Asymptotic Robust Hypothesis Testing Based on Moment Classes

Sean Meyn and Venugopal V. Veeravalli
ECE Department & Coordinated Science Lab
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
Email: meyn,vvv@uiuc.edu

***Abstract*— A robust hypothesis testing framework is introduced in which candidate hypotheses are characterized by moment classes. It is shown that there exists a test sequence that is asymptotically optimal in the min-max sense, and that it is expressed as a comparison of a log-linear combination of the constraint functions to a pre-determined threshold. Algorithms are described to compute the optimal test and applications of the robust hypothesis testing framework are discussed.**

I. Introduction

We consider the binary hypothesis testing problem based on a finite number of observations from a sequence of observations $\boldsymbol{X} = \{X_t : t = 1, \ldots\}$, taking values in a finite set $\mathcal{A}$. The goal is to classify a given set of observations into one of the two hypotheses, H_0 or H_1. In much of this paper, it is assumed that, conditioned on the hypotheses H_0 or H_1, these observations are independent and identically distributed (i.i.d.), with marginal distribution π_j under hypothesis H_j for $j = 0, 1$.

Standard approaches to deciding whether the observations come from H_0 or H_1 include the Bayesian, Neyman-Pearson, and min-max criteria (see, e.g., [1]). It is well-known that when the distributions π_0 and π_1 are specified, the optimal test under any one of these three criteria can be expressed as a Likelihood Ratio Test (LRT) [1].

In most applications, however, π_0 and π_1 are not completely specified *a priori*. Typically these distributions must be estimated through on-line or off-line measurements. The statistical modeling problem may then be solved by choosing an appropriate statistical model that is consistent with the measurements. Given that in reality the measurement data is always finite, there is a multitude of statistical models that are consistent with the data.

In some exceptional cases, which arise mainly in the context of digital communications, the distributions under both hypotheses can be estimated very accurately. For example, in the communication of binary information over the telephone line channel, the observations (under each hypothesis) are the sum of a deterministic transmitted signal (that depends on the hypothesis) and a random noise term (that is independent of the hypothesis). The statistics of the noise term are relatively time-invariant and easy to estimate accurately prior to communication. In particular, zero-mean Gaussian models for the noise are known to be accurate, and the only parameter that requires estimation is the noise variance. Hence the distributions π_0 and π_1 are essentially known.

In other cases that arise in the context of surveillance, one of the hypotheses, typically denoted by H_0, corresponds to the case “null” where the intruder is absent, and the alternative H_1 corresponds to the presence of an intruder. Such problems have recently generated much interest in the context of sensor networks (see, e.g., [2] and references therein). The default state of the system is H_0, and often there is ample opportunity to learn the observation statistics under this hypothesis. In this case, π_0 can be assumed to be known. On the other hand, the parameters associated with the intruder are typically unknown *a priori*, and hence π_1 may only be partially known.

In general, both π_0 and π_1 may only be known partially. An example of such a scenario is digital communications over wireless channels subject to severe time varying interference with a few dominant interferers as in a cellular wireless communication systems. Here the additive noise term may not be well modeled as Gaussian or even statistically time-invariant. Training to learn the statistics can only result in partial knowledge of π_0 and π_1. Additive Gaussian models are typically used to model the interference only because they result in simple detector structures.

A reasonable way to capture partial knowledge of π_0 and π_1 is through sets of distributions referred to as *uncertainty classes*. A standard approach to designing decision rules in this setting is the min-max approach, where the goal is to minimize the worst-case performance over the uncertainty classes. The decision rules thus obtained are said to be robust to the uncertainties in the probability distributions. Min-max robust detection has been the subject of numerous papers since the seminal work of Huber and Strassen [3], [4]. The solution to the robust detection problem, if one exists, is a LRT between a pair of *least favorable distributions* (LFDs) within the classes. Huber and Strassen showed that LFDs exist for several uncertainty models, such as ϵ-contaminated neighborhoods, total variation neighborhoods, band-classes, and p-point classes, or more generally when the neighborhood classes can be described in terms of alternating capacities of order 2 [4]. The proofs of existence of LFDs and the corresponding robust tests rely on the property that the uncertainty classes satisfy the *joint stochastic boundedness* assumption (as defined in [5]). Unfortunately the uncertainty models for which LFDs are known to exist have limited application in practice, and the specific uncertainty model of interest in this paper (that we describe below) does not appear to be amenable to the Huber-Strassen approach.

This paper concerns uncertainty classes obtained by specifying bounds on a finite number of moments of the distributions under the respective hypotheses. Specifically, we define the *moment classes* $\mathbb{P}_0$ and $\mathbb{P}_1$ as,

$$\mathbb{P}_j = \Big\{\pi \in \mathcal{M} : \langle \pi, f_{ij}\rangle \leq c_{ij}, i = 1, ..., n\Big\}, \qquad j = 0, 1, \tag{1}$$

where $\{f_{ij}\}$ are real-valued functions[1] on $\mathcal{A}$, $\{c_{ij}\}$ are constants, and $\mathcal{M}$ denotes the set of all distributions on $\mathcal{A}$. We use the notation $\langle \pi, f\rangle$ for the expected value of the function f with respect to the distribution π. Clearly, lower bound constraints on the moments $\langle \pi, f_{ij}\rangle$ can be accomodated in this setting by additional constraints on $-f_{ij}$. It is assumed that the sets $\mathbb{P}_0$, $\mathbb{P}_1$ are disjoint.

Our motivation for considering moment classes comes firstly from the simple observation that the most common approach to partial statistical modeling is through moments, typically mean and correlation. Probabilistic inference using moment information has a long and rich history (see discussion in [6]). The primary motivation comes from the fact that it is possible to obtain worst-case bounds on the probability of a given set, over all probability distributions in a given moment class.

As we mentioned previously, for uncertaintly classes defined by moment constraints, the min-max robust versions of the standard hypothesis testing problems do not fall under the Huber and Strassen framework. To facilitate analysis, we turn to the asymptotic setting that is described in more detail in the next section. Throughout this paper we restrict to our attention to the asymptotic version of the Neyman-Pearson (N-P) criterion for evaluating a given detector. The results of this paper can be extended to asymptotic robust versions of the Bayesian and min-max hypothesis testing problems.

II. Asymptotic N-P Hypothesis Testing

The asymptotic robust N-P criterion [7] is described as follows. For a given $N \geq 1$, suppose that a decision test ϕ_N is constructed based on the finite set of measurements $\{X_1, \dots, X_N\}$. This may be expressed as the characteristic function of a subset $A_1^N \subset \mathcal{A}^N$. The test declares that hypothesis H_1 is true if $\phi_N = 1$, or equivalently, $(X_1, X_2, \dots, X_N) \in A_1^N$. In the asymptotic setting considered in this paper, we view $N \geq 1$ as a parameter. The performance of a *sequence* of tests $\boldsymbol{\phi} \triangleq \{\phi_N : N \geq 1\}$ is reflected in the error exponents for the type-II error probability and type-I error probability, defined respectively by,

$$\begin{aligned} I_\phi^{\pi_1} &\triangleq -\liminf_{N\to\infty} \frac{1}{N}\log(\mathsf{P}_{\pi_1}(\phi_N(X_1, \dots, X_N) = 0)), \\ J_\phi^{\pi_0} &\triangleq -\liminf_{N\to\infty} \frac{1}{N}\log(\mathsf{P}_{\pi_0}(\phi_N(X_1, \dots, X_N) = 1)), \end{aligned}$$

where P_{π_j} denotes the distribution on $\boldsymbol{X}$ when $\boldsymbol{X}$ is i.i.d. with marginal π_j.

Consider first the non-robust setting in which the distributions π_0, π_1 are known exactly. The asymptotic N-P criterion for choosing an optimal test is to maximize the type-II exponent subject to a constraint on the type-I exponent. Thus, for a given constant $\eta \geq 0$ as the constraint, we have

$$\sup_\phi I_\phi^{\pi_1} \qquad \text{subject to} \qquad J_\phi^{\pi_0} \geq \eta \tag{2}$$

where the supremum is over all test sequences $\boldsymbol{\phi}$. The optimal value of the exponent $I_\phi^{\pi_1}$ in the asymptotic N-P problem is expressed in Theorem 1 below, which is a combination of results established in [7] and [8]. Before we state the theorem we define some relevant quantities.

For $\mu, \pi \in \mathcal{M}$, the *relative entropy* or *Kullback-Leibler divergence* between μ and π is defined as

$$D(\mu \parallel \pi) \triangleq \Big\langle \mu, \log\frac{\mu}{\pi}\Big\rangle \tag{3}$$

For $\pi \in \mathcal{M}$ and for $\beta \in \Re_+$, we define the *divergence sets*

$$\begin{aligned} \mathcal{Q}_\beta(\pi) &\triangleq \{\mu \in \mathcal{M} : D(\mu \parallel \pi) < \beta\}, \\ \mathcal{Q}_\beta^+(\pi) &\triangleq \{\mu \in \mathcal{M} : D(\mu \parallel \pi) \leq \beta\}. \end{aligned} \tag{4}$$

Divergence sets are convex since $D(\cdot \parallel \pi)$ is a convex function.

Given the two distributions π_0, π_1 we define the *likelihood ratio* by

$$\ell(a) \triangleq \frac{\pi_0(a)}{\pi_1(a)}, \qquad a \in \mathcal{A}. \tag{5}$$

Theorem 1: Suppose that the two distributions π_0, π_1 are known exactly. Then the following statements hold,

(i) The optimal value of $I_\phi^{\pi_1}$ in (2) is given by the minimal divergence

$$\beta^* = \inf_{\mu \in \mathcal{Q}_\eta(\pi_0)} D(\mu \parallel \pi_1) \tag{6}$$

(ii) There exists a distribution $\mu^* \in \mathcal{Q}_\eta^+(\pi_0)$ that solves (6) and this solution satisfies

$$D(\mu^* \parallel \pi_0) = \eta \,, \text{and} \quad D(\mu^* \parallel \pi_1) = \beta^*$$

(iii) The following sequence of log-likelihood ratio tests (LRTs) is optimal for (2).

$$\phi_N \triangleq \mathbb{I}\Big\{x \in \mathcal{A}^N : \frac{1}{N}\sum_{t=1}^N \log \ell(x_t) \ \leq\ \beta^* - \eta\Big\}. \tag{7}$$

where $\mathbb{I}$ is the indicator function.

Part (i) of Theorem 1 was proved by Hoeffding [7]. Results similar to (ii) and (iii) were established by Blahut[8]. Below we provide a sketch of the proof based on Sanov's Theorem and the geometry of convex sets, which will be useful when we study the robust detection problem.

[1]Note that $\{f_{ij}\}$ do not have to be polynomials.

We briefly recall Sanov's theorem here. Define the *types* for $\boldsymbol{X} = (X_1, X_2, \ldots)$ through the sequence of empirical distributions (indexed by N)

$$\gamma_N(a) = \frac{1}{N} \sum_{t=1}^{N} \mathbb{I}\{X_k = a\}, \quad a \in \mathcal{A}. \tag{8}$$

Suppose that $\boldsymbol{X}$ is i.i.d. with marginal distribution π. Sanov's Theorem states that for any closed convex set of distributions $\mathcal{E} \subseteq \mathcal{M}$,

$$\lim_{N\to\infty} -N^{-1} \log \mathsf{P}\{\gamma_N \in \mathcal{E}\} = \inf_{\mu\in\mathcal{E}} D(\mu\|\pi) \tag{9}$$

The relative entropy is jointly convex on $\mathcal{M} \times \mathcal{M}$, and hence computation of the minimum of $D(\mu\|\pi)$ amounts to solving a convex program.

The optimal value of the exponent I_ϕ in the asymptotic N-P problem is described in terms of relative entropy. It is shown in [9] that one may restrict to tests of the following form without loss of generality: for a closed set $\mathcal{E} \subseteq \mathcal{M}$,

$$\phi_N = \mathbb{I}\{\gamma_N \in \mathcal{E}\}, \tag{10}$$

where $\{\gamma_N\}$ denotes the sequence of types (8). Sanov's Theorem (9) tells us that for any test of this form,

$$I_\phi^{\pi_1} = \inf_{\mu\in\mathcal{E}^c} D(\mu\|\pi_1), \quad J_\phi^{\pi_0} = \inf_{\mu\in\mathcal{E}} D(\mu\|\pi_0).$$

The smallest set $\mathcal{E}$ that gives $J_\phi^{\pi_0} \geq \eta$ is the divergence set $\mathcal{Q}_\eta^+(\pi_0)$ (see (4)), and hence the optimum value of $I_\phi^{\pi_1}$ in (2) is the value of the convex program,

$$\begin{aligned}\beta^* &= \sup\{\beta \geq 0 : \mathcal{Q}_\eta(\pi_0) \cap \mathcal{Q}_\beta(\pi_1) = \emptyset\} \\ &= \inf_{\mu\in\mathcal{Q}_\eta(\pi_0)} D(\mu \parallel \pi_1).\end{aligned}$$

and the optimizing $\mu^* \in \mathcal{Q}_\eta^+(\pi_0) \cap \mathcal{Q}_{\beta^*}^+(\pi_1)$ clearly satisfies the conditions given in part (ii) of Theorem 1.

It is clear fromn (10) and (9) that any $\mathcal{E} \in \mathcal{M}$ such that $\mathcal{E} \supseteq \mathcal{Q}_\eta(\pi_0)$ and $\mathcal{E}^c \supseteq \mathcal{Q}_\beta^*(\pi_1)$ defines a sequence of optimal tests for (2) via $\phi_N = \mathbb{I}\{\gamma_N \in \mathcal{E}\}$. In particular, setting $\mathcal{E} = \mathcal{Q}_\eta^+(\pi_0)\}$ yeilds a *universal* test in that it does not require knowledge of π_1 [9]. However, this universal test is considerably more difficult to implement than the LRT of (7).

The key to showing that (7) is optimal is to establish that the following set

$$\mathcal{H} = \{\mu \in \mathcal{M} : \langle \mu, \log \ell\rangle = \langle \mu^*, \log \ell\rangle\}$$

is a separating set (hyperplane) for the convex sets $\mathcal{Q}_\eta^+(\pi_0)$ and $\mathcal{Q}_{\beta^*}^+(\pi_1)$. This result follows from the Kuhn-Tucker alignment conditions based on a dual functional.

Note that we can write the constant $\langle \mu^*, \log \ell\rangle$ as

$$\begin{aligned}\langle \mu^*, \log \ell\rangle &= \left\langle \mu^*, \log \frac{\mu^*}{\pi_1} \right\rangle - \left\langle \mu^*, \log \frac{\mu^*}{\pi_0} \right\rangle \\ &= \beta^* - \eta\end{aligned}$$

Thus an optimal sequence of decision rules results if we choose

$$\phi_N = \mathbb{I}\{\langle \gamma_N, \log \ell\rangle \leq \beta^* - \eta\}$$

which is the test specified in (7).

This geometry is illustrated in Figure 1.

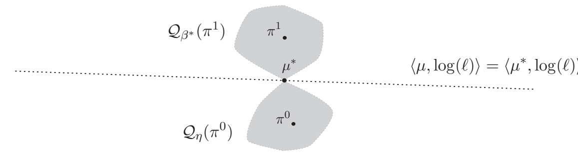

Fig. 1. The likelihood ratio test interpreted via a separating hyperplane between the convex sets $\mathcal{Q}_\eta(\pi_0)$ and $\mathcal{Q}_{\beta^*}(\pi_1)$.

III. Robust Asymptotic N-P Hypothesis Testing

We now get to the main result of this paper, which is an extension of Theorem 1 to the robust framework. When π_0 and π_1 are known to belong to the uncertainty classes $\mathbb{P}_0$ and $\mathbb{P}_1$ respectively, we impose a uniform constraint on the type-I exponent for all $\pi_0 \in \mathbb{P}_0$, and subject to this we seek a test sequence that maximizes the worst type-II exponent across $\pi_1 \in \mathbb{P}_1$. Thus, in the robust version, the asymptotic N-P criterion (2) is replaced by the following constrained optimization:

$$\sup_\phi \inf_{\pi_1\in\mathbb{P}_1} I_\phi^{\pi_1} \qquad \text{subject to} \qquad \inf_{\pi_0\in\mathbb{P}_0} J_\phi^{\pi_0} \geq \eta. \tag{11}$$

Theorem 2 below describes optimal solutions to this optimization problem.

For any moment class $\mathbb{P}$ defined as in (1), consider the divergence sets

$$\mathcal{Q}_\beta(\mathbb{P}) = \bigcup_{\pi\in\mathbb{P}} \mathcal{Q}_\beta(\pi) , \quad \text{and} \quad \mathcal{Q}_\beta^+(\mathbb{P}) = \bigcup_{\pi\in\mathbb{P}} \mathcal{Q}_\beta^+(\pi). \tag{12}$$

It is easy to show that these divergence sets are also convex based on the joint convexity of the relative entropy and the convexity of $\mathbb{P}$.

The following assumption is imposed to ensure that the solution to the robust hypothesis testing problem is non-trivial:

(A1) $\mathcal{Q}_\eta^+(\mathbb{P}_0) \cap \mathbb{P}_1 = \emptyset$.

This simply means that $D(\pi_1 \parallel \pi_0) > \eta$ for each $\pi_0 \in \mathbb{P}_0$, and $\pi_1 \in \mathbb{P}_1$.

Theorem 2: The following hold for the asymptotic robust N-P hypothesis testing problem (11) under Assumption (A1):

(i) The optimal value of $I_\phi^{\pi_1}$ in (11) is given by

$$\begin{aligned}\beta^* &\stackrel{\Delta}{=} \sup\{\beta \geq 0 : \mathcal{Q}_\eta(\mathbb{P}_0) \cap \mathcal{Q}_\beta(\mathbb{P}_1) = \emptyset\} \\ &= \inf_{\pi_1\in\mathbb{P}_1} \inf_{\mu\in\mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \parallel \pi_1).\end{aligned} \tag{13}$$

(ii) There exist $\pi_0^* \in \mathbb{P}_0, \pi_1^* \in \mathbb{P}_1$, and $\mu^* \in \mathcal{M}$ that solve (13) and this solution satisfies

$$D(\mu^*\|\pi_0^*) = \eta, \quad D(\mu^*\|\pi_1^*) = \beta^*.$$

The distributions π_0^*, π_1^* and μ^* are are related through

$$\mu^*(a) = \ell_0(a)\, \pi_0^*(a) = \ell_1(a)\, \pi_1^*(a), a \in \mathcal{A}$$

where

$$\ell_0(a) = \sum_{i=1}^{n} \lambda_i f_{i0}(a) \text{ , and } \quad \ell_1(a) = \sum_{i=1}^{n} \gamma_i f_{i1}(a)$$

for appropriately chosen $\{\lambda_i, \gamma_i, i = 1, \ldots, n\}$

(iii) The following three sequences of tests are all optimal for (11).

$$\begin{aligned} \phi_N^{(0)} &\triangleq \mathbb{I}\{x \in \mathcal{A}^N : \langle \gamma_N, \log \ell_0 \rangle \ > \ \eta\} \\ \phi_N^{(1)} &\triangleq \mathbb{I}\{x \in \mathcal{A}^N : \langle \gamma_N, \log \ell_1 \rangle \ \le \ \beta^*\} \\ \phi_N &\triangleq \mathbb{I}\{x \in \mathcal{A}^N : \langle \gamma_N, \log \ell \rangle \ \le \ \beta^* - \eta\} \end{aligned} \tag{14}$$

where $\ell(a) \triangleq \ell_1(a)/\ell_0(a)$ for $a \in \mathcal{A}$.

Analogous to the proof of Theorem 1, restricting to tests of the form $\phi_N = \mathbb{I}\{\gamma_N \in \mathcal{E}\}$, the smallest set $\mathcal{E}$ that gives $\inf_{\pi_0 \in \mathbb{P}_0} J_\phi^{\pi_0} \ge \eta$ is the divergence set $\mathcal{Q}_\eta^+(\mathbb{P}_0)$. Thus the optimum value of $I_\phi^{\pi_1}$ in (11) is as given in part (i) of Theorem 2.

Also note that any $\mathcal{E} \in \mathcal{M}$ such that $\mathcal{E} \supseteq \mathcal{Q}_\eta(\mathbb{P}_0)$ and $\mathcal{E}^c \supseteq \mathcal{Q}_\beta^*(\mathbb{P}_1)$ defines a sequence of optimal tests for (11) via $\phi_N = \mathbb{I}\{\gamma_N \in \mathcal{E}\}$. In particular, setting $\mathcal{E} = \mathcal{Q}_\eta^+(\mathbb{P}_0)\}$ yeilds a *universal* test. However, simpler optimal tests exist as seen in part (iii) of Theorem 2.

Since $\mathcal{Q}_\eta^+(\mathbb{P}_0)$ and $\mathcal{Q}_{\beta^*}^+(\mathbb{P}_1)$ are compact sets it follows from their construction that there exists $\mu^* \in \mathcal{Q}_\eta^+(\mathbb{P}_0) \cap \mathcal{Q}_{\beta^*}^+(\mathbb{P}_1)$. Moreover, by convexity there exists *some* function $h : \mathcal{A} \to \Re$ defining a separating hyperplane between the sets $\mathcal{Q}_\eta^+(\mathbb{P}_0)$ and $\mathcal{Q}_{\beta^*}^+(\mathbb{P}_1)$, satisfying

$$\begin{aligned} \mathcal{Q}_\eta(\mathbb{P}_0) &\subset \{\mu \in \mathcal{M} : \langle \mu, h \rangle < \tau\}, \\ \mathcal{Q}_{\beta^*}(\mathbb{P}_1) &\subset \{\mu \in \mathcal{M} : \langle \mu, h \rangle > \tau\}. \end{aligned}$$

The remainder of the proof consists of the identification of h and τ using the Kuhn-Tucker alignment conditions based on consideration of a dual functional as in the proof of Theorem 1. However, unlike in the case of Theorem 1, there is some flexibility in the choice of h as we see in part (iii) of Theorem 2. This geometry of the test corresponding to ℓ_0 is illustrated in Figure 2 below.

We also note that likelihood ratio between π_0^* and π_1^* can be written as:

$$\frac{\pi_0^*(a)}{\pi_1^*(a)} = \frac{\ell_1(a)}{\ell_0(a)} = \ell(a)$$

Thus it is tempting to consider the third optimal test sequence in part (iii) to be a likelihood ratio test based on π_0^* and π_1^*. However, the support of π_0^* and π_1^* may be a small subset of $\mathcal{A}$ in applications, whereas $\ell(a)$ is defined for all $a \in \mathcal{A}$.

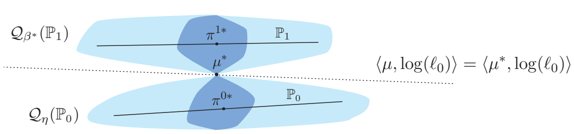

Fig. 2. The robust hypothesis testing problem. The uncertainty classes $\mathbb{P}_i$, $i = 0, 1$ are determined by a finite number of linear constraints, and the thickened regions $\mathcal{Q}_\eta(\mathbb{P}_0)$, $\mathcal{Q}_{\beta^*}(\mathbb{P}_1)$ are each convex. The linear threshold test corresponding to ℓ_0 is interpreted as a separating hyperplane between these two convex sets.

IV. Numerical Computation

The optimal π_0^*, π_1^*, μ^*, and β^* of Theorem 2 can be obtained using standard techniques from nonlinear programming. A recursive approach based on the *cutting-plane* method of Kelley [10] yields an efficient algorithm for numerical computation.

V. Conclusion

We have characterized the solution to the asymptotic robust hypothesis problem where the distributions under the two hypotheses belong to moment classes. We showed that optimal test sequences can be expressed as a comparison of a linear combination of the constraint functions to a threshold.

Potential directions for future work are drawn from both theoretical as well as practical viewpoints. Specifically, we believe the following directions would be of interest:

1) It is likely that many of our conclusions can be extended beyond the i.i.d. case to include classes of Markov processes, and beyond the finite-alphabet case.
2) We have not dealt with the problem of how to select the moment functions $\{f_{ij}\}$ for an arbitrary application. It may be possible to formulate rules for selecting these functions based on additional information about the candidate hypotheses. For instance, when the distribution π_0 is known, the functions $\{f_{i1}\}$ could be tailored so that discriminating between π_0 and $\pi_1 \in \mathbb{P}_1$ is made easier.
3) Extension to M-ary hypothesis testing ($M > 2$) is of interest, particularly in the context of channel coding.

Acknowledgment

The authors thank Charuhas Pandit for his contributions to this research. This paper is based upon work supported by the National Science Foundation under Award Nos. ECS-0217836, ITR-0085929, and the CCF-0049089.

References

[1] H. V. Poor, *An introduction to signal detection and estimation*. New York: Springer-Verlag, 2nd ed., 1994.

[2] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, pp. 407–416, 2003.

[3] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Stat.*, vol. 36, pp. 1753–1758, 1965.

[4] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Annals of Statistics*, vol. 1, pp. 251–263, 1973.

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

[6] C. Pandit, *Robust Statistical Modeling Based on Moment Classes with Applications to Admission Control, Large Deviations, and Hypothesis Testing*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2004.

[7] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Statist.*, vol. 36, pp. 369–408, 1965.

[8] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.

[9] O. Zeitouni and M. Gutman, "On universal hypotheses testing via large deviations," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 285–290, 1991.

[10] J. Kelley, Jr., "The cutting-plane method for solving convex programs," *J. Society Industrial Appl. Mathematics*, vol. 8, no. 4, pp. 703–712, 1960.