Robust Multi-Hypothesis Testing with Moment-Constrained Uncertainty Sets

Akshayaa Magesh*, Zhongchang Sun†, Venugopal V. Veeravalli*, and Shaofeng Zou†

*Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA, {amagesh2, vvv}@illinois.edu
†Department of Electrical Engineering, University at Buffalo, the State University of New York, Buffalo, NY, USA, {zhongcha,szou3}@buffalo.edu

Abstract—The problem of robust multi-hypothesis testing in the Bayesian setting is studied in this paper. Under the $m \geq 2$ hypotheses, the data-generating distributions are assumed to belong to uncertainty sets constructed through some moment functions, i.e., the sets contain distributions whose moments are centered around empirical moments obtained from some training data sequences. The goal is to design a test that performs well under all distributions in the uncertainty sets, i.e., a test that minimizes the worst-case probability of error over the uncertainty sets. Insights on the need for optimization-based approaches to solve the robust testing problem with moment constrained uncertainty sets are provided. The optimal (robust) test based on the optimization approach is derived for the case where the observations belong to a finite-alphabet. When the size of the alphabet is infinite, the optimization problem is infinite-dimensional and intractable, and therefore a tractable finite-dimensional approximation is proposed, whose optimal value converges to the optimal value of the original problem as the size of the dimension of the approximation goes to infinity. A robust test is constructed from the solution to the approximate problem, and guarantees on its worst-case error probability over the uncertainty sets are provided. Numerical results are provided to demonstrate the performance of the proposed robust test.

I. INTRODUCTION

Hypothesis testing is a fundamental problem in statistical decision-making, in which the goal is to decide between given hypotheses based on observed data. In the multi-hypothesis testing problem ($m \geq 2$ hypotheses), the likelihood ratios of the $m$ hypotheses with respect to the first one is used to construct optimal tests under various criteria (see, e.g., [1]).

In general, the distributions under the hypotheses may be unknown, and may need to be estimated from historical data. However, deviations of the estimates from the true underlying distributions can result in significant performance degradation of the likelihood ratio test constructed using the estimated distributions. The robust hypothesis testing framework was introduced in Huber’s seminal work [2] to address this problem. In the robust setting, it is assumed that under each hypothesis, the distributions belong to certain uncertainty sets, and the goal is to build a detector that performs well under all distributions in the uncertainty sets. The uncertainty sets are generally constructed as collections of distributions that lie within some neighbourhood (for instance, with respect to some discrepancy measure) of certain nominal distributions.

The uncertainty sets in [2] were constructed as epsilon-contamination sets, i.e., they contain mixture distributions where the nominal distribution is corrupted by some unknown distribution. A censored likelihood ratio test was proposed and proved to be minmax optimal, i.e., the test minimizes the worst-case error probability over the uncertainty sets. In [2], Huber additionally characterized a sufficient condition that needs to be satisfied by the worst-case distributions for general uncertainty classes in the binary hypothesis testing setting, which was later formalized as the Joint Stochastic Boundedness (JSB) condition [3] (see also [1]). This condition is quite powerful, as the optimal minmax test for a single sample or independent batch samples can be directly constructed from the worst-case distributions satisfying the JSB condition. However, except for certain structured uncertainty classes such as epsilon-contamination classes, Total Variation (TV) constrained classes, classes with alternating capacities of order 2, and bounded density variations, it is quite difficult to verify the JSB condition. Alternate optimization based approaches were therefore explored in [4]–[6] for uncertainty classes constructed using f-divergences and adversarial perturbations.

Recent works have studied the problem of constructing uncertainty sets using a data-driven approach [7]–[10], where the nominal distributions are the empirical distributions derived from training observations. In [7], [8], the Wasserstein distance was used to construct the uncertainty sets for a minmax robust Bayesian hypothesis testing problem. The main result in [8] states that for the Wasserstein uncertainty sets, there exists a pair of worst-case distributions supported on the empirical samples. The work in [9] studied the data-driven robust hypothesis testing problem with the uncertainty sets constructed using the Sinkhorn distance, where a test for minimizing the worst-case loss using an approximated smoothed error probability was proposed. In [10], the Maximum Mean Discrepancy (MMD) was used to construct the uncertainty sets. In the Bayesian setting, a tractable approximation to the minmax problem was proposed, and in the Neyman-Pearson setting, an asymptotically optimal test was proposed. The performances of the proposed tests in [8]–[10] were evaluated empirically, and theoretical guarantees on the worst-case error over the distributions in the uncertainty sets were absent. All the above-mentioned works [2], [4]–[18] consider only the binary hypothesis testing problem.
A. Our Contributions

We study the problem of robust multi-hypothesis testing with general moment-constrained uncertainty sets, i.e., the sets contain distributions whose moments are centered around empirical moments. Moment constrained sets are practical since it is generally computationally easier to calculate empirical moments, and the convergence rates of empirical moments to the true moments are faster than that of the Wasserstein or MMD distances between the empirical distributions and the underlying true distributions. Our focus is on the minmax robust formulation in the Bayesian setting. To the best of our knowledge, ours is the first work to address robust multi-hypothesis testing for more than two hypotheses. In our prior work [19], only the binary hypothesis setting was considered, and some preliminary results with weaker theoretical guarantees were presented.

- In the basic case of binary hypotheses \( m = 2 \), we provide insights on the need for optimization based approaches for the moment constrained robust hypothesis testing problem by showing that the moment-constrained uncertainty sets may not satisfy the JSB condition.
- We characterize the minmax optimal robust test for the case when the distributions under the \( m \) hypotheses are supported on a finite alphabet set \( \mathcal{X} \).
- In the case when the size of \( \mathcal{X} \) is infinite, we provide a tractable approximation of the worst-case error that converges to the optimal minmax Bayes error, and propose a robust test. We provide robust guarantees on the worst-case error of the proposed test over all distributions in the uncertainty sets. To the best of our knowledge, ours is the first work to provide such guarantees.
- We provide numerical results to demonstrate the performance of our proposed algorithms.

II. Problem Setup

Let the random variable \( X \) denote a single observation, and let \( \mathcal{X} \subset \mathbb{R}^d \) be a compact set denoting the corresponding sample space, where \( d \) is the dimension of the data. Let \( \mathcal{P} \) denote the set of all Borel probability measures on \( \mathcal{X} \). Let \( \mathcal{P}_1, \ldots, \mathcal{P}_m \subset \mathcal{P} \) denote the uncertainty sets under the \( m \) hypotheses, respectively. We construct the uncertainty sets using general moment constraints derived from observations from the hypotheses. In this paper, we use the notation \([n]\) to denote the set \( \{1, \ldots, n\} \) for any \( n \in \mathbb{N} \). Let \( \hat{x}_{1,i}, \ldots, \hat{x}_{n,i} \) for \( i \in [m] \) denote the realizations of training sequences under the \( m \) hypotheses. Let \( Q_1, \ldots, Q_m \) denote the empirical distributions corresponding to the training observations, i.e.,

\[
Q_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \delta(\hat{x}_{i,j}),
\]

where \( \delta(x) \) is the Dirac measure on \( x \). We use the empirical distributions as the nominal distributions in the construction of the uncertainty sets. Let \( \psi_k : \mathcal{X} \to \mathbb{R} ; \ k \in [K] \), denote \( K \) real-valued, continuous functions defined on the sample space. The uncertainty sets for \( i \in [m] \) are defined as follows:

\[
P_i^n = \left\{ P \in \mathcal{P} : \left| E_P[\psi_k(X)] - E_{Q_i}[\psi_k(X)] \right| \leq \eta, \ k \in [K] \right\}
\]

where \( \eta \) is a pre-specified radius of the uncertainty sets, and \( E_P[\cdot] \) denotes the expectation under the distribution \( P \). Note that we can choose the same \( \eta \) for all the constraints, since otherwise the moment functions \( \psi_k \)'s can be scaled appropriately. Let \( \eta_{\text{max}} \) denote the maximum radius so that the uncertainty sets do not overlap. Thus, it is assumed that \( \eta \in (0, \eta_{\text{max}}] \), otherwise, the problem becomes ill-defined. The \( m \) hypotheses with uncertainty sets \( P_{i,1}^n, \ldots, P_{i,m}^n \) are defined as follows:

\[
H_1 : X \sim P_1, \quad P_1 \in \mathcal{P}_1^n,
\]

\[
\vdots
\]

\[
H_m : X \sim P_m, \quad P_m \in \mathcal{P}_m^n. \tag{2}
\]

A (randomized) decision rule is defined using \( \phi(x) = (\phi_1(x), \ldots, \phi_m(x)) \), a measurable function from \( \mathcal{X} \) to the \( m \)-dimensional probability simplex \( \Delta_m \), with the interpretation that the test accepts hypothesis \( i \) with probability \( \phi_i(x) \) (see Chapter 1 of [20]). Note that \( \phi_i(x) \geq 0 \) and \( \sum_{i=1}^{m} \phi_i(x) = 1 \) for all \( x \in \mathcal{X} \). We assume uniform costs, i.e., the cost of the test declaring \( H_j \) when \( H_i \) is true is \( 1(i \neq j) \). The conditional risk \( R_i(\phi) \) associated with hypothesis \( i \), i.e., the probability of the test rejecting \( H_i \) when \( H_i \) is true, is given by

\[
R_i(\phi) = \int_{\mathcal{X}} (1 - \phi_i(x)) dP_i(x). \tag{3}
\]

In the Bayesian setting with equal priors, the probability of error of the test is given by:

\[
P_E(\phi; P_1, \ldots, P_m) = \sup_{\phi \in \mathcal{P}_1^n \cup \cdots \cup \mathcal{P}_m^n} \inf_{\phi \in \mathcal{P}_1^n \cup \cdots \cup \mathcal{P}_m^n} P_E(\phi; P_1, \ldots, P_m). \tag{5}
\]

In order to minimize the Bayes probability of error in (4) for fixed \( P_1, \ldots, P_m \), it is not necessary to consider randomized decision rules (see Section 2 of [21] for a detailed explanation).

Remark 1: The methods proposed in this paper and their analyses can be easily extended to the case with unequal priors.

III. Insights from the Binary Case

Consider the robust binary hypothesis setting, for which \( m = 2 \). The minmax robust hypothesis testing problem can equivalently be written as:

\[
\inf_{\phi} \sup_{P_0 \in \mathcal{P}_0^n, P_1 \in \mathcal{P}_1^n} P_E(\phi; P_0, P_1), \tag{6}
\]

where

\[
P_E(\phi; P_0, P_1) \triangleq \frac{1}{2} E_{P_0}[\phi(x)] + \frac{1}{2} E_{P_1}[1 - \phi(x)], \tag{7}
\]

and a test \( \phi \) : \( \mathcal{X} \to [0, 1] \) accepts \( H_0 \) with probability \( 1 - \phi(x) \), and accepts \( H_1 \) with probability \( \phi(x) \). Only for the binary case, we index the hypotheses by \( i = 0, 1 \) rather than \( i = 1, 2 \).
to be consistent with literature on binary hypothesis testing. An important result in this setting from Huber’s seminal work in [2] is the characterization of sufficient conditions for a pair of least favorable distributions for the construction of a saddle point, later formalized in [3] as the Joint Stochastic Boundedness (JSB) condition. This condition states that, a pair of uncertainty classes $\mathcal{P}_0$ and $\mathcal{P}_1$ is jointly stochastically bouded by $(Q_0, Q_1)$, if there exist distributions $Q_0 \in \mathcal{P}_0$ and $Q_1 \in \mathcal{P}_1$, such that for any distributions $P_0 \in \mathcal{P}_0$ and $P_1 \in \mathcal{P}_1$, and all $t \in \mathbb{R}$,

$$P_0 \{ \ln L_q(x) > t \} \leq Q_0 \{ \ln L_q(x) > t \}, \quad (8)$$

and

$$P_1 \{ \ln L_q(x) > t \} \geq Q_1 \{ \ln L_q(x) > t \}, \quad (9)$$

where $L_q(x) = \frac{q(x)}{\tilde{q}(x)}$ is the likelihood ratio between $Q_1$ and $Q_0$. Intuitively, this condition leads to a pair of distributions from the uncertainty sets that are closest to each other for the purpose of hypothesis testing.

Except for a few special cases of the uncertainty classes such as the epsilon-contamination classes and total variation constrained classes (see [2]), it is difficult to ascertain the existence of distributions satisfying the JSB condition. Most recent works in the robust hypothesis testing literature employ optimization-based approaches, and do not even discuss the JSB property [5], [8]–[10]. For moment-constrained uncertainty classes, we provide some insights on the existence of distributions satisfying the JSB property. In particular, we provide a counterexample to show that the JSB property may not be satisfied in a simple finite alphabet setting, where $\mathcal{P}_0$ contains a single distribution, and $\mathcal{P}_1$ contains distributions satisfying a single first-order moment constraint.

Claim 1: Consider the alphabet $\mathcal{X} = \{0, 1, 2\}$. Let $\mathcal{P}_0$ contain a single distribution $P_0 = [0.45, 0.25, 0.3]$. Let the uncertainty set $\mathcal{P}_1$ be defined as:

$$\mathcal{P}_1 = \{ P \in \mathcal{P} : E_P[\mathcal{X}] > 0.9 \}. \quad (10)$$

This pair of uncertainty sets defined on $\mathcal{X}$ does not satisfy the JSB condition in (8) and (9).

The proof of the claim is provided in [21].

IV. ROBUST MULTI-HYPOTHESIS TESTING

First, we present a result to establish the existence of a saddle point for our minmax problem. Then, we characterize the minmax optimal test in the case where the alphabet is finite. In the infinite-alphabet case, for which the optimization problem is infinite dimensional, we present a tractable finite-dimensional approximation, and bound the error due to the approximation. Henceforth, we denote $E[\psi_k(X)]$ by $E[\psi_k]$ for ease of notation.

Theorem 1: The minmax problem in (5) has a saddle-point solution $(\phi^*; P_{1,N}^{*}, \ldots, P_{m,N}^{*})$, i.e.,

$$\inf_{\phi} \sup_{P_1 \ldots P_m} E_\phi (P_1 \ldots P_m)$$

$$= \sup_{P_1 \ldots P_m} \inf_{\phi} E_\phi (P_1 \ldots P_m). \quad (11)$$

The proof of the above theorem is provided in [21].

A. Finite Alphabet: Optimal Test

Consider a finite alphabet $\mathcal{X} = \{z_1, \ldots, z_N\}$. Let $P_{i,N}$ denote a probability mass function on $\mathcal{X}$. The randomized decision rule for a finite alphabet $\mathcal{X}$ can be written as $\phi_N = [\phi_1, \ldots, \phi_m]$, where $\phi_i = [\phi_{i,N}(z_1), \ldots, \phi_{i,N}(z_N)]$ and $\sum_{i=1}^{m} \phi_{i,N}(z_j) = 1$ for $j \in [N]$. Then the minmax problem in (5) can be written as:

$$\min_{\phi_{i,N} \in [0,1]^N} \max_{i \in [m]} \{ 1 - \frac{1}{m} \sum_{i=1}^{m} \phi_{i,N} P_{i,N} \}$$

s.t.

$$\sum_{j=1}^{N} \phi_{i,N}(z_j) \psi_k(z_j) - E_{Q_i}[\psi_k] \leq \eta, \quad k \in [K]$$

$$P_{i,N}^1 = 1, \quad i \in [m], \quad \sum_{i=1}^{m} \phi_{i,N}(z_j) = 1, \quad j \in [N], \quad (12)$$

where $P_{i,N} = [p_{i,N}(z_1), \ldots, p_{i,N}(z_N)]^T$ for $i \in [m]$ and $1$ is the vector of all ones. Using the dual formulation for the inner maximization, we can reformulate the minmax problem into a tractable minimization problem.

Theorem 2: The minmax optimization problem in (12) can be reformulated as:

$$\min_{\phi_{i,N} \in [0,1]^N} \max_{i \in [m]} \{ 1 - \sum_{k=1}^{K} \sum_{i=1}^{m} \lambda_i \phi_{i,N} \psi_k - \sum_{i=1}^{m} b_i \}$$

s.t.

$$\frac{1}{m} \phi_{i,N}(z_j) - \sum_{k=1}^{K} \lambda_i \psi_k \psi_k(z_j) + \sum_{k=1}^{K} \lambda_i \psi_k \psi_k(z_j) - \mu_i \geq 0, \quad j \in [N], \quad i \in [m], \quad \sum_{i=1}^{m} \phi_{i,N}(z_j) = 1 \quad j \in [N], \quad (13)$$

where $a_{i,k} = E_{Q_i}[\psi_k] - \eta, \quad b_{i,k} = E_{Q_i}[\psi_k] + \eta$.

The proof of the above theorem is provided in [21].

The test $\phi_N$ has to be in the form of a likelihood ratio test constructed using $P_{1,N}^{*}, \ldots, P_{m,N}^{*}$ in case of no tie breaks. We can obtain the corresponding saddle point worst-case distributions $(P_{1,N}^{*}, \ldots, P_{m,N}^{*})$ by substituting $\phi_N$ in the original minmax objective (12) with an additional constraint to satisfy the saddle point property $\inf_{\phi_N} E_\phi (P_{1,N}^{*}, \ldots, P_{m,N}^{*}) = \gamma_N$, and solving the corresponding maximization problem (see Section 4 of [21] for a detailed explanation).

B. Infinite Alphabet

In this section, we consider the case when $\mathcal{X}$ is infinite. In this case, the minmax formulation in (5) is in general an infinite-dimensional optimization problem, and closed form solutions are difficult to derive. We can characterize the dual for this problem with a form similar to the optimization problem in Theorem 2.
Theorem 3: The strong dual for the minmax robust problem in the infinite alphabet case is

\[
\inf_{\phi, \lambda^e, \lambda^n} \{ 1 - \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda^e_{i,k} a_{0,k} + \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda^n_{i,k} b_{0,k} - \sum_{i=1}^{m} \mu_i, \quad \text{s.t.} \quad \frac{1}{m} \phi_i(x) - \sum_{k=1}^{K} \lambda^e_{i,k} \psi_k(x) + \sum_{k=1}^{K} \lambda^n_{i,k} \psi_k(x) - \mu_i \geq 0, \\
\forall x \in \mathcal{X}, \quad i \in [m], \quad \sum_{i=1}^{m} \phi_i(x) = 1, \quad \forall x \in \mathcal{X}.
\]

(14)

The proof of the above theorem is provided in [21]. The dual formulation in (14) is an infinite dimensional optimization problem (in the variable \( \phi \)) with infinitely many constraints. We propose a tractable finite dimensional optimization problem as an approximation to the minmax problem. Recall that \( \mathcal{X} \) is a compact set in \( \mathbb{R}^d \). For the sake of simplicity and without loss of generality, we assume that \( \mathcal{X} \subseteq [0,1]^d \). First, note that the moment defining functions \( \psi_k, k \in [K] \) are continuous functions on a compact set. Thus, they are Lipschitz functions with constants \( L_1, \ldots, L_K \), and without loss of generality, we set \( L = \max L_k = 1 \), since the moment functions can be scaled appropriately. In addition, we consider values of \( \eta \in [\Delta, \eta_0] \), where \( \eta_0 = \eta_{\text{max}} \), and \( 0 < \Delta < \eta_0 \). Let \( \epsilon > 0 \) such that \( \eta + \epsilon \leq \eta_0 \). Consider a discretization of the space \( \mathcal{X} \) through an \( \epsilon \)-net or a covering set. Indeed, we can consider a simple and efficient construction by considering a grid of equally spaced \( N = \left[ \frac{1}{\epsilon} \right] \) points \( S_N = \{z_1, \ldots, z_N\} \) such that for any \( x \in \mathcal{X}, \min_{j \in [N]} \|z_j - x\| \leq \epsilon \).

Here, \( N \) depends on \( \epsilon \), but we drop the dependence on \( \epsilon \) in the notation for \( N \) for readability. Let \( P_N \) denote all the distributions that are supported on the set \( S_N \). Define the relaxed uncertainty sets as follows for \( i \in [m] \):

\[
P_{i,N}^{\eta + \epsilon} = \{ P \in P_N : |E_P[\psi_k] - E_{Q_i}[\psi_k]| \leq \eta + \epsilon, \quad k \in [K] \}.
\]

(15)

Consider the relaxation of the minmax problem with the uncertainty sets \( P_i \) replaced by \( P_{i,N}^{\eta + \epsilon} \) as follows:

\[
\inf_{\phi_i \in [0,1]^N, i \in [m]} \sup_{P_i \in [0,1]^N, i \in [m]} \left\{ 1 - \frac{1}{m} \sum_{i=1}^{m} \phi_{i,N}^{T} P_{i,N} \right\} \quad \text{s.t.} \quad \sum_{j=1}^{N} \phi_{i,N}(z_j) \psi_k(z_j) - E_{Q_i}[\psi_k] \leq \eta + \epsilon, \quad k \in [K], \\
P_{i,N}^{T} 1 = 1, \quad i \in [m], \quad \sum_{i=1}^{m} \phi_{i,N}(z_j) = 1, \quad j \in [N].
\]

(16)

We can construct a tractable dual to the above problem similar to the finite alphabet case as follows:

\[
\inf_{\phi_i, \lambda^e, \lambda^n} \left\{ 1 - \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda^e_{i,k} (a_{0,k} - \epsilon) + \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda^n_{i,k} (b_{0,k} + \epsilon) - \sum_{i=1}^{m} \mu_i, \quad \text{s.t.} \quad \frac{1}{m} \phi_i(z_j) - \sum_{k=1}^{K} \lambda^e_{i,k} \psi_k(z_j) + \sum_{k=1}^{K} \lambda^n_{i,k} \psi_k(z_j) - \mu_i \geq 0, \\
j \in [N], \quad i \in [m], \quad \sum_{i=1}^{m} \phi_{i,N}(z_j) = 1, \quad j \in [N].
\]

(17)

This is a finite dimensional optimization problem and can be solved efficiently. With the optimal solution to the above problem \( \phi_{N,i} \), we can obtain the saddle point distributions \( P_{i,N}^{*}, \gamma_{N,i} \) as in the finite alphabet case.

Let the optimal values of the minmax problem in (5) and the approximation problem in (16) be denoted by \( \gamma \) and \( \gamma_{N,i} \), respectively. We have the following result showing the convergence of \( \gamma_{N,i} \) to \( \gamma \) and bounding the error due to the approximation.

Theorem 4: With the optimal values of the minmax Bayes formulation and its approximation denoted as \( \gamma \) and \( \gamma_{N,i} \), respectively, as \( \epsilon \to 0 \) (equivalently \( N \to \infty \)), \( \gamma_{N,i} \) converges to \( \gamma \), with \( |\gamma_{N,i} - \gamma| \leq L_0 \epsilon \).

The proof of the above theorem is provided in [21].

We will now construct a robust detection test from the solution of the approximation problem in (17). With the optimal solution to the above problem \( \phi_{N,i} \), we can obtain the saddle point distributions \( P_{i,N}^{*}, \gamma_{i,N} \) as in the finite alphabet case, i.e., \( (\phi_{N,i}, P_{i,N,i}, \gamma_{N,i}) \) is a saddle point for the relaxed problem in (16). To construct a robust test, we extend the discrete distributions to the whole space through kernel smoothing as:

\[
p_i^{*}(x) = \sum_{j=1}^{N} k(x, z_j) p_{i,N}^{*}(z_j).
\]

(18)

Popular kernels used for smoothing are k-Nearest Neighbor kernels, Gaussian kernels and the Epanechnikov (parabola) kernel. Define the pair-wise likelihood ratios as follows:

\[
\ell_i(x) = \frac{p_i^{*}(x)}{p_{i}^{*}(x)}.
\]

(19)

The proposed test is defined as follows. For \( x \in \mathcal{X} \), if \( \max_{i \in [m]} \ell_i(x) = \{i^*\} \) (i.e., there are no tie breaks), then

\[
\phi_{i}^{*}(x) = \begin{cases} 1 & \text{if } i = i^* \\ 0 & \text{otherwise} \end{cases}
\]

(20)

In case of a tie, i.e., \( \max_{i \in [m]} \ell_i(x) \) is not a singleton set, \( \phi_{i}^{*}(x) = \phi_{N,i}^{*}(z_x) \) for \( z_x \in \arg\min_{z \in S_N} \|x - z\| \). For instance,
the 1-NN kernel smoothing test can be defined as follows. The mass $p_{j,N}^*(z_j)$ is distributed uniformly on all points in the corresponding set in the partition $A_j$ for $j \in [N]$. In this case, the proposed test reduces to a 1-NN test, where for $x \in \mathcal{X}$

$$
\phi^*(x) = \phi_N^*(z_x), \quad z_x = \arg\min_{z_j \in \mathcal{S}_N} \|x - z_j\|,
$$

(21)
since $\phi_N^*$ is the likelihood ratio test with $P_1^*, \ldots, P_m^*$ when there are no tie breaks.

We present the following result that bounds the worst-case error of the 1-NN kernel smoothing test in (21) over all distributions in the uncertainty sets.

**Theorem 5:** With the optimal value of the minimax Bayes formulation as $\gamma$, let the robust test $\phi^*$ with 1-NN kernel smoothing be as defined in (21). Then,

$$
\sup_{P_i \in \mathcal{P}_m} \mathbb{E}_P(\phi^*; P_1, \ldots, P_m) - \gamma \leq L_0 \epsilon.
$$

(22)

The proof of the above theorem is provided in [21].

**Remark 2:** The key idea we used in the proof of Theorem 5 is that the 1-NN test at any point $x$ is the same as that at the point $z_x$ in the $\epsilon$-net $\mathcal{S}_N$ closest to $x$, and thus the Lipschitz property of the moment functions can be used to bound the error probability of the 1-NN test for all distributions in the uncertainty sets. However, for other kernels, the test at $x$ depends on not just $z_x$, but additional points from $\mathcal{S}_N$ as well. This makes it challenging to bound the error probability in a similar manner, since the Lipschitz property of the moment functions is no longer enough to obtain such a bound.

As in classical hypothesis testing, for a batch of i.i.d. observations $\{x_j\}_{j=1}^s$, we propose a test that chooses the hypothesis with the maximum sum of log-likelihood ratios $\sum_{j=1}^s \log \ell_i(x_j)$. Note that this test is not necessarily optimal; we evaluate the performance on i.i.d. samples empirically in Section V.

**V. EXPERIMENTAL RESULTS**

In this section, we provide some simulation results for real data. In our experiments, we compare different kernels to extend the discrete distributions to the whole space: the k-NN kernel for different values of $k$, the Gaussian kernel and the parabola kernel (detailed definitions provided in [21]). Additionally, we compare our proposed test with a heuristic test defined as follows. Let $\hat{P}_s = \frac{1}{s} \sum_{j=1}^s \delta_{x_j}$ be the empirical distribution of the batch samples $x^* = \{x_j\}_{j=1}^s$. Consider the test statistics $T_i(x^*) = \sum_{k=1}^K |E_{\hat{P}_s}[\psi_k] - E_{Q_i}[\psi_k]|^2$. We propose a heuristic test for i.i.d. observations that chooses the hypothesis $i^* \in \arg\min_{i \in [m]} T_i(x^*)$, with ties broken arbitrarily. In [21], we show that the proposed heuristic test is exponentially consistent. Therefore, it serves as a good benchmark to compare with our proposed test.

**Remark 3:** It is to be noted that different constructions of uncertainty sets have their own advantages and disadvantages, and choosing a particular one depends on the application. It is not meaningful to compare the empirical performances of robust tests for different uncertainty sets, as it might not be possible to ascertain that the different sets are equivalent in some sense for a fair comparison.

We provide some results for real data experiments to compare the performance of different tests. We use a dataset collected with the Actitracker system [22]–[24] to form the hypotheses. For the case with three hypotheses, we use 25 samples of the walking data from user 669, 25 samples of the jogging data from user 685 and 25 samples of the sitting data from user 594 to construct the uncertainty sets. For the case with four hypotheses, we use 25 samples of the walking data from user 669, 25 samples of the jogging data from user 685, 15 samples of the sitting data from user 594 and 15 samples of the lying down data from user 1603 to construct the uncertainty sets. We use the coordinate wise first and second moments as the constraints function. We plot the error probability as a function of the testing sample size. From Fig. 1, we have that our proposed test with k-NN, Gaussian and parabola smoothing performs better than the heuristic test.

**ACKNOWLEDGMENT**

The work of Z. Sun and S. Zou was supported in part by the National Science Foundation under Grants 2106560 and 2112693. The work of A. Magesh and V.V. Veeravalli was supported by the National Science Foundation under grant #2106727.
REFERENCES

[1] P. Moulin and V. V. Veeravalli, *Statistical Inference for Engineers and Data Scientists*. Cambridge University Press, 2018.

[2] P. J. Huber, “A robust version of the probability ratio test,” *The Annals of Mathematical Statistics*, pp. 1753–1758, 1965.

[3] V. V. Veeravalli, T. Basar, and H. V. Poor, “Minmax robust decentralized detection,” *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 35–40, 1994.

[4] B. C. Levy, “Robust hypothesis testing with a relative entropy tolerance,” *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 413–421, 2009.

[5] G. Gül and A. M. Zoubir, “Minmax robust hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5572–5587, 2017.

[6] Y. Jin and L. Lai, “On the adversarial robustness of hypothesis testing,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 515–530, 2021.

[7] R. Gao, L. Xie, Y. Xie, and H. Xu, “Robust hypothesis testing using Wasserstein uncertainty sets,” in *Proceedings of Advances in Neural Information Processing Systems*, 2018, pp. 7902–7912.

[8] L. Xie, R. Gao, and Y. Xie, “Robust hypothesis testing with Wasserstein uncertainty sets,” *arXiv preprint arXiv:2105.14348*, 2021.

[9] J. Wang and Y. Xie, “A data-driven approach to robust hypothesis testing using sinkhorn uncertainty sets,” in *Proceedings of IEEE International Symposium on Information Theory*. IEEE, 2022, pp. 3315–3320.

[10] Z. Sun and S. Zou, “Kernel robust hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 69, no. 10, pp. 6619–6638, 2023.

[11] H. Rieder, “Least favorable pairs for special capacities,” *The Annals of Statistics*, vol. 5, no. 5, pp. 909–921, 1977.

[12] T. Bednarski, “On solutions of minmax test problems for special capacities,” *Z. Wahrscheinlichkeitstheorie verb Gebiete*, vol. 43, pp. 49–55, 1978.

[13] R. Hafner, “Simple construction of least favourable pairs of distributions and of robust tests for prokhorov-neighborhoods,” *Series Statistics*, vol. 13, no. 1, pp. 33–46, 1982.

[14] S. Kassam, “Robust hypothesis testing for bounded classes of probability densities (corresp.),” *IEEE Transactions on Information Theory*, vol. 27, no. 2, pp. 242–247, 1981.

[15] R. Hafner, “Construction of minmax-tests for bounded families of probability-densities,” *Metrika*, vol. 40, no. 1, pp. 1–23, 1993.

[16] K. Vastola and H. Poor, “On the p-point uncertainty class (corresp.),” *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 374–376, 1984.

[17] C. Pandit, S. Meyn, and V. Veeravalli, “Asymptotic robust neyman-pearson hypothesis testing based on moment classes,” in *Proceedings of International Symposium on Information Theory*. IEEE, 2004, p. 220.

[18] A. Magesh, Z. Sun, V. V. Veeravalli, and S. Zou, “Robust hypothesis testing with moment constrained uncertainty sets,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[19] J. O. Berger, *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[20] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, “Design considerations for the wisdm smart phone-based sensor mining architecture,” in *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, 2011, p. 25–33.

[21] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, p. 74–82, 2011.

[22] G. M. Weiss and J. W. Lockhart, “The impact of personalization on smartphone-based activity recognition,” in *Proceedings of AAAI Workshop on Activity Context Representation: Techniques and Languages*, 2012, pp. 98–104.