Fig. 1a. It is easy to see that, for $n = 3$, there are sharp jumps in the maximum achievable expected cost under mismatch as the maximum achievable number of adjacent transpositions increases. A cost of 2ϵ is the maximum achievable with 1 transposition, while 3ϵ is the maximum with 2, and 4ϵ with 3. Although this phenomenon does appear to some degree in higher dimensions, it is not necessarily true that the maximum achievable expected cost under mismatch is strictly increasing in the maximum number of achievable adjacent transpositions. In Example 2, for instance, only $2n-3$ adjacent transpositions are needed to achieve an expected cost which is arbitrarily close to the maximum, while $n(n-1)/2$ adjacent transpositions are possible.

V. Conclusions and Future Work

In this paper, we analyzed the cost of guesswork under mismatch from a non-asymptotic perspective. We showed that the total variation distance exactly captures the maximum possible expected cost under mismatch between two distributions if the number of symbols is considered constant. The connections developed between permutation metrics and guesswork highlight the combinatorial aspects of the problem and readily suggest possible extensions.

We notably only investigated the first moment of guesswork, which facilitated the connection with the Kendall tau distance. A natural next step would be to consider higher moments and formulate a more general framework for combinatorial analysis. Similarly, it would be interesting to consider how different statistical metrics or simplex geometries, such as those discussed in [13], can be related to guesswork under mismatch.

References

[1] J. L. Massey, "Guessing and entropy," in *Proceedings of IEEE International Symposium on Information Theory*, 1994, p. 204.

[2] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[3] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[4] R. Kumar and S. Vassilvitskii, "Generalized distances between rankings," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 571–580.

[5] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1041–1056, 1998.

[6] R. Sundaresan, "Guessing under source uncertainty," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 269–287, 2006.

[7] S. Salamatian, L. Liu, A. Beirami, and M. Médard, "Mismatched guesswork and one-to-one codes," in *IEEE Information Theory Workshop*, 2019, pp. 1–5.

[8] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 70–78, 2010.

[9] M. M. Christiansen and K. R. Duffy, "Guesswork, large deviations, and Shannon entropy," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 796–802, 2012.

[10] A. Beirami, R. Calderbank, M. M. Christiansen, K. R. Duffy, and M. Médard, "A characterization of guesswork on swiftly tilting curves," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2850–2871, 2018.

[11] A. Beirami, R. Calderbank, M. Christiansen, K. Duffy, A. Makhdoumi, and M. Médard, "A geometric perspective on guesswork," in *53rd Annual Allerton Conference on Communication, Control, and Computing*, 2015, pp. 941–948.

[12] C. J. Colbourn and J. H. Dinitz, *Handbook of Combinatorial Designs*. CRC Press, 2010.

[13] F. Nielsen and K. Sun, "Clustering in Hilbert's projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices," in *Geometric Structures of Information*. Springer, 2019, pp. 297–331.

Appendix A
An Example Concerning Definition 5

Example 3. Let $n = 4$ and p, q be distributions given by:

	1	2	3	4
p	0.40	0.30	0.20	0.10
q	0.60	0.20	0.10	0.10

The distribution p has a unique optimal guessing function:

	1	2	3	4
G_p	1	2	3	4

The distribution q has two optimal guessing functions:

	1	2	3	4
G_q	1	2	3	4
G'_q	1	2	4	3

In this case, $\Delta_p(G_p, G_q) = 0$ and $\Delta_p(G_p, G'_q) = 0.1$. Hence, $\delta(p, q) = 0$, due to the minimum in Definition 5. □

Appendix B
Proof of Theorem 2 for odd n

The setup for handling distributions over an odd number of symbols is largely the same as for an even number, except for the fact that we instead constructing an odd tournament design, yielding an $(n-1)/2 \times n$ array of disjoint pairs. This would only give a bound of $2n\epsilon$ if we were to apply Lemma 2.

The necessary additional detail is that, if n is odd, for any set of disjoint pairs there must exist at least one element of $[n]$ which does not appear in any pair. In this setting, we can consider *pairs of sets* which leave out different symbols.

Definition 8. Let n be odd and let M_1 and M_2 be sets of disjoint pairs such that $|M_1| = |M_2| = (n-1)/2$ and there exist distinct, unique elements $k_1, k_2 \in [n]$ such that $k_1 \notin m_1$ for all $m_1 \in M_1$ and $k_2 \notin m_2$ for all $m_2 \in M_2$. The pair (k_1, k_2) is defined to be the *bridge pair* for M_1 and M_2. □

Example 4. For the sets listed in Example 1, the pair $(2, 5)$ is the bridge pair for M_2 and M_5, while $(3, 4)$ is the bridge pair for M_3 and M_4. Indeed, the sets are labeled such that (i, j) is the bridge pair for M_i and M_j. □

Note that it does not make sense to consider Definition 8 when n is even, if the sets in question are not of maximal size, or if the sets in question do not leave out distinct elements. When a bridge pair does exist, however, it is by definition unique.

The following lemma is an analogue of Lemma 2 which accounts for bridge pairs.

Lemma 3. *Let n be odd and let M_1 and M_2 be sets of disjoint pairs for which there exists a bridge pair (k_1, k_2). Let $M = M_1 \cup M_2 \cup \{(k_1, k_2)\}$. Let p and q be distributions such that $p_i \geq p_j$ and $q_i \leq q_j$ for all $(i, j) \in M$. If $\mathrm{d_{TV}}(p, q) \leq \epsilon$, then $\sum_{(i,j)\in M}(p_i - p_j) \leq 4\epsilon$.* □

Proof: Suppose that $\sum_{(i,j)\in M}(p_i - p_j) > 4\epsilon$. Then,

$$\begin{aligned}
4\,\mathrm{d_{TV}}(p, q) &= \sum_{i=1}^{n} |p_i - q_i| + \sum_{i=1}^{n} |p_i - q_i| \\
&\geq \sum_{(i,j)\in M_1} |p_i - p_j + q_j - q_i| \\
&\quad + \sum_{(i,j)\in M_2} |p_i - p_j + q_j - q_i| \qquad (7)\\
&\quad + |p_{k_1} - p_{k_2} + q_{k_2} - q_{k_1}| \\
&\geq \sum_{(i,j)\in M} (p_i - p_j) \qquad (8)\\
&> 4\epsilon,
\end{aligned}$$

where (7) follows from the triangle inequality and the fact that, by the definition of a bridge pair, each element of $[n]$ appears twice amongst all the pairs in M, while (8) follows from the assumption that $p_i \geq p_j$ and $q_i \leq q_j$ for all $(i, j) \in M$. ■

To prove Theorem 2 for odd n, we use a similar construction of sets of disjoint pairs based on a tournament design. For odd n, we can treat one of the sets of disjoint pairs as a set of bridge pairs and save the extra factor of n which arises in an odd tournament design by applying Lemma 3 in place of Lemma 2.

Proof (of Theorem 2, cont.): Recall that we may assume without loss of generality that $p_i \geq p_j$ for all $i < j$. If n is odd, then we can construct an odd tournament design, i.e., all $\binom{n}{2}$ possible pairs can be arranged into an $\frac{n-1}{2} \times n$ array such that every element is contained in at most one cell of each column and there is exactly one element missing from each column [12, §51.1]. This is equivalent to constructing n sets of disjoint pairs $M_1, \ldots, M_n$ such that there exists a distinct bridge pair between any two of them.

In particular, if we label the sets $M_1, \ldots, M_n$ according to the elements which are missing from each set, then each pair $(i, j) \in M_1$ is a bridge pair for M_i and M_j. It is also follows from the properties of the odd tournament design that no two elements of M_1 will be bridge pairs for the same sets. Let $M_{i,j} = M_i \cup M_j \cup \{(i, j)\}$. Then,

$$\begin{aligned}
\delta(p, q) &\leq \sum_{i<j} (p_i - p_j) \qquad (9)\\
&= \sum_{(i,j)\in M_1} \sum_{(k,l)\in M_{(i,j)}} (p_k - p_l) \qquad (10)\\
&\leq \sum_{(i,j)\in M_1} 4\epsilon \qquad (11)\\
&= 2(n-1)\epsilon,
\end{aligned}$$

where (10) follows from the bridge pair construction and (11) follows from Lemma 3, using the fact that equality in (9) is achieved when $q_i < q_j$ for all $i < j$. ■