PARTITIONING INEQUALITIES IN
PROBABILITY AND STATISTICS[1]

By THEODORE P. HILL

Georgia Institute of Technology

This article surveys fair–division or cake–cutting inequalities in probability and statistics, including bisection inequalities, basic fairness inequalities, convexity tools, superfairness inequalities, and partitioning inequalities in hypotheses testing and optimal stopping theory. The emphasis is measure theoretic, as opposed to game theoretic or economic, and a number of open problems in the area are mentioned.

1. Introduction

The main purpose of this article is to present a unified study of a class of partitioning inequalities in the theories of probability and statistics; it is not meant to be a complete review of the subject. The emphasis is measure theoretic with emphasis on both constructive (algorithmic) and non–constructive techniques, including generalizations of classical "cake–cutting" inequalities, the ham sandwich theorem, and classical statistical problems such as Fisher's Problem of the Nile, the problem of similiar regions, and the classification problem.

The overall framework is as follows. There are a finite number of (countably additive) probability measures $\mu = (\mu_1, \ldots, \mu_n)$ defined on the *same* measurable space $(\Omega, \mathcal{F})$, and a class Π of $\mathcal{F}$–measurable partitions of Ω is specified. (Recall that $(A_i)_1^k$ is an (ordered) $\mathcal{F}$–partition of Ω if $\cup_1^k A_i = \Omega, A_i \cap A_j = \emptyset$ for $i \neq j$, and $A_i \in \mathcal{F}$ for all i.) From this collection of partitions Π a single partition is sought which will satisfy some objective such as bisection or minimax–risk property. It may help the reader to keep in mind either a cake–cutting or a hypotheses–testing interpretation of this setting throughout the following sections.

In the cake–cutting interpretation, Ω is a cake which must be divided among n people having values $\mu_1, \ldots, \mu_n$ (that is, $\mu_j(A)$ is the relative value

---

of piece (measurable subset) A to person j), and Π describes the permissible divisions (e.g., into parallel slices, or convex connected pieces, or general Borel sets). The basic measure–theoretic assumptions of nonnegativity and (countable) additivity seem natural in this setting: the value of any piece is at least zero (otherwise the piece could be discarded); and the value of the union of several disjoint pieces is the sum of the values of the individual pieces.

In the hypotheses testing interpretation, a single observation is made of a random variable X taking values in Ω, and it must then be decided from which of n known distributions $\mu_1, \ldots, \mu_n$ the observation came; this is known as the *classification problem*. In this case, the decision rule "if $X \in A_i$, guess that the distribution of X is μ_i" corresponds to a partition of Ω, and the expected *risk* associated with this decision given that the true distribution is μ_i is given by $1 - \mu_i(A_i) = P(X \notin A_i \mid$ true distribution of X is $\mu_i)$.

The organization of this article is as follows: Section 2 addresses bisection results including ham sandwich theorems and medians; Section 3 the basic fairness inequalities; Section 4 the convexity tools, with special emphasis on Lyapounov's theorem and extensions; Section 5 various superfairness inequalities; Section 6 inequalities in statistical decision theory related to the classification problem; Section 7 partitioning inequalities in optimal-stopping theory; and Section 8 a list of open problems.

It should be emphasized that the results in this article are focused on probabilistic and statistical partitioning inequalities, and do not include discussion of related fair–division results in other areas such as combinatorics (cf. Alon and West (1986)), economics (cf. Kirman (1981), Samuelson (1980), Svensson (1983), Weller (1985), Young (1987)) or game theory (cf. Kuhn (1973), Legut (1990)).

2. Bisection

The Bisection Problem is the question of the existence of a single ($\mathcal{F}$–measurable) subset A of Ω which bisects Ω simultaneously with respect to each of the measures $\mu_1, \ldots, \mu_n$, that is

$$(1) \qquad\qquad \mu_i(A) = 1/2 \quad \text{for all } i = 1, \ldots, n.$$

In general such a set does not exist (e.g., if $\mu_1 = \mu_2$ is a Dirac measure assigning mass 1 to a single point ω in Ω), but the ham sandwich theorem of Steinhaus (cf. Stone and Tukey (1942)) guarantees that there is even a *half-space* simultaneous bisection in certain cases, namely

(2)   if $\mu_1, \ldots, \mu_n$ are uniformly distributed (probability measures) on bounded Borel subsets of $\Omega = \mathbb{R}^n$, there is halfspace H^+ satisfying $\mu_i(H^+) = 1/2$ for all $i = 1, \ldots, n$.

The classical proof of (2) uses the Borsuk–Ulam fixed point theorem, and relies heavily on the fact that the number of measures is no more than the dimension of the space. The hypotheses of (2) can be weakened (Stone and Tukey (1942)) to "$\mu_1, \ldots, \mu_n$ are probability measures on $\mathbb{R}^n$ satisfying $\mu_i(H) = 0$ for every hyperplane H and all i," but the conclusion may fail if the measures have atoms.

On the other hand, hyperplane bisection in a *median* sense is always possible for arbitrary (including atomic) probability measures in this setting. Say that a hyperplane $H = \sum_{i=1}^{n} a_i x_i - b$ in $\mathbb{R}^n$ is a *hyperplane median* for μ if $\mu(H^+) \geq 1/2$ and $\mu(H^-) \geq 1/2$, where $H^+ = \{\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n : \sum_{i=1}^{n} a_i r_i \geq b\}$ and $H^- = \{\mathbf{r} \in \mathbb{R}^n : \sum_{i=1}^{n} a_i r_i \leq b\}$. Using the Borsuk-Ulam theorem applied to a "midpoint-median" function, it was shown in Hill (1988a) that

(3)   every collection $\mu_1, \ldots, \mu_n$ of arbitrary Borel probability measures on $\Omega = \mathbb{R}^n$ always has a common hyperplane median.

Using countable additivity, it can be seen that if μ–bisection of every measurable set is possible, then the range of μ is convex (and conversely); cf. Dubins and Spanier (1961) and Section 4 below. Stone and Tukey (1942) have shown that for any two nonatomic Borel measures on the closed unit circle S^1 there is always an *interval* which bisects each of the measures simultaneously. For an interesting example of combinatorial bisection also based on the Borsuk-Ulam theorem, the reader is referred to Alon and West (1986).

For general $\alpha \in [0, 1]$ the question of the existence of a set A satisfying $\mu_i(A) = \alpha$ for all i (as opposed to the exact bisection in (1)), is called the Problem of Similar Regions (Feller (1938)), as such a set A is in some gross sense a smaller copy of Ω itself, and this question has been related to the efficiency of tests of statistical hypotheses by Neyman and Pearson (1933). In contrast to (2), it is not always possible to find a hyperplane H satisfying $\mu_i(H^+) = \alpha$ for $\alpha \neq 1/2$, even if the measures are continuous (Hill (1988a)). On the other hand, it follows from convexity (Section 4 below) that if the measures are all nonatomic, then for each $\alpha \in [0, 1]$ there is a measurable set A_α satisfying $\mu_i(A_\alpha) = \alpha$ for all $i \leq n$.

3. Basic Fairness Inequalities

The basic question of the existence of a *fair division* is that of the existence of an ($\mathcal{F}$–measurable) n–partition $(A_i)_{i=1}^n$ of Ω satisfying

$$(4) \qquad \mu_i(A_i) \geq 1/n \quad \text{for all } i = 1, \ldots, n.$$

The cake–cutting interpretation of (4) says that cutting the cake into pieces $A_1, \ldots, A_n$ and distributing it so that the ith piece is given to person i guarantees that each person receives a portion which he considers, by his own measure, to be at least one n^{th} of the total. The hypotheses–testing interpretation of (4) is that the decision rule corresponding to the partition $(A_i)_1^n$ has expected risk at most $(n-1)/n$. Just as bisections (1) do not exist in general, neither do fair divisions in the sense of (4). They do, however, if all the measures are nonatomic (Steinhaus (1949)), and more generally if at most one of the measures has atoms (Hill and Kennedy (1990)):

(5) if $\mu_1, \ldots, \mu_{n-1}$ are nonatomic, then there is a measurable partition $(A_i)_1^n$ with $\mu_i(A_i) \geq 1/n$ for all $i \leq n$.

For $n = 2$, the demonstration of (5) is the classical "cut–and–choose" algorithm: the first person (the person with the nonatomic measure) identifies a subset A which bisects his own measure of Ω (i.e., $\mu_1(A) = 1/2$), and the second person chooses between A and the complement of A. Kirby (1988) has recently applied this idea to obtain an algorithm for nuclear arms reduction.

For $n > 2$, there are several well known algorithms to demonstrate (5). One algorithm, called a "sliding knife" solution, is a modification of an algorithm of Knaster and Steinhaus (1946, 1953) by Dubins and Spanier (1961). Although stated under the hypothesis that *all* n measures are continuous (that is, absolutely continuous with respect to Lebesgue measure on $\Omega \subset \mathbb{R}^n$), the procedure also works if at most one of the measures has atoms. In this algorithm, a long knife is passed slowly parallel to itself over the cake Ω until one of the participants feels that the increasing portion under the knife is exactly one n^{th} the total value, at which point he says "stop," and the cake is cut at that point and that slice is given to the person who said stop (ties are broken in any manner), and the remaining $n - 1$ participants continue the process. For continuous measures, any starting orientation of the knife will suffice, and for the more general nonatomic case, it follows from Jones (1989) that almost all starting angles will suffice (a "sufficient" starting angle is one in which at most one measure of the corresponding increasing slice at that angle is discontinuous).

Fink (1964) has given an algorithm demonstrating (5) which generalizes the cut–and–choose algorithm and which has two advantages over the

sliding–knife solution: first, implementation of the algorithm does not require *a priori* knowledge of the number of participants; and second, the algorithm is essentially finite, as opposed to the continuous–evaluation method of the sliding–knife solution. In Fink's algorithm the first person bisects the cake Ω according to his own measure. The second person arriving chooses between the two pieces cut by the first player, and if a third person arrives then each of these first two players *trisects* his own portion, and the third person selects one portion from each. The algorithm continues in this manner (e.g., quadrisection at the next stage), until no new arrivals appear and the algorithm termminates. (Note that the single measure with atoms must "arrive" last in order to guarantee the solution in (5).) A variation of this algorithm requiring at most $\mathcal{O}(n \log n)$ cuts for parallel slices in $\mathbb{R}^n$ is described in Even and Paz (1984).

The bound $1/n$ in (5) is easily seen to be best possible (taking $\mu_1 = \mu_2 = \ldots = \mu_n$), and the corresponding best possible bound for nonatomic finite (e.g., non–probability) measures is one–n^{th} the harmonic mean of the total masses (Hill (1985)), a probabilistic analog of which is

(6)   if $X_1, X_2, \ldots, X_n$ are nonnegative continuous random variables on $(\Omega, \mathcal{F}, P)$ with finite means, then there is a measurable partition $(A_i)_1^n$ of Ω satisfying $\int_{A_i} X_i \geq ((EX_1)^{-1} + \ldots + (EX_n)^{-1})^{-1}$ for all i, and this bound is best possible.

Although an algorithmic proof of (6) is possible in some special cases, such as when the $\{EX_i\}$ are all rational numbers, proofs of the general case seem to rely on non–constructive results such as the convexity conclusions in Section 4 below.

If more than one of the measures $\mu_1, \ldots, \mu_n$ has atoms, then the conclusion of (5) may fail, but if an upper bound is known for the maximum atom size of the measures, the following best possible fairness bound is known (Hill (1987a)); taking limits as $\alpha \to 0$ yields (5) as a corollary.

(7)   If $\mu_1, \ldots, \mu_n$ each have atoms at most $\alpha > 0$, then there exists a measurable partition $(A_i)_1^n$ of Ω satisfying $\mu_i(A_i) \geq V_n(\alpha)$ for all $i = 1, \ldots, n$, where $V_n : [0,1] \to [0, n^{-1}]$ is the unique nonincreasing function satisfying $V_n(\alpha) = 1 - k(n-1)\alpha$ for $\alpha \in [(k+1)k^{-1}((k+1)n-1)^{-1}, (kn-1)^{-1}]$. Moreover, this bound is attained for all n and α.

The function V_n is piecewise linear and satisfies $V_n(\alpha) \nearrow n^{-1}$ as $\alpha \searrow 0$; see Hill (1987a) for the graphs of V_2 and V_3. The proof of (7) is largely combinatorial and nonconstructive in nature.

Analogs of the fairness inequalities in this section for game theory and economics can be found in Crawford (1977), Crawford and Heller (1979), Demko and Hill (1988), Kuhn (1973), and Legut (1985).

4. Convexity Tools

For nonatomic measures, the basic partitioning–inequality tool is the celebrated convexity theorem of Lyapounov (1940) which states that the range of every countably–additive, finite–dimensional, vector–valued measure is closed and convex. Many proofs of this theorem and various generalizations have appeared (e.g. Armstrong and Prikry (1981), Artstein (1990), Blackwell (1951a), Dvoretzky, Wald and Wolfowitz (1951), Elton and Hill (1987), Gouweleeuw (1991), Halmos (1948), Karlin (1953), Lindenstrauss (1966), Margolies (1978)); that of Lindenstrauss (1966) based on the Krein–Milman Theorem being perhaps the shortest, and that of Artstein (1990) perhaps the most elementary. Since $0 = \mu(\emptyset)$ and $1 = (1, 1, \ldots, 1) = \mu(\Omega)$ are in the range of μ, Lyapounov's theorem immediately guarantees that if the measures are all nonatomic, measurable bisecting sets (1) and fair divisions ((4) with *equality*) always exist; a little extra effort yields (6) (also with equality).

In non–probabilistic applications, the convexity theorem has had widespread application (see Akemann and Anderson (1990)) in combinatorics, control theory (to prove the basic bang–bang principle; LaSalle (1960)) differential equations, economics, functional analysis, graph theory, and logic. Another curious probabilistic implication of the convexity theorem apparently first obtained by Blackwell (1951b) is: given any finite collection of continuous (Borel) probability distributions on the real line, there is a sub–σ–algebra $\mathcal{G}$ of the Borels with the property that, restricted to $\mathcal{G}$, those measures are *identical* non–atomic probability measures.

A generalization of Lyapounov's convexity theorem due to Dvoretzky, Wald, and Wolfowitz (1951) (see also Dubins and Spanier (1961)) which is particularly useful in the present setting is the following:

(8)   if $\mu_1, \ldots, \mu_n$ are nonatomic, then for each k,

$$\{(\mu_i(A_j))_{i=1}^n {}_{j=1}^k : (A_j)_1^k \text{ is a measurable } k\text{–partition of } \Omega\}$$

is a compact and convex subset of $n \times k$ real matrices.

The proof of (8) is based upon Lyapounov's theorem and an idea of stringing together measures attributed to Blackwell (1951a). Application of (8) yields an affirmative solution to R. A. Fisher's "Problem of the Nile" (1936) for nonatomic measures $\mu_1, \ldots, \mu_n$; namely, the existence for each natural number k of a measurable k–partition $(A_j)_1^k$ satisfying

$$\mu_i(A_j) = 1/k \text{ for all } i \leq n \text{ and all } j \leq k,$$

and more generally, for each k and each set of positive numbers $\alpha_1, \ldots, \alpha_k$ with $\sum \alpha_i = 1$, a k–partition satisfying $\mu_i(A_j) = \alpha_j$ for all $i \leq n$ and all

$j \leq k$. Some information concerning the minimal number of "cuts" required to obtain such a partition (in the case $k = n$; $\alpha_j = 1/k$) is contained in Legut (1991) and in Stromquist and Woodall (1985); Legut (1991) gives a qualitative characterization of the partitions for continuous measures.

The "closed" conclusion of Lyapounov's theorem holds even if the measures have atoms, but the "convexity" conclusion fails in general. On the other hand, if a bound on the maximum atom size is known, then the following generalization by Elton and Hill (1987) of the convexity theorem gives a bound on how non–convex the range may be; intuitively, if the atoms are all very small, the range of $\boldsymbol{\mu}$ will be close to convex:

(9)   if $\mu_1, \ldots, \mu_n$ each have atoms at most $\alpha > 0$, then the Hausdorff distance from the range $\boldsymbol{\mu}$ to its convex hull is at most $\alpha n/2$.

(Recall that the range of $\boldsymbol{\mu}$ is the subset of $\mathbb{R}^n$ given by $\{\boldsymbol{\mu}(A) : A \in \mathcal{F}\}$, and the Hausdorff distance between $S_1 \subset S_2$ is $d(S_1, S_2) = \sup_{x \in S_2} \inf_{y \in S_1} |x - y|$.)

Thus (9) affords approximate solutions to the bisection and fair–division problems (as well as the Problem of the Nile), in the case of measures with atoms no bigger than α. For example, it implies that if μ_1, μ_2, μ_3 each have atoms no larger than $1/100$, then there is an "almost bisecting" set A satisfying $97/200 \leq \mu_i(A) \leq 103/200$ for all $i \leq 3$.

The convexity conclusion of Lyapounov's theorem may fail if the number of measures is infinite (Feller (1938)). If the hypothesis of countable additivity is weakened to finite additivity, the convexity conclusion still holds (Margolies (1978), Armstrong and Prikry (1981)).

5. Superfairness Inequalities

If $\mu_1, \ldots, \mu_n$ are *identical* measures, then any partition $(A_i)_1^n$ satisfying the fairness inequality (4) necessarily holds with equality for all i. On the other hand,

(10)   if $\mu_1, \ldots, \mu_n$ are nonatomic and $\mu_i \neq \mu_j$ for some $i \neq j$, then there is a measurable partition $(A_i)_1^n$ of Ω satisfying $\mu_i(A_i) > 1/n$ for all $i \leq n$.

The result (10) was apparently first stated by Knaster and Steinhaus (1953), proved by Urbanik (1955) for the case the measures all have the same null sets, and proved independently by Dubins and Spanier (1961) for the general nonatomic case. Although both these proofs used Lyapounov's Convexity Theorem (note that the strict inequality conclusion $> 1/n$ requires a much more subtle argument than the weak inequality $\geq 1/n$), Woodall

(1980) has modified Fink's (1964) fair–division algorithm to yield an algorithm to generate the superfair partition appearing in (10). Woodall's algorithm, however, requires more information than just $\mu_i \neq \mu_j$ for some $i \neq j$; it requires knowledge of a set A, real numbers $\alpha \neq \beta$, and exact indices i and j satisfying $\mu_i(A) = \alpha \neq \beta = \mu_j(A)$.

Woodall's algorithm proving (10) does not give any bound strictly greater than $1/n$, but such bounds are possible if the total masses of the supremum or infimum of the μ_i's are known. (Here $\bigvee_{i=1}^{n} \mu_i$ is the smallest measure dominating each μ_i, and $\bigwedge_{i=1}^{n} \mu_i$ is the largest measure dominated by each μ_i; it is an easy exercise to show that such measures always exist.) The superfair inequality of Elton, Hill and Kertz (1986)

(11)   if $\mu_1, \ldots, \mu_n$ are nonatomic, then there is a measurable partition $(A_i)_1^n$ of Ω satisfying $\mu_i(A_i) \geq (n - M + 1)^{-1}$, where M is the total mass of $\bigvee_{i=1}^{n} \mu_i$,

is best possible, and improves (10) since $M \geq 1$, with equality if and only if $\mu_1 = \ldots = \mu_n$. The proof in Elton, Hill and Kertz (1986) is partly constructive, and Legut (1988) contains an easy nonconstructive proof based on Lyapounov's convexity theorem. Using the convexity theorem and an "inversion principle," an analog of (11) for the infimum was obtained by Hill (1987b):

(12)   if $\mu_1, \ldots, \mu_n$ are nonatomic, then there exists a measurable partition $(A_i)_1^n$ of Ω satisfying $\mu_i(A_i) \geq (n + m - 1)^{-1}$, where m is the total mass of $\bigwedge_{i=1}^{n} \mu_i$.

The bound in (12) is also best possible, and improves (10) since $m \leq 1$, with equality if and only if $\mu_1 = \ldots = \mu_n$. If $n = 2$, (11) and (12) are equivalent since $m + M = 2$, but for $n > 2$ neither (11) nor (12) implies the other.

The superfairness inequality (10) can also be generalized in another direction. If $\{\alpha_i\}_1^n$ are nonnegative numbers with $\alpha_1 + \ldots + \alpha_n = 1$, then the same hypothesis as in (10) guarantees the existence of measurable partition $(A_i)_1^n$ of Ω satisfying $\mu_i(A_i) > \alpha_i$ for all $i \leq n$ (cf. Dubins and Spanier (1961)). The weights $\{\alpha_i\}$ can be viewed in the cake–cutting framework as representing non–uniform shares to which each participant is entitled and in the hypotheses–testing framework as non–uniform loss functions.

6. Inequalities in Hypotheses Testing

In the classification problem setting, using decision rule (partition) $(A_i)_1^n$ against $\mu_1, \ldots, \mu_n$ results in a maximum expected risk of $\max_{i \leq n} P(X \notin A_i \mid \mathrm{dist}(X) = \mu_i)$, and the objective is to minimize this risk. That is, a partition $(A_i)_1^n$ of Ω is sought which will attain the *minimax risk*

$$R(\boldsymbol{\mu}) = \inf\{\max_{i \leq n} P(X \notin A_i \mid \mathrm{dist}(X) = \mu_i) : (A_i)_1^n \text{ is an } \mathcal{F}\text{-partition of } \Omega\}.$$

Since $\inf\max P(X \notin A_i \mid \mathrm{dist}X = \mu_i) = \inf\max(1 - \mu_i(A_i)) = 1 - \sup\min \mu_i(A_i)$, it follows that $R(\boldsymbol{\mu}) = 1 - \max\{\min_{i \leq n} \mu_i(A_i) : (A_i)_1^n \text{ is an } \mathcal{F}\text{-partition of } \Omega\}$, so the fairness and superfairness inequalities above can all be translated immediately into minimax risk inequalities. For example, the analog of (11) (cf. Elton, Hill and Kertz (1986)) is

(13)   if $\mu_1, \ldots, \mu_n$ are nonatomic, then the minimax risk R satisfies $n^{-1}(n - M) \leq R(\boldsymbol{\mu}) \leq (n - M + 1)^{-1}(n - M)$, where M is the total mass of $\bigvee_{i=1}^{n} \mu_i$.

Both bounds in (13) are sharp (the lower bound is easy, and the upper bound follows from (11)), and are attained.

In a similar application of the convexity theorem to the classification problem, Dvoretzky, Wald, and Wolfowitz (1951) showed that if $\mu_1, \ldots, \mu_n$ are nonatomic, then given any *randomized* decision function and loss functions $L_{i,j}$ there exists a non–randomized decision function (i.e., partition) with exactly the same expected risks for each i.

In a special case of the classification problem called the *location–parameter* problem, $\Omega = \mathbb{R}^1$ and the measures μ_i are translates of one another (for example μ_i is $N(a_i, 1)$ for each i), and the classification problem is now that of guessing which parameter was underlying the observation X. Based on concentration parameters of the (location–parameter) distributions, sharp bounds for the minimax risk were obtained by Hill and Tong (1989) using the convexity theorem. For example, letting $\rho(\mu_1, d)$ denote the *tail–d concentration* of μ_1 (see Hill and Tong (1989)),

(14)   if μ_1 is continuous and $\mu_i(A) = \mu_i(A - (i - 1)d)$ for all A and $i = 1, \ldots, n$, then $R(\boldsymbol{\mu}) \leq \left(\sum_{j=1}^{n-1} q^j\right) / \left(\sum_{j=0}^{n-1} q^j\right)$, where $q = 1 - \rho(\mu_1, d)$. Moreover this bound is best possible and is attained for all n, all d and all $q < 1$.

And, letting $\lambda(\mu, d)$ denote the *Levy d–concentration* $\sup_x \mu([x, x + d])$ of μ,

(15) if μ_1 is continuous and $\mu_2(A) = \mu_1(A - d)$ for all A, then there is a test for testing $H_1 : \text{dist}(X) = \mu_1$ versus $H_2 : \text{dist}(X) = \mu_2$ which satifies

$$\max\{\alpha, \beta\} \leq (1 - \lambda)/(2 - \lambda)$$

(where α and β are the type I and type II errors, respectively, and $\lambda = \lambda(\mu_1, d)$.) Moreover, this bound is attained for all d and all λ.

Again, a key element in the proof of (15) is the convexity theorem; Legut and Wilczynski (1991) have found improvements of (14) using a similar argument.

The convexity theorem was also used by Hill (1987a) to establish a proportionality principle for partitioning problems, which essentially says that in a general class of partitioning problems, the worst case is when the measures are proportional. One corollary of that principle related to the classification problem is

(16) if $X_1, X_2, \ldots, X_n$ are independent continuous random variables on $(\Omega, \mathcal{F}, P)$, then for each positive integer $k \leq n$ and each set of k distinct integers $K = \{1 \leq i_1 < \ldots < i_k \leq n\}$, there is a real Borel set B satisfying

$$P(X_i \in B \text{ if and only if } i \in K) \geq \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k},$$

and this bound is best possible.

7. Partitioning Inequalities in Optimal–Stopping Theory

The classical problem in optimal–stopping theory is: given a sequence of integrable random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ on $(\Omega, \mathcal{F}, P)$, find a stopping time t which maximizes EX_t. Here the stopping times are required to be adapted to an increasing filtration of σ–algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$, where typically $\mathcal{F}_j$ is the σ–algebra $\sigma(X_1, \ldots, X_j)$ generated by $X_1, \ldots, X_j$. In other words, a stopping time t corresponds to an n–partition $(A_i)_1^n$ of Ω satisfying $A_j \in \mathcal{F}_j$ for each $j \leq 1$; the correspondence is simply $\{t = j\} = A_j$. Thus the set Π of allowable partitions of Ω for optimal stopping is more restricted, by these σ–algebra constraints, than for the above fair–division problems. Still, many of the convexity tools apply in optimal stopping, where for example it was shown by Hill and Pestien (1983) using Lyapounov's theorem and an idea of Blackwell (1951a) that even in a finitely additive setting, the function $t \mapsto EX_t$ has convex range on the nonatomic components of the distributions.

A useful analog to the Lyapounov convexity result for optimal stopping is (Hill and Kennedy (1990))

(17)   if $X_1, \ldots, X_n$ are integrable random variables on $(\Omega, \mathcal{F}, P)$, and X_1 is continuous, then the stopping time range of $\mathbf{X}$

$$\left\{ \left(\int_{t=1} X_1, \ldots, \int_{t=n} X_n \right) : t \text{ is a stopping time for } \mathbf{X} \right\}$$

is a closed convex subset of $\mathbb{R}^n$.

With *no* restrictions on X_1, the stopping time range may fail to be convex, although the *randomized*–stopping rule range is always convex. Note the contrast between (17), where the nonatomic assumption is only required for the first distribution, and (5), where nonatomicity is required for all but one of the distributions.

Using (17), the separating hyperplane theorem and the classical prophet inequality of Krengel and Sucheston (1978), the following sharp minimax partitioning inequality (18) in a prophet problem (optimal stopping) setting was proved in Hill and Kennedy (1990). Here $\mathcal{T}$ is the set of stopping times for $X_1, \ldots, X_n$, and $\mathcal{S}$ is the set of stopping functions which are $\mathcal{F}_n$ measurable, that is, $s \in \mathcal{S}$ has the property that the decision to stop at time j may depend on the *whole* sequence $X_1, \ldots, X_n$ as opposed to just the first j variables. In this sense, a player allowed to use stopping times from $\mathcal{S}$ is like a prophet, in that he can use information about future variables to decide when to stop (cf. Hill and Kennedy (1990) for the formal definition of $\mathcal{S}$).

(18)   If $X_1, X_2, \ldots, X_n$ are integrable nonnegative random variables on $(\Omega, \mathcal{F}, P)$ and X_1 is continuous, then

$$\sup_{s \in \mathcal{S}} \min_i \int_{s=i} X_i \leq 2 \sup_{t \in \mathcal{T}} \min_i \int_{t=i} X_i,$$

and the bound 2 is best possible. If X_1 is not continuous, the inequality may fail, but does hold (and is best possible) if $\mathcal{T}$ is replaced with the collection of *randomized* stopping times.

A probabilistic interpretation of (18) is that if the objective is to maximize the minimum expected reward of stopping at $i = 1, \ldots, n$ then a prophet (or player with complete foresight) may never do better than twice that of an ordinary player restricted to using non–anticipatory stopping times. The convexity result (17) is also used in Hill and Kennedy (1990) to prove a stopping time analog of the partitioning principle in Hill (1988b); typical corollaries of which are (6) and

(19)   if $X_1, \ldots, X_n$ are integrable nonnegative random variables on $(\Omega, \mathcal{F}, P)$, then there is a randomized stopping time t satisfying

$$\prod_{i=1}^{n} \int_{t=i} X_i \geq n^{-n} \prod_{i=1}^{n} EX_i,$$

and this bound is best possible.

8. Open Problems

The main purpose of this section is to record a number of open problems related to the partitioning inequalities mentioned above.

PROBLEM 1. Find a finite algorithm for generating a hyperplane median guaranteed by (3) for general distributions.

PROBLEM 2. Find an *efficient* algorithm for generating a hyperplane median based on a finite set of data points. (If there are k n–dimensional data points in $\mathbb{R}^n$, then it is easy to see that at least one of the $\binom{k}{n}$ hyperplanes will be a median, but checking all possible hyperplanes is certainly not optimal.)

The algorithms of Steinhaus, Banach and Knaster and of Fink guarantee a fair solution (i.e., $\mu_i(A_i) \geq 1/n$ for all i), but do not guarantee a *first–choice* solution in which each participant gets the piece he values most highly, i.e., satisfying

(20) $\mu_i(A_i) \geq \mu_i(A_j)$ for all $i \leq n$ and $j \leq n$.

Of course convexity (e.g., (8)) guarantees the existence of a partition satisfying $\mu_i(A_j) = 1/n$ for all $i \leq n$ and all $j \leq n$ (and hence satisfying (20)) but Gamow and Stern (1958) raised the question of finding an *algorithm* generating a first–choice solution. Stromquist (1980) and Woodall (1980) independently found an algorithm for $n = 3$.

PROBLEM 3. Find (or demonstrate non–existence of) a finite algorithm yielding a first–choice partition (20) for $n \geq 4$.

PROBLEM 4. Find the best possible inequality generalizing (6) and (7), that is, find the largest constant $k = k(n, \alpha, \|\mu_1\|, \ldots, \|\mu_n\|)$ so that if $\mu_1, \ldots, \mu_n$ have atoms at most α, then there is a partition $(A_i)_1^n$ of Ω satisfying $\min_i \mu_i(A_i) \geq k$.

PROBLEM 5. Find necessary and sufficient geometric conditions on a set $S \subset \mathbb{R}^n$ so that S is the range of a nonatomic vector (probability) measure. (By

Lyapounov's theorem, S must be convex and compact; by general principles, S must be centrally symmetric, contain the origin, and lie in the positive orthant. For $n = 2$, these five conditions are also sufficient, but they are not sufficient for $n > 2$. Bolker (1969, 1971) attributes this question to Blaschke, and proves non–geometric characterizations.)

PROBLEM 6. Find necessary and sufficient conditions on $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ so that the range of $\boldsymbol{\mu}$ is convex. (By Lyapounov's theorem, nonatomicity of the $\{\mu_i\}$ suffices, but even for $n = 1$ it is not necessary, as can be seen by looking at a purely atomic measure with atoms of size $\frac{1}{2^k}$ for all $k \geq 1$. Gouweleeuw (1991) has partial results in this direction.)

PROBLEM 7. Find natural topological and geometric generalizations of the convexity theorem; for example, if Ω is a polyhedron in $\mathbb{R}^n$, is the range of nonatomic measures over subpolyhedra convex? What about open simply connected sets? (The problem seems to be that taking limits is not always possible here; the limit of polygons need not be a polygon, nor must the limit of connected sets be connected. Approximate convexity is known in some cases, such as the fair–border results in Hill (1983) and Beck (1987), and Samuelson (1980) has a geometric fair–division scheme for coastal water-ways. Note that if Ω is a convex subset of $\mathbb{R}^n$, the sliding–knife fair–division algorithm, as well as algorithms of Stromquist (1980) and Woodall (1980) generate partitions consisting of convex pieces. Gardner's problem (1978) is purely geometric.)

PROBLEM 8. Find the best possible constant in the generalization of Lya-pounov's theorem to measures with atoms (9). (The bound $\alpha n/2$ is not sharp for small n, but is known to be of the correct order in n; the best possible bound is at least $n/8$ for general n and at least $n/4$ if n is a power of 2 (Elton and Hill (1987).)

PROBLEM 9. Find the best possible bound generalizing both superfairness inequalities (11) and (12); that is, find the largest $k = k(n, m, M)$ depending on the number of measures n, and the masses m and M of the infimum and supremum, respectively, so that if $\mu_1, \ldots, \mu_n$ are nonatomic with $\| \vee \mu_i \| = M$ and $\| \wedge \mu_i \| = m$, then there is a partition $(A_i)_1^n$ satisfying $\mu_i(A_i) \geq k(n, m, M)$.

PROBLEM 10. Find finite algorithms for generating the superfair partitions guaranteed by (11) and (12).

PROBLEM 11. Prove (or give a counterexample) that the minimax risk result (14) holds if ρ is replaced by the Levy concentration λ.

PROBLEM 12. For a general class of problems, find a proportionality prin-ciple for general measures (the proportionality principle in Hill (1988b) was

for nonatomic measures, but in many partitioning problems for general measures (e.g., Hill (1987a)), the worst case is also known to be when all the measures are proportional, i.e., *equal*, in the case of probability measures. A basic superfairness property generalizing (10) should hold.)

Acknowledgements

REFERENCES

AKEMANN, C. A. AND ANDERSON, J. (1990). Lyapounov theorems for operator algebras, preprint. University of California at Santa Barbara.

ALON, N. AND WEST, D. B. (1986). The Borsuk–Ulam theorem and bisection of necklaces. *Proc. Amer. Math. Soc.* **98** 623–628.

ARMSTRONG, T. AND PRIKRY, K. (1981). Liapounoff's theorem for nonatomic finitely–additive, bounded, finite–dimensional, vector–valued measures. *Trans. Amer. Math. Soc.* **266** 499–514.

ARTSTEIN, Z. (1990). Yet another proof of the Lyapounov convexity theorem. *Proc. Amer. Math. Soc.* **108** 89–91.

BECK, A. (1987). Constructing a fair border. *Amer. Math. Monthly* **94** 157–162.

BLACKWELL, D. (1951a). On a theorem of Lyapounov. *Ann. Math. Statist.* **22** 112–114.

BLACKWELL, D. (1951b). The range of certain vector integrals. *Proc. Amer. Math. Soc.* **2** 390–395.

BOLKER, E. D. (1969). A class of convex bodies. *Trans. Amer. Math. Soc.* **145** 323–345.

BOLKER, E. D. (1971). The zonoid problem. *Amer. Math. Monthly* **78** 529–531.

CRAWFORD, V. P. (1977). A game of fair division. *Rev. Econ. Studies* **44** 235–247.

CRAWFORD, V. P. AND HELLER, W. P. (1979). Fair division with indivisible commodities. *J. Econ. Th.* **21** 10–27.

DEMKO, S. AND HILL, T. P. (1988). Equitable distribution of indivisible objects. *Math. Soc. Sci.* **16** 1–14.

DUBINS, L. E. AND SPANIER, E. H. (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. AND HILL, T. P. (1987). A generalization of Lyapounov's convexity theorem to measures with atoms. *Proc. Amer. Math. Soc.* **99** 297-304.

ELTON, J., HILL, T. P. AND KERTZ, R. P. (1986). Optimal–partitioning inequalities for nonatomic probability measures. *Trans. Amer. Math. Soc.* **296** 703–725.

EVEN, S. AND PAZ, A. (1984). A note on cake–cutting. *Discrete Appl. Math.* **7** 285–296.

FELLER, W. (1938). Note on regions similar to the sample space. *Statistical Research Memoirs.*

FINK, A. M. (1964). A note on the fair division problem. *Math. Magazine* **37** 341–342.

FISHER, R. A. (1936). Uncertain inference. *Proc. Amer. Acad. Arts Sci.* **71** 245–258.

GAMOW, G. AND STERN, M. (1958). *Puzzle Math.* Viking, New York.

GARDNER, M. (1978). The curious cake cut. *Aha! Insight!* 56–57, W. H. Freeman, New York.

GOUWELEEUW, J. (1991). A generalization of Lyapounov's theorem. Tech. Report No. WS–381, Free University, Amsterdam.

HALMOS, P. R. (1948). The range of a vector measure. *Bull. Amer. Math. Soc.* **54** 416–421.

HILL, T. P. (1983). Determining a fair border. *Amer. Math. Mo.* **90** 438–442.

HILL, T. P. (1985). Equipartitioning common domains of non–atomic measures. *Math. Z.* **189** 415–419.

HILL, T. P. (1987a). Partitioning general probability measures. *Ann. Probab.* **15** 804–813.

HILL, T. P. (1987b). A sharp partitioning–inequality for non–atomic probability measures based on the mass of the infimum of the measures. *Prob. Th. and Rel. Fields* **75** 143–147.

HILL, T. P. (1988a). Common hyperplane medians for random vectors. *Amer. Math. Mo.* **95** 437–441.

HILL, T. P. (1988b). A proportionality principle for partitioning problems. *Proc. Amer. Math. Soc.* **103** 288–293.

HILL, T. P. AND KENNEDY, D. P. (1990). Optimal stopping problems with generalized objective functions. *J. Appl. Prob.* **28** 828–838.

HILL, T. P. AND PESTIEN, V. (1983). The advantage of using non–measurable stop rules. *Ann. Probab.* **11** 442–450.

HILL, T. P. AND TONG, Y. L. (1989). Optimal–partitioning inequalities in classification and multi–hypotheses testing. *Ann. Statist.* **17** 1325–1334.

JONES, M. (1989). *Universal constants in optimal stopping theory.* Ph.D. Dissertation, Georgia Institute of Technology.

KARLIN, S. (1953). Extreme points of vector functions. *Proc. Amer. Math. Soc.* **4** 603–610.

KARLIN, S. AND STUDDEN, W. (1966). *Tchebycheff Systems with Applications in Analysis and Statistics.* Wiley, New York.

KIRBY, R. (1988). A new proposal for nuclear arms negotiation, preprint. Department of Mathematics, University of California at Berkeley.

KIRMAN, A. (1981). Measure theory with applications to economics. In *Handbook of Mathematical Economics.* Vol. I, Ch. 5, K. Arrow and M. Intrilgator, eds., North–Holland, Amsterdam.

KNASTER, B. (1946). Sur le probleme du partage pragmatique de H. Steinhaus. *Comptes Rendus de la Societé Polonaise de Mathématique, Annales de la Societé Polonaise de Mathematique* **19** 228–230.

KNASTER, B. AND STEINHAUS, H. (1953). Sur le partage pragmatique. *Comptes Rendus de la Societé des Sciences et des Lettres de Wroclaw.* Communication No. 1, 2.

KRENGEL, U. AND SUCHESTON, L. (1978). On semiamarts, amarts, and processes with finite value. *Probability on Banach Spaces.* J. Kuelbs, ed., Marcel Dekker, New York.

KUHN, H. W. (1973). On games of fair division. In *Essays in Modern Math. Econ.* M. Shubik, ed., Princeton University Press, Princeton, NJ. 29–37.

LASALLE, J. P. (1960). The time optimal control problem. In *Contributions to the Theory of Nonlinear Oscillators V.* Princeton University Press, Princeton, NJ.

LEGUT, J. (1985). The problem of fair division for countably many participants. *J. Math. Anal. Appl.* 109 83–89.

LEGUT, J. (1987). A game of fair division with continuum of players. *Colloq. Math.* 53 323-331.

LEGUT, J. (1988). Inequalities for α–optimal partitioning of a measurable space. *Proc. Amer. Math. Soc.* 104 1249–1251.

LEGUT, J. (1990). On totally balanced games arising from cooperation in fair division. *Games and Econ. Behavior* 2 47–60.

LEGUT, J. (1991). On dividing the unit interval. Preprint, Technical University of Wroclaw.

LEGUT, J. AND WILCZYNSKI, M. (1988). Optimal partitioning of a measurable space. *Proc. Amer. Math. Soc.* 104 262–264.

LEGUT, J. AND WILCZYNSKI, M. (1991). Inequalities for the minimax risk in multihypotheses testing. Preprint, Technical University of Wroclaw.

LINDENSTRAUSS, J. (1966). A short proof of Liapounoff's convexity theorem. *J. Math. Mech.* 15 971–972.

LYAPOUNOV, A. (1940). Sur les fonctions–vecteurs complétement additives. *Bull. Acad. Sci. URSS* 4 465–478.

MARGOLIES, D. (1978). *A study of finitely additive measures as regards amenable groups, Liapounov's theorem, and the elimination of infinite integrals via nonstandard real numbers.* Ph.D. Dissertation, University of California at Berkeley.

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

NEYMAN, J. AND PEARSON, E. (1933). On the problem of most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. London Ser. A* 231 289–337.

REBMAN, K. (1979). How to get (at least) a fair share of the cake. *Mathematical Plums, Dolciani Math. Expositions* #4. R. Honsberger, ed., 22–37.

SAMUELSON, W. (1980). The object distribution problem revisited. *Quart. J. Econ.* 85–90.

SAMUELSON, W. (1985). Dividing coastal waterways. *J. Conflict Resolution* 29 83–111.

STEINHAUS, H. (1948). The problem of fair division. *Econometrika* 16 101–104.

STEINHAUS, H. (1949). Sur la division pragmatique. *Econometrika (Supplement)* 17 315–319.

STEINHAUS, H. (1960). *Mathematical Snapshots.* Oxford University Press, New York.

STONE, A. H. AND TUKEY, J. W. (1942). Generalized "sandwich" theorems. *Duke Math. J.* 9 356–359.

STROMQUIST, W. (1980). How to cut a cake fairly. *Amer. Math. Monthly* **87** 640-644.

STROMQUIST, W. AND WOODALL, D. R. (1985). Sets on which several measures agree. *J. Math. Anal. Appl.* **108** 241–248.

SVENSSON, L. G. (1983). On the existence of fair allocations. *J. Econ.* **43** 301–308.

URBANIK, K. (1955). Quelques theorémès sur les mesures. *Fund. Math.* **41** 150–162.

WELLER, W. (1985). Fair division of a measurable space. *J. Math. Econ.* **14** 5–17.

WOODALL, D. R. (1980). Dividing a cake fairly. *J. Math. Anal. Appl.* **78** 233–247.

WOODALL, D. R. (1985). A note on the cake–division problem. Preprint, University of Nottingham.

YOUNG, H. P. (1987). On dividing an amount according to individual claims or liabilities. *Math. Oper. Res.* **12** 398–414.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA 30332–0160