The Annals of Statistics
1989, Vol. 17, No. 3, 1325–1334

OPTIMAL-PARTITIONING INEQUALITIES IN CLASSIFICATION AND MULTI-HYPOTHESES TESTING

BY THEODORE P. HILL[1] AND Y. L. TONG[2]

Georgia Institute of Technology

Optimal-partitioning and minimax risk inequalities are obtained for the classification and multi-hypotheses testing problems. Best possible bounds are derived for the minimax risk for location parameter families, based on the tail concentrations and Lévy concentrations of the distributions. Special attention is given to continuous distributions with the maximum likelihood ratio property and to symmetric unimodal continuous distributions. Bounds for general (including discontinuous) distributions are also obtained.

1. Preliminaries. The statistical classification problem, in its standard form, deals with optimal decision rules for classifying an observation into one of several specified populations. The problem is closely related to the following multi-hypotheses testing problem: For $n \geq 2$, let $F_1, \dots, F_n$ be given (univariate) distributions. Let X be a random variable with distribution F. In testing the hypotheses

$$H_i\colon F = F_i, \qquad i = 1, \dots, n, \tag{1.1}$$

a decision rule corresponds to a measurable partition $\{A_i\}_{i=1}^n$ of the real line such that H_i is accepted iff $X \in A_i$. The main purpose of this paper is to use optimal-partitioning results for densities with the monotone likelihood ratio (MLR) property together with convexity to derive some best-possible inequalities for the minimax risk, in terms of two probability-concentration parameters (the tail-d concentration, Definition 2.1 below and the Lévy concentration, Definition 2.4) of continuous distributions, for general location parameter families and for symmetric unimodal densities (Section 2). Analogous results for discontinuous distributions are then given (Section 3).

For the objective of minimizing the largest probability of misclassification, the standard classification problem is equivalent to many "fair-division" problems in which there are n probability measures $\mu_1, \dots, \mu_n$ defined on the same space, and the objective is to partition the space so as to maximize the minimum share, i.e., to find an ordered measurable partition $(A_1^*, \dots, A_n^*)$ which attains or nearly attains

$$C^*(\mu) = \sup\left\{ \min_{1 \leq i \leq n} \mu_i(A_i)\colon (A_1, \dots, A_n) \text{ is a measurable partition of } \Omega \right\},$$

Received April 1987; revised July 1988.
[1]Research partially supported by NSF Grants DMS-84-01604, DMS-86-01608 and DMS-87-01691.
[2]Research partially supported by NSF Grants DMS-85-02346 and DMS-88-01327.
AMS 1980 *subject classifications.* 60E15, 62H30, 28B05.
Key words and phrases. Optimal-partitioning inequalities, minimax risk, classification and discriminant analysis, multi-hypotheses testing, convexity theorem, concentration function, tail concentration.

where $\mu_i(A)$ represents the value of portion A to the ith individual [the reader is referred to Dubins and Spanier (1961) for details]. Consequently, the minimax-risk results derived below may also be interpreted as results for the fair-division problem and, in fact, it is in the fair-division or C^*-terminology that most of the proofs will be given.

Throughout this paper, $F_1, \ldots, F_n$ are distinct distribution functions with corresponding densities $f_1, \ldots, f_n$ and probability measures $\mu_1, \ldots, \mu_n$, respectively; $\mathbb{B}$ are the Borels on $\mathbb{R}$ and Π_n is the collection of ordered $\mathbb{B}$-measurable n-partitions of $\mathbb{R}$, that is,

$$\Pi_n = \left\{(A_i)_{i=1}^n \colon A_i \in \mathbb{B}\ \forall\, i,\ \bigcup_{i=1}^{n} A_i = \mathbb{R} \text{ and } A_i \cap A_j = \phi \text{ if } i \neq j\right\}.$$

For brevity, write $\mathbf{A} = (A_1, \ldots, A_n)$ for $(A_i)_1^n \in \Pi_n$, $\mathbf{F} = (F_1, \ldots, F_n)$, $\mathbf{f} = (f_1, \ldots, f_n)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and $\boldsymbol{\mu}(\mathbf{A}) = (\mu_1(A_1), \ldots, \mu_n(A_n)) \in [0,1]^n$. The *partition range* of $\boldsymbol{\mu}$, $\mathrm{PR}(\boldsymbol{\mu})$ is the subset of $[0,1]^n$ defined by $\mathrm{PR}(\boldsymbol{\mu}) = \{\boldsymbol{\mu}(\mathbf{A})\colon \mathbf{A} \in \Pi_n\}$.

One of the main tools in this paper is the following generalization of Lyapounov's convexity theorem.

Proposition 1.1 [Dvoretzky, Wald and Wolfowitz (1951)]. (i) $\mathrm{PR}(\boldsymbol{\mu})$ *is compact.*

(ii) *If* $\mu_1, \ldots, \mu_n$ *are nonatomic, then* $\mathrm{PR}(\boldsymbol{\mu})$ *is convex.*

Definition 1.2. A partition $\mathbf{A} \in \Pi_n$ is *optimal* for $\boldsymbol{\mu}$ if

$$\boldsymbol{\mu}(\mathbf{A}^*) = C^*(\boldsymbol{\mu}). \tag{1.2}$$

Corollary 1.3 [cf. Dubins and Spanier (1961)]. *Optimal partitions exist for all* $\boldsymbol{\mu}$ *(and* $\mathbf{F}$*).*

Definition 1.4. A partition $\mathbf{A} = (A_1, \ldots, A_n)$ has *equal risks* for $\boldsymbol{\mu}$ if

$$\mu_1(A_1) = \cdots = \mu_n(A_n). \tag{1.3}$$

The next two results follow from a standard "mass-shifting" argument.

Theorem 1.5. *If* $F_1, \ldots, F_n$ *are continuous, then there exists an optimal partition* $\mathbf{A}^* \in \Pi_n$ *with equal risks for* $\boldsymbol{\mu}$. *If, in addition, the* $\{F_i\}$ *have common support, then every optimal partition for* $\mathbf{F}$ *has equal risks.*

It should be observed that for some discrete distributions, no optimal partition has equal risks. Also, for continuous distributions that do not have common support, not all optimal partitions may have equal risks as can be easily seen by considering partially overlapped uniform distributions.

For distributions whose density functions possess the MLR property, the following result is a direct consequence of Karlin and Rubin (1956).

THEOREM 1.6. *If $F_1, \ldots, F_n$ are continuous with densities $\{f_i\}_1^n$ having the MLR property, then there exist real numbers $d_1^* < \cdots < d_{n-1}^*$ so that*

$$\mathbf{A}^* = ((-\infty, d_1^*], (d_1^*, d_2^*], \ldots, (d_{n-1}^*, \infty)) \tag{1.4}$$

is optimal for $F_1, \ldots, F_n$ and has equal risks, that is,

$$\int_{-\infty}^{d_1^*} f_1 = \int_{d_1^*}^{d_2^*} f_2 = \cdots = \int_{d_{n-1}^*}^{\infty} f_n = C^*(\mu). \tag{1.5}$$

2. Location parameter families for continuous distributions. In this section, F is a continuous distribution function, $F_i(x) = f(x - \theta_i)$, $i = 1, \ldots, n$, and $\theta_1 < \theta_2 < \cdots < \theta_n$ are the location parameters. The functional form of F and the values of the θ_i's are assumed to be known, so, without loss of generality θ_1 is assumed to be 0. The main purpose is to derive some best possible universal lower bounds for the smallest probability of correct decision (or equivalently, upper bounds for the largest probability of misclassification). The bounds will be given in terms of two concentration parameters of the densities $\{f_i(x) = f(x - \theta_i)\}$ and most of the results will be stated for the equally spaced configuration with $\theta_i = (i - 1)d$ for fixed $d > 0$. A more general result is then obtained under the additional assumption that $\{f(x - \theta_i)\}$ possesses the MLR property.

DEFINITION 2.1. The *tail-d concentration* of F, $\rho(F, d)$, is defined by

$$\rho(F, d) = \max\{\mu((-\infty, \operatorname{ess\,inf} F + d]), \mu([\operatorname{ess\,sup} F - d, \infty))\}. \tag{2.1}$$

Note that if F_1, F_2 are two continuous distributions such that $F_1(x) = F_2(x - d)$ and $a = \operatorname{ess\,inf} F_1$, $b = \operatorname{ess\,sup} F_1$, then

$$\mu_1((b, b + d]) = \mu_2([a, a + d)) = 0.$$

So under an optimal classification rule $\mathbf{A}^* = (A_1^*, A_2^*)$ one has $(b, b + d] \subset A_2^*$ a.s. and $[a, a + d) \subset A_1^*$ a.s. Furthermore, note that $\rho(F, d) = 0$ if and only if $\operatorname{ess\,inf} F = -\infty$ and $\operatorname{ess\,sup} F = \infty$.

THEOREM 2.2. *If F is continuous and $F_i(x) = F(x - (i - 1)d)$ for $i = 1, \ldots, n$, then*

$$C^*(\mu) \geq \left(1 + \sum_{j=1}^{n-1} q^j\right)^{-1}, \tag{2.2}$$

where $q = 1 - \rho(F, d)$. Moreover, this bound is best possible and is attained for all n, all d and all $q < 1$.

PROOF. If $\rho(F, d) = 0$, then the bound in (2.2) is n^{-1} and the inequality (2.2) holds for *any* continuous distributions $\mu_1, \ldots, \mu_n$, as follows easily from Proposition 1.1(ii). (Even more is true: Neyman's (1946) solution of Fisher's "Problem of the Nile" [Fisher (1936)] even shows that $\mathbf{A}$ may be chosen so that $\mu_i(A_j) = n^{-1}$ for all i and j.) On the other hand, if $\rho(F, d) = 1$, then the bound in (2.2) is 1, which follows trivially since in this case the distributions $F_1, \ldots, F_n$ have essentially disjoint support. Suppose that $\rho = \rho(F, d) \in (0,1)$. Since $\rho > 0$ implies $\operatorname{ess\,inf} F > -\infty$ or $\operatorname{ess\,sup} F < \infty$, it may be assumed (by translation) that one of these, say $\operatorname{ess\,inf} F$, is zero and also that $\mu_1((-\infty, d]) = \rho$.

For each $k = 1, \ldots, n$, let $\mathbf{A}_k \in \Pi_n$ be defined by

$$\mathbf{A}_1 = ((-\infty, \infty), \phi, \ldots, \phi)$$

and

$$\mathbf{A}_k = ((-\infty, d], (d, 2d], \ldots, ((k-1)d, \infty), \phi, \ldots, \phi) \quad \text{for } k > 1.$$

Then

$$\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathrm{PR}(\mu), \tag{2.3}$$

where $\mathbf{a}_k = \mu(\mathbf{A}_k) = (\rho, \ldots, \rho, 1, 0, \ldots, 0)$ is the vector in $\mathbb{R}^n$ with 1 in the kth coordinate and preceded by $k-1$ entries of ρ. Let $\beta_k = q^{n-k}/(1 + \sum_{j=1}^{n-1} q^j)$ for $k = 1, \ldots, n$. By (2.3) and Proposition 1.1(ii),

$$\mathbf{a} = \sum_{k=1}^{n} \beta_k \mathbf{a}_k \in \mathrm{PR}(\mu)$$

and an easy calculation shows that each entry of $\mathbf{a}$ is $(1 + \sum_{j=1}^{n-1} q^j)^{-1}$, which establishes (2.2).

To see that (2.2) is best possible for $q = 1$, let $F_1 = F_{1,M}$ be uniformly distributed on $[-M, M]$. Then as $M \to \infty$, $\rho(F, d) \to 0$ and $C^*(\mu) \to n^{-1}$. For $q = 0$, any distribution with support in $[0, d/2]$ attains the bound in (2.2). That (2.2) is attained for all n, all d and all $q \in (0,1)$ is shown by the next example. □

EXAMPLE 2.3. Let $F(x) = 1 - e^{-x}$ for $x > 0$ and for fixed $n > 1$ and $d > 0$ let $F_i(x) = F(x - (i-1)d)$ for $i = 1, \ldots, n$. Then the corresponding density functions are negative exponential with location parameters $\theta_i = (i-1)d$, i.e.,

$$f_i(x) = \exp(-(x - (i-1)d)) \quad \text{for } x \geq (i-1)d$$

and zero otherwise for $i = 1, \ldots, n$. Clearly $\{f_i\}$ has the MLR property, so by Theorem 1.6 there exist positive constants $d_1^* < d_2^* < \cdots < d_{n-1}^* < \infty$ satisfying (1.5).

It is also easy to see that $C^* > \rho$, so $d_1^* > d$ and inductively $d_k^* > kd$ for all $k > 1$. This implies that

$$f_1 = qf_2 = q^2 f_3 = \cdots = q^{j-1} f_j \quad \text{on } (d_{j-1}^*, d_j^*) \tag{2.4}$$

for $j = 2, \ldots, n$ ($d_0^* = 0$, $d_n^* = \infty$). Together (2.4) and (1.5) imply

$$q^{-j+1} \int_{d_{j-1}^*}^{d_j^*} f_1 = C^*(\mu) \quad \text{for } j = 1, \ldots, n. \tag{2.5}$$

Since $\sum_{j=1}^{n} \int_{d_{j-1}^*}^{d_j^*} f_1 = 1$, it follows from (2.5) that $C^*(\mu) = (1 + \sum_{j=1}^{n-1} q^j)^{-1}$.

If $n = 2$, the location parameter classification problem is precisely the problem of testing a simple null hypothesis H_1: $\theta = \theta_1$ against a simple alternative

H_2: $\theta = \theta_2$, where $\theta_2 = \theta_1 + d$ for some $d > 0$. In this case a sharp bound for the minimax risk in terms of the Lévy-concentration function is given by the next theorem.

DEFINITION 2.4. The *Lévy concentration* for F, d is

$$\lambda(F, d) = \sup_x \{F(x + d) - F(x)\} \in (0, 1].$$

THEOREM 2.5. *Let X have a continuous distribution $F(x - \theta)$ with Lévy concentration $\lambda = \lambda(F, d)$ and let θ_1, θ_2 be location parameters such that $\theta_2 - \theta_1 = d > 0$. Then there exists a test for testing*

$$H_1: \theta = \theta_1 \quad \textit{versus} \quad H_2: \theta = \theta_2 \tag{2.6}$$

which satisfies

$$\max\{\alpha, \beta\} \le (1 - \lambda)/(2 - \lambda), \tag{2.7}$$

where α, β are the type I *and type* II *errors, respectively. Moreover, this bound is attained for all d and all λ.*

REMARK. We note that, by definition, $\lambda = \lambda(F, d) \ge \rho(F, d)$ for all F and all $d > 0$ and equality holds for monotone density functions. If $n = 2$, then $1 - (1 + q)^{-1} = (1 - \rho)/(2 - \rho) \ge (1 - \lambda)/(2 - \lambda)$ always holds. Thus the bound in (2.7) is sharper than that in (2.2).

PROOF OF THEOREM 2.5. For notational convenience assume $\theta_1 = 0$. We show that there is a test with

$$C^*(\mu) = \min\{1 - \alpha, 1 - \beta\} \ge (2 - \lambda)^{-1}. \tag{2.8}$$

For fixed $d > 0$ and $\lambda = \lambda(F, d) \in (0, 1]$, $\lambda(F, d)$ is always attained [see, e.g., Theorem 1.1.8 of Hengartner and Theodorescu (1973)]. That is, there is a real number γ satisfying

$$F(\gamma + d) - F(\gamma) = \mu_F(\gamma, \gamma + d) = \lambda.$$

Let $r_1 = \mu_F(-\infty, \gamma)$ and $r_2 = \mu_F(\gamma + d, \infty)$ and assume without loss of generality (by symmetry) that $r_2 \ge r_1$. Considering the partitions $\mathbf{A}$ and $\mathbf{B} \in \Pi_2$ given by

$$\mathbf{A} = ((-\infty, \gamma + d), [\gamma + d, \infty)), \qquad \mathbf{B} = ((-\infty, \infty), \phi)$$

implies that

$$\{(\lambda + r_1, \lambda + r_2), (1, 0)\} \subset \mathrm{PR}(\mu). \tag{2.9}$$

Let

$$\beta = (r_2 - r_1)/(2r_2 + \lambda) = 1 - (2r_2 + \lambda)^{-1} \in [0, 1).$$

Then by (2.9) and Proposition 1.1(ii),

$$\beta(1, 0) + (1 - \beta)(\lambda + r_1, \lambda + r_2) = ((\lambda + r_2)/(2r_2 + \lambda), (\lambda + r_2)/(2r_2 + \lambda))$$

is in $\mathrm{PR}(\boldsymbol{\mu})$. But $\lambda + r_1 + r_2 = 1$ and $r_2 \geq r_1$, which implies $(1-\lambda)/2 \leq r_2 \leq 1-\lambda$. Consequently, for r_2 in this range, one has $(\lambda + r_2)/(2r_2 + \lambda) \geq (2-\lambda)^{-1}$ and (2.8) follows. That the lower bound $(2-\lambda)^{-1}$ is attained is shown by Example 2.3 with $n = 2$. □

Recall that a continuous distribution function F is said to be symmetric about b and unimodal if its density f satisfies

$$f(b+y) = f(b-y) \quad \text{for all } y \text{ and } \quad f(b+y) \downarrow 0 \quad \text{as} \quad y \to \infty.$$

The family of symmetric and unimodal distributions plays an important role in statistical applications. The next theorem gives a best possible bound, in terms of the Lévy concentration, for the minimax risk for the location-parameter problem with continuous, symmetric, unimodal distributions.

THEOREM 2.6. *If F is continuous, symmetric about b for some b and unimodal, and if $F_i(x) = F(x-(i-1)d)$ for fixed $d > 0$ and $i = 1, \dots, n$, then*

$$C^*(\boldsymbol{\mu}) \geq \left(1 + 2\sum_{j=1}^{m-1} \tau^j + (k+1)\tau^m\right)^{-1}, \tag{2.10}$$

where m is the largest integer less than or equal to $n/2$, $k = n - 2m$, $\tau = (1-\lambda)/(1+\lambda)$ and $\lambda = \lambda(F, d)$. Moreover, this bound is attained for all n, d and λ.

PROOF. CASE 1. $n = 2m$ for some $m \geq 1$. Using the symmetry of F and Definition 2.4, it is easy to see that

$$\{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subset \mathrm{PR}(\boldsymbol{\mu}), \tag{2.11}$$

where

$$\begin{aligned} \mathbf{v}_1 &= ((1+\lambda)/2, \lambda, \dots, \lambda, (1+\lambda)/2), \\ \mathbf{v}_2 &= (0, (1+\lambda)/2, \lambda, \dots, \lambda, (1+\lambda)/2, 0), \dots, \\ \mathbf{v}_m &= (0, \dots, 0, (1+\lambda)/2, (1+\lambda)/2, 0, \dots, 0). \end{aligned}$$

[For example, $\mathbf{v}_2 = (\mu_1(\phi), \mu_2((-\infty, b + 3d/2]), \mu_3((b + 3d/2, b + 5d/2]), \dots, \mu_n(\phi))$.] For $\tau = (1-\lambda)/(1+\lambda)$ define

$$\beta_j = \tau^{j-1} \Bigg/ \left(\sum_{i=0}^{m-1} \tau^i\right) \quad \text{for } j = 1, \dots, m.$$

Then $\beta_j \geq 0$ and $\sum_{j=1}^m \beta_j = 1$. It follows from (2.11) and Proposition 2.1(ii) that $\sum_{j=1}^m \beta_j \mathbf{v}_j = (c, c, c, \dots, c) \in \mathrm{PR}(\boldsymbol{\mu})$, where $c = (1 + 2\sum_{j=1}^{m-1} \tau^j + \tau^m)^{-1}$.

CASE 2. $n = 2m + 1$ for some $m \geq 1$. Proceed as in Case 1 using the additional vector $\mathbf{v}_{m+1} = (0, 0, \dots, 0, 1, 0, \dots, 0)$.

To see that these bounds are attained for all n, d and λ, consider the continuous symmetric (about $d/2$) unimodal distribution F with right-half density given by

$$f(x) = \lambda\tau^j \quad \text{for } x \in [jd, (j+1)d) \quad \text{for } j = 0, 1, 2, \dots .$$

Letting $f_i(x) = f(x - (i-1)d)$ for $i = 1, \dots, n$, note that $\{f_1, \dots, f_n\}$ have common support and the MLR property. Then proceed as in Example 2.3, using Theorem 1.6 to show that the bound in (2.10) is attained. □

REMARK. The authors believe that, for all $n > 2$, the conclusion of Theorem 2.2 is true even if $q = 1 - \rho(F, d)$ is replaced by $q = 1 - \lambda(F, d)$, which is a stronger result since $\lambda(F, d) \geq \rho(F, d)$. The Lévy concentration $\lambda(F, d)$ is, as is the variance, some gauge of how spread out the distribution F is, and analogous bounds for the minimax risk in terms of the *variance* of the distribution are also possible. Although the best possible bounds are not known to the authors, the bounds in Theorems 2.5 and 2.6 may be used to obtain corresponding minimax-risk inequalities in terms of the variance by applying inequalities of Lévy [e.g., Hengartner and Theodorescu (1973), pages 26–30] which give bounds on λ in terms of the variance and vice versa.

Thus far we have considered distributions with equally spaced location parameters, i.e., $\theta_i = (i-1)d$, $i = 1, \dots, n$. In the following we extend the results given in Theorems 2.2 and 2.6 to yield lower bounds for the more general case. Toward this end we first observe a lemma concerning a monotonicity property of the optimal partitioning problem.

LEMMA 2.7. *Let δ_i, δ_i' for $i = 1, \dots, n-1$ be positive real numbers and define $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, $\boldsymbol{\theta}' = (\theta_1', \dots, \theta_n')$, where $\theta_1 = \theta_1' = 0$, $\theta_i = \sum_{j=1}^{i-1} \delta_j$, $\theta_i' = \sum_{j=1}^{i-1} \delta_j'$ for $i = 2, \dots, n$. Let F be a continuous distribution function and define $\mathbf{F}$ and $\mathbf{F}'$ by*

$$\mathbf{F} = \big(F(x), F(x - \theta_2), \dots, F(x - \theta_n)\big),$$

$$\mathbf{F}' = \big(F(x), F(x - \theta_2'), \dots, F(x - \theta_n')\big).$$

Let $C^*(\mu_{\boldsymbol{\theta}})$ and $C^*(\mu_{\boldsymbol{\theta}'})$ correspond to the minimax risks when the true distribution vector is $\mathbf{F}$ or $\mathbf{F}'$, respectively. If the density functions of $\mathbf{F}$ and $\mathbf{F}'$ have the MLR property and if $\delta_i \leq \delta_i'$ for $i = 1, \dots, n-1$, then $C^*(\mu_{\boldsymbol{\theta}}) \leq C^*(\mu_{\boldsymbol{\theta}'})$.

PROOF. By induction it suffices to show that the statement holds for $\delta_I < \delta_I'$ for an arbitrary but fixed I and $\delta_i = \delta_i'$ for $i \neq I$. Let

$$\mathbf{A}_{\boldsymbol{\theta}}^* = \big((-\infty, d_1^*], (d_1^*, d_2^*], \dots, (d_{n-1}^*, \infty)\big)$$

denote an optimal solution when the true distribution vector is $\mathbf{F}$. Let $\Delta = \delta_I' - \delta_I$ and define a partition $\mathbf{A}' = (A_1', \dots, A_n')$ such that

$$A_i' = (d_{i-1}^*, d_i^*] \quad \text{for } i \leq I - 1,$$

$$A_I' = (d_{I-1}^*, d_I^* + \Delta],$$

$$A_i' = (d_{i-1}^* + \Delta, d_i^* + \Delta] \quad \text{for } i > I.$$

Then

$$\int_{A_I'} dF(x - \theta_I') \geq \int_{A_I^*} dF(x - \theta_I),$$

$$\int_{A_i'} dF(x - \theta_i') = \int_{A_i^*} dF(x - \theta_i) \quad \text{for } i \leq I - 1$$

and, by the translation invariance property,

$$\int_{A_i'} dF(x - \theta_i') = \int_{A_i^*} dF(x - \theta_i) \quad \text{for } i > I.$$

Since $\mathbf{A}'$ is not necessarily an optimal partition when the true distribution vector is $\mathbf{F}'$, the proof is complete. □

Combining Lemma 2.7 with Theorems 2.2 and 2.6, one immediately obtains the following theorems which apply to all location parameter families of distributions when the densities possess the MLR property.

THEOREM 2.2′. *Let F be a continuous distribution, $F_i(x) = F(x - \theta_i)$ for $\theta_1 < \theta_2 < \cdots < \theta_n$, and let $d = \min_{1 \leq i \leq n-1}(\theta_{i+1} - \theta_i)$. Let $q = 1 - \rho(F, d)$ where $\rho(F, d)$ denotes the tail-d concentration and let $C^*(\mu)$ be the minimax risk when the true distribution is given by $\mathbf{F} = (F(x - \theta_1), F(x - \theta_2), \dots, F(x - \theta_n))$. If $\{f(x - \theta)\colon \theta \in \Lambda\}$ has the MLR property, then the inequality in* (2.2) *holds. Moreover, this lower bound is best possible and is attained for all n, d and $q < 1$ when $\theta_{i+1} = \theta_i + d$, $i = 1, \dots, n - 1$.*

THEOREM 2.6′. *Assume that the conditions in Theorem 2.2′ are satisfied and that $F(x)$ is symmetric about b for some b and is unimodal. Let $\tau = (1 - \lambda)/(1 + \lambda)$ where $\lambda = \lambda(F, d)$ is the Lévy concentration for $F(x)$ and $d = \min_{1 \leq i \leq n-1}(\theta_{i+1} - \theta_i)$. Then the inequality in* (2.10) *holds. Moreover, the lower bound is best possible and is attained for all n, d and λ when $\theta_{i+1} = \theta_i + d$, $i = 1, \dots, n - 1$.*

3. Discontinuous distributions. Without continuity of F, the above minimax-risk inequalities may fail, but analogous inequalities may be derived as a function of the maximum discontinuity (atom size) of the distribution via the following generalizations of Lyapounov's convexity theorem and Proposition 1.1(ii).

PROPOSITION 3.1 [Elton and Hill (1987)]. *If all the atoms of each μ_i have mass less than or equal to α, then the Hausdorff distance from the range of μ to its convex hull is at most $\alpha n/2$.*

THEOREM 3.2. *If no F_i has discontinuity greater than α, then the Hausdorff distance from* $\mathrm{PR}(\mu)$ *to its convex hull is at most $\sqrt{2}\,\alpha n$.*

PROOF. Fix $(A_i)_1^n$ and $(B_i)_1^n$ in Π_n and $t \in (0,1)$. Define the $2n$-dimensional vector-valued Borel measure $\mathbf{m}$ by

$$\mathbf{m}(E) = (\mu_1(E \cap A_1), \dots, \mu_n(E \cap A_n), \mu_1(E \cap B_1), \dots, \mu_n(E \cap B_n)).$$

Since each μ_i has only atoms of mass less than or equal to α, so does $\mathbf{m}$. It then follows from Proposition 3.1 that there is an $E \in \mathbb{B}$ with $\|\mathbf{m}(E) - t\mathbf{m}(\mathbb{R})\| \le \alpha n$, that is,

$$\left[\sum_{i=1}^{n} (\mu_i(E \cap A_i) - t\mu_i(A_i))^2 + \sum_{i=1}^{n} (\mu_i(E \cap B_i) - t\mu_i(B_i))^2\right]^{1/2} \le \alpha n. \tag{3.1}$$

Since $|\mu_i(E \cap B_i) - t\mu_i(B_i)| = |\mu_i(B_i \setminus E) - (1-t)\mu_i(B_i)|$, it follows from (3.1) that

$$\begin{aligned} &\sum_{i=1}^{n} (\mu_i(A_i \cap E) - t\mu_i(A_i))^2 \\ &\qquad + \sum_{i=1}^{n} (\mu_i(B_i \setminus E) - (1-t)\mu_i(B_i))^2 \le \alpha^2 n^2. \end{aligned} \tag{3.2}$$

Letting $E_i = (A_i \cap E) \cup (B_i \setminus E)$ it follows from (3.2) that

$$\begin{aligned} &\sum_{i=1}^{n} (\mu_i(E_i) - t\mu_i(A_i) - (1-t)\mu_i(B_i))^2 \\ &\qquad = \|\mathbf{a} + \mathbf{c} - (\mathbf{b} + \mathbf{d})\|^2 \le 2(\|\mathbf{a} - \mathbf{b}\|^2 + \|\mathbf{c} - \mathbf{d}\|^2) \le 2\alpha^2 n^2, \end{aligned} \tag{3.3}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ are defined in the obvious manner. Taking square roots of both sides of (3.3) completes the proof, since $\mathbf{A}$, $\mathbf{B}$ and t were arbitrary. □

REMARK. The idea of stringing together vector measures is attributed by Dubins and Spanier (1961) to Blackwell.

Typical of an application of Theorem 3.2 to the classification problem is the following analog of Theorem 2.2.

THEOREM 3.3. *If $d > 0$, $F_i(x) = F_1(x - (i-1)d)$ for $i = 1, 2, \dots, n$ and if F_1 has maximum discontinuity α, then*

$$C^*(\mu) \ge \left(1 + \sum_{j=1}^{n-1} q^j\right)^{-1} - \sqrt{2}\,\alpha n,$$

where $q = 1 - \rho(F, d)$.

PROOF. Immediate from Theorems 2.2 and 3.2. □

Acknowledgments. The authors are grateful to John Elton for a suggestion concerning the proof of Theorem 3.2 and to the Associate Editor and the referee for a number of suggestions.

REFERENCES

DUBINS, L. and SPANIER, E. (1961). How to cut a cake fairly. *Amer. Math. Monthly* **68** 1–17.

DVORETZKY, A., WALD, A. and WOLFOWITZ, J. (1951). Relations among certain ranges of vector measures. *Pacific J. Math.* **1** 59–74.

ELTON, J. H. and HILL, T. P. (1987). A generalization of Lyapounov's convexity theorem to measures with atoms. *Proc. Amer. Math. Soc.* **99** 297–304.

FISHER, R. A. (1936). Uncertain inference. *Proc. Amer. Acad. Arts Sci.* **71** 245–257.

HENGARTNER, W. and THEODORESCU, R. (1973). *Concentration Functions.* Academic, New York.

KARLIN, S. and RUBIN, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272–299.

NEYMAN, J. (1946). Un théorème d'existence. *C. R. Acad. Sci. Paris* **222** 843–845.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332