ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096843

ROBUST HYPOTHESIS TESTING WITH MOMENT CONSTRAINED UNCERTAINTY SETS

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

†University of Illinois at Urbana-Champaign
‡University at Buffalo, the State University of New York
Email:amagesh2@illinois.edu zhongcha@buffalo.edu vvv@illinois.edu szou3@buffalo.edu

ABSTRACT

The problem of robust binary hypothesis testing is studied. Under both hypotheses, the data-generating distributions are assumed to belong to uncertainty sets constructed through moments; in particular, the sets contain distributions whose moments are centered around the empirical moments obtained from training observations. The goal is to design a test that performs well under all distributions in the uncertainty sets, i.e., minimize the worst-case error probability over the uncertainty sets. In the finite-alphabet case, the optimal test is obtained. In the infinite-alphabet case, a tractable approximation to the worst-case error is derived that converges to the optimal value A test is further constructed to generalize to the entire alphabet. An exponentially consistent test for testing batch samples is also proposed. Numerical results are provided to demonstrate the performance of the proposed robust tests.

Index Terms— Moment robust test, Bayesian setting, tractable approximation, converge, exponentially consistent.

1. INTRODUCTION

Binary hypothesis testing is a fundamental statistical decision-making problem in which the goal is to decide between two given hypotheses based on observed data [1–3]. The two hypotheses H_0 and H_1 are generally referred to as the null and the alternate hypotheses, respectively. The likelihood ratio between the distributions under the two hypotheses can be used to construct the optimal test under various settings. However, in general, these distributions may be unknown and need to be estimated from historical data. Deviations from the true distributions can result in significant performance degradation in likelihood ratio tests. The *robust* hypothesis testing framework [4–19] can be used to alleviate this performance degradation. In the robust setting, it is assumed that 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 a collection of distributions that lie within the neighbourhood (with respect to some discrepancy measure) of certain *nominal* distributions. The epsilon-contaminated uncertainty sets were studied by Huber [4], and a censored likelihood ratio test was proposed and proved to be minimax optimal. Moment-constrained uncertainty sets under the Neyman-Pearson setting were studied in [12]. The above works assume the nominal distributions to be known or estimated from historical data. More recent works have studied the problem of constructing uncertainty sets using a data-driven approach [18, 19], where the nominal distributions are the empirical distributions derived from training observations. In [18], the Wasserstein distance was used to construct uncertainty sets. The minimax problem in the Bayesian setting was considered, and a computationally tractable reformulation and the optimal robust test were characterized [18]. In [19] the maximum mean discrepancy (MMD) was used to construct uncertainty sets. In the Bayesian setting, a tractable approximation to the minimax problem was proposed, and in the Neyman-Pearson setting, an asymptotically optimal test was proposed [19].

In this paper, we study the problem of robust hypothesis testing with moment-constrained uncertainty sets, i.e., the sets contain distributions whose moments are centered around the empirical moments. Mean constrained and variance constrained uncertainty sets can be viewed as special cases. Moment constrained sets are practical as it is computationally easy to calculate empirical moments. We study the minimax formulation in the Bayesian setting. First, we present the results for the case when the distributions under the two hypotheses are supported on a finite alphabet set $\mathcal{X}$. We then extend the study to the case when $\mathcal{X}$ is infinite (the infinite-alphabet case contains the continuous-alphabet case and discrete infinite-alphabet case) and present the optimal test. In the infinite-alphabet case, we provide a tractable approximation of the worst-case error that converges to the optimal value, and propose a test that generalizes to the entire alphabet. An exponentially consistent test for testing i.i.d. batch samples is also proposed. We provide numerical results to demonstrate the performance of our proposed algorithms.

*Equal contribution.

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.

2. PROBLEM SETUP

Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact set denoting the 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 $\mathcal{P}_0, \mathcal{P}_1 \subset \mathcal{P}$ denote the uncertainty sets under the null and alternate hypotheses, respectively. We construct the uncertainty sets using general moment constraints derived from observations from the two hypotheses. Let $\hat{\mathbf{x}}_\mathbf{0} = (\hat{x}_{0,0}, \dots, \hat{x}_{0,m})$ and $\hat{\mathbf{x}}_\mathbf{1} = (\hat{x}_{1,0}, \dots, \hat{x}_{1,n})$ denote the training sequences under the two hypotheses. Let $\hat{Q}_0 = \frac{1}{m}\sum_{j=1}^m \delta_{\hat{x}_{0,j}}$ and $\hat{Q}_1 = \frac{1}{n}\sum_{j=1}^n \delta_{\hat{x}_{1,j}}$ denote the empirical distributions corresponding to the training observations, where δ_x corresponds to 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, where $[K] = \{1, ..., K\}$. The uncertainty sets for $i = 0, 1$ are defined as follows:

$$\mathcal{P}_i^\theta = \left\{ P \in \mathcal{P} : \left| E_P[\psi_k] - E_{\hat{Q}_i}[\psi_k] \right| \le \theta,\ \ k \in [K] \right\}, \tag{1}$$

where θ is the pre-specified radius of the uncertainty sets. It is assumed that the uncertainty sets do not overlap, i.e.,

$$\theta < \max_{k\in[K]} \frac{\left| E_{\hat{Q}_1}[\psi_k] - E_{\hat{Q}_0}[\psi_k] \right|}{2}$$

Given a new sample $x \in \mathcal{X}$, the robust hypothesis problem is defined as follows:

$$\begin{aligned} H_0 : x \sim P_0, \quad P_0 \in \mathcal{P}_0^\theta \\ H_1 : x \sim P_1, \quad P_1 \in \mathcal{P}_1^\theta. \end{aligned} \tag{2}$$

A test $\phi : \mathcal{X} \to [0, 1]$ accepts H_0 with probability $\phi(x)$, and accepts H_1 with probability $1 - \phi(x)$. Let

$$\begin{aligned} P_F(\phi) &\triangleq \sup_{P_0 \in \mathcal{P}_0^\theta} E_{P_0}\big[\phi(x)\big], \\ P_M(\phi) &\triangleq \sup_{P_1 \in \mathcal{P}_1^\theta} E_{P_1}\big[1 - \phi(x)\big] \end{aligned} \tag{3}$$

denote the worst-case probability of false alarm (type-I error probability) and the worst-case probability of miss detection (type-II error probability) for the test ϕ. In the Bayesian setting with equal priors, the probability of error is given by:

$$P_E(\phi) \triangleq \frac{1}{2} E_{P_0}\big[\phi(x)\big] + \frac{1}{2} E_{P_1}\big[1 - \phi(x)\big]. \tag{4}$$

The goal in the Bayesian setting is to solve

$$\inf_\phi \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} P_E(\phi). \tag{5}$$

Note that our analysis is easily generalized to unequal priors.

In this paper, we assume that any distributions $P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta$ admit probability density functions (PDFs) p_0, p_1, with respect to a common reference measure μ.

3. FINITE ALPHABET: OPTIMAL TEST

An important result in arriving at a solution to the minimax robust hypothesis testing problem involves interchanging the infimum and supremum in (5).

Theorem 1.

$$\inf_\phi \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} P_E(\phi) = \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} \inf_\phi P_E(\phi). \tag{6}$$

The advantage of applying the minimax theorem is that the inner minimization problem in (6) corresponds to the optimal Bayes error in classical binary hypothesis testing, and is achieved by the likelihood ratio test. The problem then reduces to a single maximization problem as follows:

$$\begin{aligned} &\inf_\phi \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} P_E(\phi) = \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} \inf_\phi P_E(\phi) && (7) \\ &= \frac{1}{2} \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} \int_\mathcal{X} \min\{p_0(x), p_1(x)\}\, dx && (8) \\ &= \frac{1}{2} \sup_{P_0 \in \mathcal{P}_0^\theta, P_1 \in \mathcal{P}_1^\theta} 1 - \mathrm{TV}(P_0, P_1), && (9) \end{aligned}$$

where $\mathrm{TV}(P_0, P_1)$ is the total variation between P_0 and P_1.

We first consider the case when the alphabet size is finite, i.e., $|\mathcal{X}| < \infty$. Let $\mathcal{X} = \{z_1, \dots, z_N\}$. The following theorem gives the optimal test in this case.

Theorem 2. *Let $|\mathcal{X}| = N < \infty$. Let $P_0^* = (p_0^*(z_1), \dots, p_0^*(z_N))$, $P_1^* = (p_1^*(z_1), \dots, p_1^*(z_N))$ be the optimal solution to the following optimization problem:*

$$\begin{aligned} \max_{P_0, P_1 \in \mathbb{R}^N} &\sum_{j=1}^N \min\{p_0(z_j), p_1(z_j)\} \\ s.t. \quad &\left| \sum_{j=1}^N p_i(z_j)\psi_k(z_j) - E_{\hat{Q}_0}[\psi_k] \right| \le \theta,\ \ k \in [K] \\ &\sum_{j=1}^N p_i(z_j) = 1 \\ &0 \le p_i(z_j) \le 1, \quad j = 1, \dots, N, \quad i = 0, 1. \end{aligned} \tag{10}$$

Then, the likelihood ratio test between P_0^* and P_1^* achieves the optimal minimax Bayes risk in (7).

The result in the above theorem follows from using (8) in the finite alphabet case, which characterizes the minimax

Bayes error for the likelihood ratio test. Note that the optimization problem in Theorem 2 can be solved effectively using available solvers for finite dimension convex optimization problem with linear constraints.

4. INFINITE ALPHABET

In this section, we consider the case when $\mathcal{X}$ is infinite. 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 $\mathcal{X} \subseteq [0,1]^d$. In this case, the minimax formulation in (8) is in general an infinite-dimensional optimization problem, and closed form solutions are difficult to derive. We propose a tractable finite dimension optimization problem as an approximation to the minimax problem, and construct a robust detection test based on the solution to the approximation. In addition, we also quantify the error arising from the approximation of the original problem formulation.

First, note that the moment defining functions $\psi_k,\ k = 1,\ldots,K$ 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 can set $L = \max_k L_k = 1$. In addition, we consider values of $\theta \in [0,\theta_0]$, where

$$\theta_0 < \theta_{\max} = \max_{k\in[K]} \frac{\left|E_{\hat{Q}_1}[\psi_k] - E_{\hat{Q}_0}[\psi_k]\right|}{2}.$$

Let $\epsilon > 0$ such that $\theta + \epsilon \leq \theta_0$. Consider a discretization of the space $\mathcal{X}$ through an ϵ-net or a covering set. Indeed we can consider a simple and efficient construction by considering a grid of equally spaced $N = \lceil \frac{1}{\epsilon^d} \rceil$ points $\mathcal{S}_N = \{z_1,\ldots,z_N\}$ such that for any $x \in \mathcal{X}$,

$$\min_{i=1,\ldots,N} \|z_i - x\| \leq \epsilon. \tag{11}$$

Here, N depends on ϵ, and we ignore the dependence on ϵ in the notation for N for readability. Let $\mathcal{P}_N$ denote all the distributions that are supported on the set $\mathcal{S}_N$. Define the relaxed uncertainty sets as follows for $i = 0,1$:

$$\mathcal{P}_{i,N}^{\theta+\epsilon} = \left\{P \in \mathcal{P}_N : \left|E_P[\psi_k] - E_{\hat{Q}_i}[\psi_k]\right| \leq \theta+\epsilon,\ \ k \in [K]\right\}. \tag{12}$$

Consider the maximization problem in (8) with the uncertainty sets $\mathcal{P}_{0,N}^{\theta+\epsilon}, \mathcal{P}_{1,N}^{\theta+\epsilon}$:

$$\frac{1}{2} \sup_{P_0\in\mathcal{P}_{0,N}^{\theta+\epsilon}, P_1\in\mathcal{P}_{1,N}^{\theta+\epsilon}} \sum_{i=1}^{N} \min\{p_0(z_i), p_1(z_i)\}, \tag{13}$$

which can be written as

$$\begin{aligned} \sup_{P_0,P_1\in\mathbb{R}^N} & \sum_{j=1}^{N} \min\{p_0(z_j), p_1(z_j)\} \\ s.t. \quad & \left|\sum_{j=1}^{N} p_i(z_j)\psi_k(z_j) - E_{\hat{Q}_0}[\psi_k]\right| \leq \theta+\epsilon, \quad k\in[K], \\ & \sum_{j=1}^{N} p_i(z_j) = 1, \\ & 0 \leq p_0(z_j) \leq 1, \quad j = 1,\ldots,N, \quad i = 0,1. \end{aligned} \tag{14}$$

This is a finite dimension convex optimization problem with linear constraints that can be solved efficiently. We first show that the optimal value of the tractable relaxation in (13) converges to the optimal value of our original problem (8), and quantify the error introduced by the relaxation. We then propose a robust detection test for our original problem based on the solution to (13), and quantify the error due to the approximation. The following lemma will be useful in the proof of convergence of the approximation in (13).

Lemma 1. *Let*

$$g(\theta) := \sup_{P_0\in\mathcal{P}_0^\theta, P_1\in\mathcal{P}_1^\theta} \int_{\mathcal{X}} \min\{p_0(x), p_1(x)\}\, dx. \tag{15}$$

The function $g(\theta)$ is continuous in θ.

The proof of the above Lemma follows from showing that $g(\theta)$ is concave in θ. Thus, we have that $g(\theta)$ is a concave function on a open interval $(0,\theta_{\max})$, and hence Lipschitz on a closed interval $[0^+,\theta_0]$. Let the Lipschitz constant be denoted by L_0. Define

$$\gamma = \frac{1}{2} \sup_{P_0\in\mathcal{P}_0^\theta, P_1\in\mathcal{P}_1^\theta} \int_{\mathcal{X}} \min\{p_0(x), p_1(x)\}\, dx, \tag{16}$$

$$\gamma_\epsilon = \frac{1}{2} \sup_{P_0\in\mathcal{P}_{0,N}^{\theta+\epsilon}, P_1\in\mathcal{P}_{1,N}^{\theta+\epsilon}} \sum_{i=1}^{N} \min\{p_0(z_i), p_1(z_i)\}. \tag{17}$$

Theorem 3. *With the optimal values of the minimax Bayes formulation and its approximation denoted as in* (16) *and* (17) *respectively, as $\epsilon \to 0$ (equivalently $N \to \infty$), γ_ϵ converges to γ, with $|\gamma_\epsilon - \gamma| \leq L_0\epsilon$.*

Let $P_{0,N}^*, P_{1,N}^*$ be the solution to the optimization problem in (14). Recall the partition $\{\mathcal{A}_1,\ldots,\mathcal{A}_N\}$ defined by the set $\mathcal{S}_N$ on $\mathcal{X}$ such that for any $j = 1,\ldots,N$, if $x \in \mathcal{A}_j$, then

$$\|x - z_j\| \leq \epsilon. \tag{18}$$

In order to construct a robust detection test, we extend these discrete distributions defined on $\mathcal{S}_N$ to the whole space $\mathcal{X}$ as P_0^*, P_1^*. For $i = 0,1$, we distribute the probability mass of the point z_j onto points in the set $\mathcal{A}_j$ through a common channel. For instance, we can distribute the mass $p_{i,N}^*(z_j)$ uniformly on all points in the set $\mathcal{A}_j$ for $j = 1,\ldots,N$, i.e., for $x \in \mathcal{X}$

$$p_i^*(x) = \sum_{j=1}^{N} \frac{p_{i,N}^*(z_j)\mathbf{1}_{\{x\in\mathcal{A}_j\}}}{\int_{\mathcal{A}_j} dx}. \tag{19}$$

Then, we can define the robust test with $\ell(x) = \frac{p_1^*(x)}{p_0^*(x)}$ as:

$$\phi^*(x) = \begin{cases} 1, & \text{if} \quad \log \ell(x) \geq 0 \\ 0, & \text{if} \quad \log \ell(x) < 0. \end{cases} \tag{20}$$

Let the Bayes error for the test ϕ^* be denoted by $P_E(\phi^*)$. The following theorem quantifies the optimality gap between $P_E(\phi^*)$ and the optimal minimax error in (16).

Theorem 4. *Let the robust test ϕ^* be as defined in* (20). *Then,*

$$|P_E(\phi^*) - \gamma| \leq L_0 \epsilon. \tag{21}$$

5. DIRECT ROBUST TEST FOR BATCH SAMPLES

Let $x_1^s = (x_1, \ldots, x_s)$ be a sequence of i.i.d. observations, where s is the sample size. One way to test this batch sample is to extend the log likelihood ratio test proposed in (20). However, it is difficult to analyze the error exponent of the test. In this section, we propose a test for testing batch samples, and show that it is exponentially consistent as $s \to \infty$ in the Bayesian setting.

Let $\hat{P}_s = \frac{1}{s} \sum_{j=1}^{s} \delta_{x_j}$ be the empirical distribution of the batch sample, and consider the test statistic

$$\begin{aligned} T(x_1^s) = &\sum_{k=1}^{K} \left| E_{\hat{P}_s}[\psi_k] - E_{\hat{Q}_0}[\psi_k] \right|^2 \\ &- \sum_{k=1}^{K} \left| E_{\hat{P}_s}[\psi_k] - E_{\hat{Q}_1}[\psi_k] \right|^2 . \end{aligned} \tag{22}$$

We propose the following test for sequence of observations:

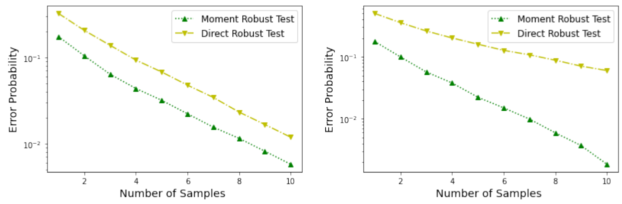
$$\phi_s(x_1^s) = \begin{cases} 1, & \text{if} \quad T(x_1^s) \geq 0 \\ 0, & \text{if} \quad T(x_1^s) < 0. \end{cases} \tag{23}$$

Theorem 5. *The test in* (23) *is exponentially consistent.*

6. SIMULATION RESULTS

In this section, we provide some numerical results. We first compare our moment robust test ϕ^* with the direct robust test ϕ_s using synthetic data. For hypothesis H_0, the uncertainty set is constructed using 20 observations collected from a multi-variate Gaussian distribution with mean $[0, 0, 0, 0]$ and covariance matrix $\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. For H_1, the uncertainty set is constructed using 20 observations collected from a multi-variate Gaussian distribution with a different mean $[0.8, 0.8, 0.8, 0.8]$ and covariance matrix $0.5\mathbf{I}$. The constraint functions are chosen as the mean and variance of each dimensions. We define the uncertainty sets as the collections of distributions such that the mean and variance for each dimensions lie in a certain range centered around the empirical mean and variance. We use true distributions to generate the test sample and plot the error probability as a function of testing sample size n.

From Fig. 1(a), it can be seen that the error probabilities for the moment robust test and the direct robust test decay exponentially with the testing sample size, which demonstrates the exponentially consistency of the direct robust test. Moreover, the moment robust test performs better than the direct robust test.



(a) Synthetic Data (b) Real Data

Fig. 1. Comparison of the Moment Robust Test and the Direct Robust Test

We then use the real data to compare our moment robust test ϕ^* with the direct robust test ϕ_s. We use a dataset collected with the Actitracker system [20–22] to form the hypotheses. For hypothesis H_0, the jogging data of the person 685 is used to construct the uncertainty sets. For hypothesis H_1, the walking data of the person 669 is used to construct the uncertainty set. The uncertainty sets are defined as in the synthetic data case. We plot the error probability as a function of sample size n. In Fig. 1(b), it can be seen that the moment robust test performs better than the direct robust test and the error probabilities for the moment robust test and the direct robust test decay exponentially with the testing sample size, which demonstrates the exponentially consistency of the direct robust test.

7. CONCLUSION

In this paper, we studied the robust hypothesis testing problem, with uncertainty sets constructed through moments. We focused on the Bayesian setting, where the goal is to minimize the worst-case error probability over the uncertainty sets. We proposed the optimal test for the finite-alphabet case, and a tractable approximation of the worst-case error probability that converges to the optimal value of the original problem for the infinite-alphabet case. Based on the tractable approximation, a moment robust test was constructed. We also proposed an exponentially consistent test for testing batch samples and provided numerical results to demonstrate the performance of the proposed robust tests. The detailed proofs of the results in this paper can be found in [23]. Extensions of the results in this paper to uncertainty sets constructed through matrix-valued moment functions, and results for the asymptotic Neyman-Pearson setting are also given in [23].

8. REFERENCES

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

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: John Wiley & Sons, 2006.

[3] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

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

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

[6] F. Österreicher, "On the construction of least favourable pairs of distributions," *Z. Wahrscheinlichkeitstheorie verw Gebiete*, vol. 43, pp. 49–55, 1978.

[7] T. Bednarski, "On solutions of minimax test problems for special capacities," *Z. Wahrscheinlichkeitstheorie verw Gebiete*, vol. 58, pp. 397–405, 1981.

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

[9] 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.

[10] R. Hafner, "Construction of minimax-tests for bounded families of probability-densities," *Metrika*, vol. 40, no. 1, pp. 1–23, 1993.

[11] 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.

[12] 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.

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

[14] M. Barni and B. Tondi, "The source identification game: An information-theoretic perspective," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 450–463, 2013.

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

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

[17] Y. Jin and L. Lai, "On the adversarial robustness of hypothesis testing," *IEEE Transaction on Signal Processing*, to appear, 2020.

[18] L. Xie, R. Gao, and Y. Xie, "Robust hypothesis testing with wasserstein uncertainty sets," *arXiv preprint arXiv:2105.14348*, 2021.

[19] Z. Sun and S. Zou, "Kernel robust hypothesis testing," *arXiv preprint arXiv:2203.12777*, 2022.

[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," *AAAI Workshop on Activity Context Representation: Techniques and Languages*, pp. 98–104, 2012.

[23] A. Magesh, Z. Sun, V. V. Veeravalli, and S. Zou, "Robust hypothesis testing with moment constrained uncertainty sets," 2022. [Online]. Available: https://arxiv.org/abs/2210.12869