Lemma 5: For a given k, Q_D is a concave function of Q_F.

Proof: It suffices to show that $\frac{d^2 Q_D}{dQ_F^2} \leq 0$. Recalling from the proof of Lemma 1, we have

$$\frac{dQ_D}{dQ_F} = \frac{dQ_D}{d\lambda} \bigg/ \frac{dQ_F}{d\lambda} = e^{r(\lambda,\, k)}.$$

Therefore,

$$\begin{aligned} \frac{d^2 Q_D}{dQ_F^2} &= \frac{de^{r(\lambda,\, k)}}{dQ_F} = \frac{de^{r(\lambda,\, k)}}{d\lambda} \bigg/ \frac{dQ_F}{d\lambda} \\ &= e^{r(\lambda,\, k)} \cdot \frac{dr(\lambda,\, k)}{d\lambda} \bigg/ \frac{dQ_F}{d\lambda}. \end{aligned}$$

Since $\frac{dP_F}{d\lambda} \leq 0$, we have $\frac{dQ_F}{d\lambda} \leq 0$. Recalling that $\frac{dr(\lambda,\, k)}{d\lambda} \geq 0$, we have $\frac{d^2 Q_D}{dQ_F^2} \leq 0$. □

The concavity of the ROC for a fixed k-out-of-n fusion rule ensures that the Lagrange multiplier method can be used to uniquely determine the optimal threshold in the case considered here.

V. Summary

We considered the problem of distributed binary hypothesis testing with independent identical sensors. The goal was to find the optimal k-out-of-n fusion rule and the optimal likelihood ratio threshold test for the sensors according to a performance criterion.

For the Bayesian detection problem, we showed that the objective function possesses the property of quasi-convexity, which secures the unique (global) optimum. We then developed a SECANT type of algorithm to efficiently compute the optimum. For Neyman–Pearson detection problem, we showed that the quasi-convexity exists and gives good reason for the use of the Lagrange multiplier method.

References

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

[2] J. Han, P. K. Varshney, and R. Srinivasan, “Distributed binary integration,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 2–8, Jan. 1993.

[3] F. Gini, F. Lombardini, and L. Verrazzani, “Decentralized CFAR detection with binary integration in Weibull clutter,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 396–407, Apr. 1997.

[4] J. V. DiFranco and W. L. Rubin, *Radar Detection*. Englewood Cliffs, NJ: Prentice-Hall, 1968.

[5] S. A. Kassam, “Optimum quantization for signal detection,” *IEEE Trans. Commun.*, vol. COM-25, pp. 479–484, May 1977.

[6] H. V. Poor and J.B. Thomas, “Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems,” *IEEE Trans. Commun.*, vol. COM-25, pp. 893–900, Sept. 1977.

[7] J. N. Tsitsiklis, “Decentralized detection by a large number of sensors,” *Math. Contr., Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[8] P. Chen and A. Papamarcou, “New asymptotic results in parallel distributed detection,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1847–1863, Nov. 1993.

[9] W. Shi, T. W. Sun, and R. D. Wesel, “Quasiconvexity and optimal binary fusion for distributed detection with identical sensors in generalized Gaussian noise,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 446–450, Jan. 2001.

[10] L. Hörmander, *Notions of Convexity*. Boston, MA: Birkhäuser, 1994.

[11] W. H. Press *et al.*, *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge Univ. Press, 1986.

[12] J. N. Tsitsiklis, “Decentralized detection,” *Adv. Statist. Signal Processing*, vol. 2, pp. 297–344, 1993.

[13] Y. I. Han and T. Kim, “Randomized fusion rules can be optimal in distributed Neyman–Pearson detectors,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 1281–1288, July 1997.

[14] Q. Zhang, “Efficient computational algorithms for the design of distributed detection networks,” Ph.D. dissertation, Syracuse Univ., Syracuse, NY, 2000.

A Note on Robust Hypothesis Testing

Luc Devroye, László Györfi, *Fellow, IEEE*, and Gábor Lugosi, *Member, IEEE*

***Abstract*—We introduce a simple new hypothesis testing procedure, which, based on an independent sample drawn from a certain density, detects which of k nominal densities is the true density closest to, under the total variation (L_1) distance. We obtain a density-free uniform exponential bound for the probability of false detection.**

***Index Terms*—Robust detection, robust hypotheses testing.**

I. Result

A model of robust hypothesis testing may be formulated as follows: let $f^{(1)}, \ldots, f^{(k)}$ be fixed densities on $\mathbb{R}^d$ which are the nominal densities under k hypotheses. We observe independent and identically distributed (i.i.d.) random vectors $X_1, \ldots, X_n$ according to a common density f. Under the hypothesis H_j $(j = 1, \ldots, k)$ the density f is a distorted version of $f^{(j)}$. This notion may be formalized in various ways. In this correspondence, we assume that the true density lies within a certain total variation distance of the underlying nominal density. More precisely, we assume that there exists a positive number ϵ such that for some $j \in \{1, \ldots, k\}$

$$\left\| f - f^{(j)} \right\| \leq \Delta_j - \epsilon$$

where

$$\Delta_j \stackrel{\text{def}}{=} (1/2) \min_{i \neq j} \| f^{(i)} - f^{(j)} \|.$$

Here $\|f - g\| = \int |f - g|$ denotes the L_1 distance between two densities. Thus, we formally define the k hypotheses by (see Fig. 1)

$$H_j = \left\{ f \colon \left\| f - f^{(j)} \right\| \leq \Delta_j - \epsilon \right\}, \qquad j = 1, \ldots, k.$$

The goal is to construct tests which, with high probability, assign to the observed sample the index j of the correct nominal density, that is, determines to which hypothesis H_j the density f belongs to.

Manuscript received August 1, 2000; revised July 1, 2001. The work of G. Lugosi was supported by DGI under Grant BMF2000-0807.

L. Devroye is with the School of Computer Sciences, McGill University, Montreal, QC H3A 2K6, Canada (e-mail: luc@kriek.cs.mcgill.ca).

L. Györfi is with the Department of Computer Science and Information Theory, Technical University of Budapest, Budapest, Hungary (e-mail: gyorfi@inf.bme.hu).

G. Lugosi is with the Department of Economics, Pompeu Fabra University, 08005 Barcelona, Spain (e-mail: lugosi@upf.es).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(02)05147-7.

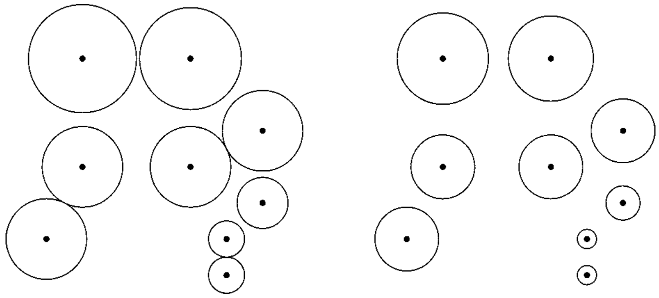

Fig. 1. The hypothesis classes $\boldsymbol{H}_j$ are illustrated here for $\boldsymbol{k = 9}$ with $\boldsymbol{\epsilon = 0}$ on the left and $\boldsymbol{\epsilon > 0}$ on the right. The centers of the balls represent the nominal densities $\boldsymbol{f}^{(j)}$.

Perhaps the most standard testing method is maximum likelihood, which accepts the jth nominal density $f^{(j)}$ if

$$\prod_{\ell=1}^{n} \frac{f^{(j)}(X_\ell)}{f^{(i)}(X_\ell)} > 1, \qquad \text{for all } i \neq j.$$

It is easy to see that this method is not robust in the sense that arbitrarily small deviations from the nominal density may cause a catastrophic behavior. We provide a simple example in Section II. In the special case when $k = 2$, a remarkable result of Huber [11] shows that a simple modification of the maximum-likelihood test is optimal in the minimax sense, that is, it minimizes the worst case probability of error in the given model. More precisely, Huber's test uses the modified likelihood ratio

$$\prod_{\ell=1}^{n} \max\left[c,\, \min\left(c',\, \frac{f^{(1)}(X_\ell)}{f^{(2)}(X_\ell)}\right)\right]$$

for constants c, c' which depend on the nominal densities. The disadvantage of Huber's test is that the values of these constants are given implicitly only and determining them may be problematic, especially when $d > 1$. Also, Huber's result does not cover the case $k > 2$ and it does not provide nonasymptotic bounds for the probability of error.

Other attempts for constructing robust tests involve nonparametric estimates of the underlying density f and decisions based on its distance from the nominal densities (see, e.g., [8], [14]). However, due to the fact that the L_1 error of any density estimate is bounded away from zero for some densities at any sample size (see [2]), it seems unlikely that the error of these tests can be bounded uniformly for any $f \in \bigcup_{j=1}^{k} H_j$, or at least such a result seems to be very difficult to prove.

The purpose of this correspondence is to introduce a new, simple, explicit testing procedure with a uniform nonasymptotic exponential bound for the probability of error. For surveys on robust statistics we refer the reader to [11] and [9].

In order to define the proposed test, introduce the empirical measure

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{X_i \in A}$$

where $\mathbb{I}$ denotes the indicator function and A is a Borel set. Let $\mathcal{A}$ denote the collection of $k(k-1)/2$ sets of the form

$$A_{i,j} = \left\{x: f^{(i)}(x) > f^{(j)}(x)\right\}, \qquad 1 \le i < j \le k.$$

The proposed test is the following: accept hypothesis H_j if

$$\max_{A\in\mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| = \min_{i=1,\ldots,k} \max_{A\in\mathcal{A}} \left| \int_A f^{(i)} - \mu_n(A) \right|.$$

(In case there are several indexes achieving the minimum, choose the smallest one.) The main result of this correspondence is as follows.

Theorem 1: For any $f \in \bigcup_{j=1}^{k} H_j$

$$\mathbb{P}\{\text{error}\} \le 2k(k-1)^2 e^{-n\epsilon^2/2}.$$

Proof: Without loss of generality, assume that $f \in H_1$. Recall that by Scheffé's theorem half of the L_1 distance equals the total variation distance

$$\begin{aligned}\|f-g\| &= 2 \sup_{A\subset\mathbb{R}^d} \left| \int_A f - \int_A g \right| \\ &= 2\left(\int_{\{x:\, f(x)>g(x)\}} f - \int_{\{x:\, f(x)>g(x)\}} g \right)\end{aligned}$$

where the supremum is taken over all Borel sets of $\mathbb{R}^d$. By Scheffé's theorem

$$\begin{aligned}2\max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(1)} \right| &\le \left\| f - f^{(1)} \right\| \\ &\le \Delta_1 - \epsilon \\ &\le \tfrac{1}{2} \left\| f^{(1)} - f^{(j)} \right\| - \epsilon \\ &= \max_{A\in\mathcal{A}} \left| \int_A f^{(1)} - \int_A f^{(j)} \right| - \epsilon \\ &\le \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(1)} \right| \\ &\quad + \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(j)} \right| - \epsilon\end{aligned}$$

by the triangle inequality. Rearranging the obtained inequality, we get that

$$\max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(1)} \right| \le \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(j)} \right| - \epsilon.$$

Therefore,

$$\begin{aligned}\mathbb{P}\{\text{error}\} &= \mathbb{P}\Bigg\{ \exists\, j > 1: \max_{A\in\mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| \\ &\qquad < \max_{A\in\mathcal{A}} \left| \int_A f^{(1)} - \mu_n(A) \right| \Bigg\} \\ &\le (k-1) \max_{j>1} \mathbb{P}\Bigg\{ \max_{A\in\mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| \\ &\qquad < \max_{A\in\mathcal{A}} \left| \int_A f^{(1)} - \mu_n(A) \right| \Bigg\} \\ &= (k-1) \max_{j>1} \mathbb{P}\Bigg\{ \max_{A\in\mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| \\ &\quad - \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(1)} \right| < \max_{A\in\mathcal{A}} \left| \int_A f^{(1)} - \mu_n(A) \right| \\ &\quad - \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(1)} \right| \Bigg\} \\ &\le (k-1) \max_{j>1} \mathbb{P}\Bigg\{ \max_{A\in\mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| \\ &\quad - \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(j)} \right| + \epsilon < \max_{A\in\mathcal{A}} \left| \int_A f^{(1)} - \mu_n(A) \right| \\ &\quad - \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(1)} \right| \Bigg\} \\ &\qquad \text{(by the inequality derived above)} \\ &\le (k-1) \max_{j>1} \mathbb{P}\Bigg\{ \left| \max_{A\in\mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| \right. \\ &\quad \left. - \max_{A\in\mathcal{A}} \left| \int_A f - \int_A f^{(j)} \right| \right| > \frac{\epsilon}{2} \Bigg\}\end{aligned}$$

$$
\begin{aligned}
&+ (k-1)\mathbb{P}\left\{\left|\max_{A\in\mathcal{A}}\left|\int_A f^{(1)} - \mu_n(A)\right|\right.\right.\\
&\qquad \left.\left. - \max_{A\in\mathcal{A}}\left|\int_A f - \int_A f^{(1)}\right|\right| > \frac{\epsilon}{2}\right\}\\
&\le 2(k-1)\mathbb{P}\left\{\max_{A\in\mathcal{A}}\left|\int_A f - \mu_n(A)\right| > \frac{\epsilon}{2}\right\}\\
&\qquad \text{(by a double application of the triangle inequality)}\\
&\le 2(k-1)|\mathcal{A}|\max_{A\in\mathcal{A}}\mathbb{P}\left\{\left|\int_A f - \mu_n(A)\right| > \frac{\epsilon}{2}\right\}\\
&\le 2k(k-1)^2 e^{-n\epsilon^2/2},
\end{aligned}
$$

where in the last step we used Hoeffding's inequality [10]. □

II. Discussion

Methodology: The methodology of the proposed test is close in spirit to Yatracos' minimum distance parametric density estimate, see [13], [5]–[7].

Computation: The hypothesis-testing method proposed above is computationally quite simple. The sets $A_{i,j}$ and the integrals $\int_A f^{(j)}$ may be computed and stored before seeing the data. Then one merely needs to calculate $\mu_n(A)$ for all $A \in \mathcal{A}$ and compute the test statistics requiring $O(nk^2 + k^2 \log k)$ time. In many applications $k = 2$. In these cases, the test becomes especially simple as the class $\mathcal{A}$ contains just one set.

Robustness: Note that the theorem does not require any assumption for the nominal densities. (In fact, the result may be formulated in a similar fashion without even assuming the existence of the densities.) The test is robust in a very strong sense: we obtain uniform exponential bounds for the probability of failure under the sole assumption that the distorted density remains within a certain total variation distance of the nominal density.

Additive Noise: We illustrate the power of the proposed method on a very simple example showing that the test has an exponentially small probability of error if the nominal density is corrupted by an arbitrary additive noise of a sufficiently small support. Consider k nominal densities $f^{(1)}, \ldots, f^{(k)}$ and assume that the observations are distributed according to one of the nominal densities corrupted by an additive noise. Thus, assume that the X_i's are distributed according to density $f = f^{(1)} \star g$, where the nominal density $f^{(1)}$ is now assumed to be Lipschitz (i.e., $|f^{(1)}(x) - f^{(1)}(y)| \le c|x-y|$ for some $c > 0$ for all $x, y \in \mathbb{R}$), supported on the bounded set $[-M, M]$, and the density g of the additive noise is assumed to have support in the interval $[-r, r]$, where r is thought of as a small number. The other $k-1$ nominal densities are arbitrary. Then, according to the theorem, the proposed test is correct with probability larger than $1 - 2k(k-1)^2 e^{-n\epsilon^2/2}$ as long as $\|f - f^{(1)}\| \le \Delta_1 - \epsilon$. But

$$
\begin{aligned}
\left\|f - f^{(1)}\right\| &= \int\left|\int f^{(1)}(x-y)g(y)\,dy - \int f^{(1)}(x)g(y)\,dy\right|dx\\
&\le \iint\left|f^{(1)}(x-y) - f^{(1)}(x)\right|g(y)\,dy\,dx\\
&\le \int_{-M-r}^{M+r}\int c|y|g(y)\,dy\,dx\\
&\le 2c(M+r)r.
\end{aligned}
$$

Thus, the condition is satisfied if r is so small that

$$r \le (\Delta_1 - \epsilon)/2c(M+r).$$

This is the only assumption on the noise density g, otherwise it may be completely arbitrary! (Note that boundedness of the support of g is not a necessary condition; we assumed it to simplify the example.)

Maximum Likelihood Does Not Work: Here we show a simple example to demonstrate that the maximum-likelihood test does not share the proved property of the proposed test. Indeed, consider the case when $k = 2$, and the two nominal densities are standard normal and standard Cauchy densities, that is,

$$f^{(1)}(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2} \quad \text{and} \quad f^{(2)}(x) = \frac{1}{\pi(1+x^2)}.$$

Assume that the data are distributed according to $f = f^{(1)} \star g_c$, where the density of the additive noise is Cauchy

$$g_c(x) = \frac{1}{\pi c(1 + (x/c)^2)}$$

where c is a small positive constant. It is well known (see, e.g., [4]) that $\|f^{(1)} - f^{(1)} \star g_c\| \to 0$ as $c \to 0$, and, therefore, for a sufficiently small c, the L_1 distance between f and $f^{(1)}$ can be made arbitrarily small, in particular, $\|f^{(1)} - f\| < \|f^{(1)} - f^{(2)}\|/2 - \epsilon$. Nevertheless, it is easy to show that for any small c, the probability of error of the maximum-likelihood detector converges to one. Indeed, on the one hand

$$
\begin{aligned}
\mathbb{E}\left\{\frac{1}{n}\sum_{\ell=1}^{n}\log f^{(1)}(X_\ell)\right\} &= \int f(x)\log f^{(1)}(x)\,dx\\
&= -\log\sqrt{2\pi} - \tfrac{1}{2}\int f(x)x^2\,dx\\
&= -\infty
\end{aligned}
$$

and on the other hand

$$
\begin{aligned}
\mathbb{E}\left\{\frac{1}{n}\sum_{\ell=1}^{n}\log f^{(2)}(X_\ell)\right\} &= \int f(x)\log f^{(2)}(x)\,dx\\
&= -\log\pi - \int f(x)\log(1+x^2)\,dx\\
&> -\infty.
\end{aligned}
$$

Therefore, the strong law of large numbers implies that for sufficiently large n, the maximum-likelihood detector errs with probability one.

Tests Based on Density Estimates: An alternative way of performing robust tests is based on estimating the density. Indeed, such methods have been proposed in the literature, see [8], [14]. These tests cannot compete with the simplicity of the proposed method, and no uniform exponential bound for their probability of error is available. However, hypothesis testing based on density estimates may be necessary if even larger hypothesis classes need to be considered. A stronger notion of robust hypothesis testing is obtained if one requires good testing whenever the true density is closer to the nominal density than to any other density in the finite collection. Formally, this leads to the hypotheses

$$\overline{H}_j = \left\{f\colon \left\|f - f^{(j)}\right\| < \min_{i\ne j}\left\|f - f^{(i)}\right\|\right\}, \qquad j = 1, \ldots, k$$

that is, the sets $\overline{H}_j$ form a Voronoi partition of the set of all densities. This problem may be solved by using a nonparametric estimate f_n of f and accepting $\overline{H}_j$ if $\|f_n - f^{(j)}\|$ is minimal among the $\|f_n - f^{(i)}\|$, $i = 1, \ldots, k$. (Break ties by selecting the smallest index.) A suitable choice is the kernel estimate defined by

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

where $K\colon \mathbb{R}^d \to \mathbb{R}^+$ is a fixed kernel function with $\int K = 1$, $h > 0$ is a smoothing factor, and $K_h(\cdot) = (1/h^d)K(\cdot/h)$. If h is chosen such that $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$, then it is well known (see [4]) that the estimate is universally consistent, that is, $\mathbb{E}\|f_n - f\| \to 0$ for any density. Also, Devroye [3] shows that for any $\epsilon > 0$

$$\mathbb{P}\{\|f_n - f\| - \mathbb{E}\|f_n - f\| \geq \epsilon\} \leq e^{-n\epsilon^2/2}.$$

Using these properties, it is easy to see that the testing method based on the kernel density estimate is consistent in the sense that the probability of error converges to zero exponentially for all $f \in \bigcup_{j=1}^{k} \overline{H}_j$. In order to show this, suppose that $f \in \overline{H}_1$, and put

$$\epsilon = \min_{j>1} \left\|f - f^{(j)}\right\| - \left\|f - f^{(1)}\right\|.$$

Then

$$\begin{aligned}
\mathbb{P}\{\text{error}\} &\leq \mathbb{P}\left\{\exists\, j > 1\colon \left\|f_n - f^{(1)}\right\| \geq \left\|f_n - f^{(j)}\right\|\right\} \\
&\leq (k-1)\max_{j>1} \mathbb{P}\left\{\left\|f_n - f^{(1)}\right\| \geq \left\|f_n - f^{(j)}\right\|\right\} \\
&\leq (k-1)\max_{j>1} \mathbb{P}\left\{\|f_n - f\| + \left\|f - f^{(1)}\right\| \right. \\
&\qquad\qquad \left. \geq \left\|f - f^{(j)}\right\| - \|f_n - f\|\right\} \\
&\leq (k-1)\mathbb{P}\{2\|f_n - f\| \geq \epsilon\} \\
&= (k-1)\mathbb{P}\{\|f_n - f\| - \mathbb{E}\|f_n - f\| \\
&\qquad\qquad \geq \epsilon/2 - \mathbb{E}\|f_n - f\|\} \\
&\leq (k-1)e^{-n/2([\epsilon/2 - \mathbb{E}\|f_n - f\|]^+)^2}
\end{aligned}$$

where the last inequality follows from the previously mentioned inequality of Devroye [3]. The consistency of f_n assures that for a sufficiently large n, $\mathbb{E}\|f_n - f\| < \epsilon/4$ and for such n, $\mathbb{P}\{\text{error}\} \leq (k-1)e^{-n\epsilon^2/32}$. However, since $\mathbb{E}\|f_n - f\|$ may tend to zero at an arbitrarily slow rate (see [2]), the error exponent is not uniform: it depends on f. It is known (see [1], [12]) that for the hypotheses $\overline{H}_j$ it is impossible to construct a test with a uniform error exponent.

Acknowledgment

The authors wish to thank the reviewers for drawing their attention to relevant literature.

References

[1] A. R. Barron, "Uniformly powerful goodness of fit tests," *Ann. Statist.*, vol. 17, pp. 107–124, 1989.

[2] L. Devroye, "On arbitrary slow rates of global convergence in density estimation," *Z. Wahscheinlichkeitstheorie und verwandte Gebiete*, vol. 62, pp. 475–483, 1983.

[3] ——, "Exponential inequalities in nonparametric estimation," in *Nonparametric Functional Estimation*. ser. NATO ASI, G. Roussas, Ed. Dordrecht, The Netherlands: Kluwer, 1991, pp. 31–44.

[4] L. Devroye and L. Györfi, *Nonparametric Density Estimation: The $\boldsymbol{L}_1$ View*. New York: Wiley, 1985.

[5] L. Devroye and G. Lugosi, "A universally acceptable smoothing factor for kernel density estimates," *Ann. Statist.*, vol. 24, pp. 2499–2512, 1996.

[6] ——, "Nonasymptotic universal smoothing factors, kernel complexity, and Yatracos classes," *Ann. Statist.*, vol. 25, pp. 2626–2635, 1997.

[7] ——, *Combinatorial Methods in Density Estimation*. New York: Springer-Verlag, 2000.

[8] L. Györfi and E. C. van der Meulen, "A consistent goodness of fit test based on the total variation distance," in *Nonparametric Functional Estimation and Related Topics*, G. Roussas, Ed. Dordrecht, The Netherlands: Kluwer, 1991, pp. 631–646.

[9] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.

[10] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, vol. 58, pp. 13–30, 1963.

[11] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[12] L. Le Cam, "Convergence of estimates under dimensionality restrictions," *Ann. Statist.*, vol. 1, pp. 38–53, 1973.

[13] Y. G. Yatracos, "Rates of convergence of minimum distance estimators and Kolmogorov's entropy," *Ann. Statist.*, vol. 13, pp. 768–774, 1985.

[14] S. M. Zabin and G. A. Wright, "Nonparametric density estimation and detection in impulsive interference channels—Part II: Detectors," *IEEE Trans. Commun.*, vol. 42, pp. 1698–1711, Feb./March/Apr. 1994.

Linear MMSE Multiuser Receivers: MAI Conditional Weak Convergence and Network Capacity

Junshan Zhang, *Member, IEEE,* and
Edwin K. P. Chong, *Senior Member, IEEE*

***Abstract*—We explore the performance of minimum mean-square error (MMSE) multiuser receivers in wireless systems where the signatures are modeled as random and take values in complex space. First we study the conditional distribution of the output multiple-access interference (MAI) of the MMSE receiver. By appealing to the notion of *conditional weak convergence*, we find that the conditional distribution of the output MAI, given the received signatures and received powers, converges in probability to a proper complex Gaussian distribution that does not depend on the signatures. This result indicates that, in a large system, the output interference of the MMSE receiver is approximately Gaussian with high probability, and that systems with MMSE receivers are robust to the randomness of the signatures. Building on the Gaussianity of the output interference, we then take the quality of service (QoS) requirements as meeting the signal-to-interference ratio (SIR) constraints and identify the network capacity of single-class systems with random spreading. The network capacity is expressed uniquely in terms of the SIR requirements and received power distributions. Compared to the network capacity corresponding to the optimal signature allocation, we conclude that at the cost of transmission power, the gap between the network capacity corresponding to optimal signatures and that corresponding to random signatures can be made arbitrarily small. Therefore, from the viewpoint of network capacity, systems with MMSE receivers are robust to the randomness of signatures.**

***Index Terms*—Central limit theorem, conditional weak convergence, martingale difference array, minimum mean-square error (MMSE) receiver, proper complex random variable, random signature.**

I. Introduction

Consider a K-user communication system equipped with linear minimum mean-square error (MMSE) multiuser receivers.[1] We focus primarily on the following discrete-time synchronous baseband model

Manuscript received May 1, 2000; revised September 1, 2001. The work of E. K. P. Chong was supported in part by the National Science Foundation under Grant ECS-9501652. The material in this paper was presented in part at the 38th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 4–6, 2000.

J. Zhang is with the Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: junshan.zhang@asu.edu).

E. K. P. Chong is with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523-1373 USA (e-mail: echong@engr.colostate.edu).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(02)05162-3.

[1] As in [3], we assume coherent demodulation in this paper.