The Annals of Applied Probability
1995, Vol. 5, No. 3, 834–853

TARGET SHOOTING WITH PROGRAMMED RANDOM VARIABLES[1]

BY GRAHAM BRIGHTWELL, TEUNIS J. OTT AND PETER WINKLER

London School of Economics and Political Science, Bellcore and AT & T Bell Laboratories

We consider the following distributed optimization problem: Given a set $X_1, \dots, X_n$ of pairwise independent random variables and a target value T, a subset of the X_i's must be selected whose sum is close to T. However, no cooperation is permitted in determining the set; each variable must be "programmed" in advance, joining or not joining according to its own value. Such conditions may arise, for example, when supply of some commodity is controlled at several random sources. Under these general conditions we show that the mean square error in approximating T is always minimized by programming each X_i to join if its value is between 0 and θ_i, for appropriate choice of thresholds θ_i.

When in addition the variables are identically distributed, we give conditions under which the θ_i's must be equal. The case of uniform distribution on $[0, 1]$ (in which our conditions fail) is analyzed in detail, showing the rather bizarre behavior of the θ_i's which may take place in general as the target value is gradually changed.

We analyze also the problem in which the variables are permitted to contribute any *part* of themselves to the sum; here it turns out that in an optimal strategy, each program will be of the form "contribute the minimum of X_i and γ_i," with all the γ_i's equal in the i.i.d. case.

Finally, we show how the original target shooting problem can be generalized to a kind of load balancing, where variables are assigned to different buckets, each with its own target, and the penalty is a weighted sum of squared errors. The surprising result here is that when the weights are equal, an optimal solution assigns variables only according to their signs.

1. Introduction. The setting is Hollywood, California. A movie is being made and it is time to film the balloon scene. The prop engineer wants 1500 pounds of people on the balloon, for optimum performance. The director, megaphone in hand, is facing 46 "extras" of unknown weight. What should he tell them?

The setting is Waco, Texas. Each of six oil wells is connected by pipelines to a single storage tank, and also to a local refinery, and at any time can direct all of its output to one or to the other. The refinery is designed to handle a steady flow of 280 barrels per minute, and discrepancies must be

Received March 1994; revised December 1994.

[1]The bulk of the research explicated herein was carried out while the first author was visiting the second and third at Bellcore, under the auspices of DIMACS.

AMS 1991 *subject classification*. Primary 62C99.

Key words and phrases. Distributed optimization, load balancing.

corrected by expensive trucking of oil to or from the storage tank, but the well outputs are variable. How should the individual well outputs be directed?

The above situations are two examples of what we call *target shooting* (a third, motivating, example concerns load balancing in a telephone signalling network; unfortunately the details are beyond the scope of this work). An instance of a target shooting problem consists of a set of random variables, pairwise independent and of known distribution. In the balloon case these are the weights of the extras; in the oil case, the outputs of the wells. Some of the variables are to be selected, with the object of having their sum be close to a given target: in the balloon case 1500 pounds; in the oil case, 280 barrels per minute. The *error* or *discrepancy* in target shooting is the difference between the target value and the actual sum of the selected variables; we seek to minimize the expected value of the square of this quantity.

Many different kinds of constraints for the selected process are imaginable. Here are four possibilities, beginning with the ideal case where all the values of the random variables are known in advance:

1. *Offline*. The selection is made with knowledge of the values of each X_i. Although optimal selection can be NP-hard [2], the results even with a primitive algorithm can be expected to be excellent. (For example, with the distributions identical, continuous and fixed, the expected error will be exponentially small relative to n.) In our examples, one might imagine that the director has the time to ask each extra for his or her weight and that oil routing is controlled remotely by the refinery.
2. *Online*. The values of the X_i's are revealed in subscript order, and after each revelation that random variable must be selected or rejected. Imagine, for example, that the movie extras are weighted one by one and each told immediately whether to get on the balloon or go home. In the oil case, the refinery pipeline is routed serially through the wells and selection is made at each well on the basis of both output level of that well and current flow in the pipeline. In the online case one may typically expect errors on the order of c/n.
3. *Online initial segment*. Variables $X_1, \dots, X_k$ must be selected for some k, with the partial sums as online information. For example, the movie extras are lined up and marched onto the balloon, until the balloon's weight reaches (or nears) the target value. Here the expected error will be essentially constant relative to n.
4. *Fixed selection*. Any set of random variables may be selected, but with no advance knowledge of their values; that is, the extras are selected blindly and the wells are set up with fixed routing. Here of course the central limit theorem applies and we may expect discrepancies on the order of $c/\sqrt{n}$.

The conditions considered here fit somewhere between cases 2 and 4 above, though it will be seen that performance is not much better than in case 4. Suppose that each movie extra is presumed to know his or her own weight, but that the director must instruct everyone immediately. He may say

"Extras number 1 through 10, get on the balloon," as he might in case 4 above, but he can also say, "Everyone whose weight is between 130 and 145 pounds get on the balloon" or even something like, "Extras 1 through 5 get on the balloon, 6 through 10 get on only if you weigh less than 90 pounds, and the rest of you go home." Each extra is in effect a processor in a distributed system with limited information (its own weight) and no communication; this is the point of view taken by Papadimitriou and Yannakakis [3], who study an intriguing bin-packing problem under these (among other) conditions. (For a survey of approximation algorithms for bin-packing, the reader is referred to [1].)

In the oil case, each well is provided with a valve which can direct flow either to the refinery or to storage, depending on the current output of that well.

Thus, in our "programmed random variables" case, a *strategy* consists of a list of instructions to the random variables, each telling a random variable whether or not to join, based on its own value. We may thus identify an instruction with a subset S_i of the range of X_i, the instruction to X_i then being, "join if your value falls in the set S_i."

One might equally well consider more general "mixed" strategies, in which an instruction takes the form of a function f_i and variable X_i joins randomly with probability $f_i(X_i)$, but it is easy to see that any mixed strategy can be replaced by a pure one with the same or better performance. Nonetheless, mixed strategies can be useful in the analysis, as in the proof of Theorem 2.4 below.

Our first general result (Theorem 2.4) states that assuming only pairwise independence of the random variables X_i, there is always an optimal solution, and moreover there is an optimal solution in which each set S_i is a closed interval, one of whose endpoints is 0. Thus, when (as in our examples) the random variables are nonnegative, all instructions can be taken to be of the form "join if your value is at most θ_i."

One might expect that if the random variables are identically distributed, then the θ_i thresholds should all be equal, and that is indeed the case (Theorem 3.1) when the r.v.'s have a density function f satisfying $|x|f(x) \leq 1/2$.

In the general i.i.d. case, however, the θ_i's tend to be equal only for low values of the target; as the target value increases, instructions diversify. The behavior of the thresholds as a function of the target value is not generally either continuous or monotone. The case of uniform distributions on $[0, 1]$ is presented in Theorem 4.2 and the accompanying figure; it illustrates both the elegant and the grotesque features that are manifested in general.

We remark here that it is not simply the square penalty function that causes the bizarre behavior: calculations using a linear penalty function revealed a somewhat similar pattern, at least for small values of n. The linear penalty function may be more realistic in some circumstances (e.g., our oil well example), but, as will be seen, the selection of the penalty function as the square of the discrepancy is central to our analysis.

In Section 5, we consider a variation of target shooting which we call the partial contribution case: here each random variable may contribute any part of itself to the sum. The random variables X_i are taken to be nonnegative and an instruction becomes a function g_i from the range of X_i to the nonnegative reals such that $g(x) \le x$ for all x.

This makes no sense in the balloon scenario, but it is reasonable that the oil wells might be equipped with values which can direct any part of the flow to the local refinery and the rest to storage, depending on current output. We spend less time with the partial contribution cases, not because it is less plausible, but because the results are cleaner.

It turns out (Theorem 5.1) that in the partial contribution case, all instructions in an optimal solution may be assumed to be of the form "contribute the minimum of your value and γ_i." Moreover, we see in Theorem 5.2 that behavior in the i.i.d. case is much better than before: the cutoffs γ_i are all equal in the (essentially unique) optimal solution.

In the final section, we generalize the original target-shooting problem by stipulating that each random variable may be assigned to any of the k buckets. Each bucket will have its own target and its own weight, the final penalty now being the weighted sum of the squared errors; thus target shooting is the case $k = 2$ with weights 1 and 0.

When the weights are instead equal, we obtain a load-balancing problem with an unexpectedly simple solution. In particular, if the X_i's are nonnegative, there is an optimal solution in which the instructions require no input: that is, each variable is assigned to some bucket, regardless of the value of the variable.

2. General results. Let $X_1, \dots, X_n$ be pairwise independent random variables, taking values in $\mathbf{R}$. In this section, we will not assume that they are identically distributed. For S a measurable subset of $\mathbf{R}$, let $X_i^{(S)}$ be the random variable defined by

$$X_i^{(S)} = \begin{cases} X_i, & X_i \in S, \\ 0, & \text{otherwise.} \end{cases}$$

For T a real number and $S_1, \dots, S_n$ subsets of $\mathbf{R}$, define

$$\Delta_T(S_1, \dots, S_n) = \mathbf{E}\left(T - \sum_1^n X_i^{(S_i)}\right)^2.$$

We seek to minimize $\Delta_T(S_1, \dots, S_n)$. It is perhaps not entirely obvious that the minimum is attained: we shall prove this in due course.

In any case, we can rewrite the objective mean square error as a sum of squared bias plus variance, and since pairwise independence of the X_i's implies that the $X_i^{(S_i)}$ are uncorrelated, the variance of the sum of the $X_i^{(S_i)}$ equals the sum of their variances. Hence we have the following lemma.

LEMMA 2.1.

$$\Delta_T(S_1,\dots,S_n) = \left(T - \sum_1^n \mathbf{E}X_i^{(S_i)}\right)^2 + \sum_1^n \sigma^2\left(X_i^{(S_i)}\right).$$

Note that the above requires only pairwise independence of the X_i and that it reduces the objective function Δ_T to a function depending only on the expectation and variance of the individual $X_i^{(S_i)}$. Thus if $(X_1,\dots,X_n)$ and $(X_1',\dots,X_n')$ are two n-tuples of pairwise independent random variables, such that X_i and X_i' have the same distribution for each i, then the objective function Δ_T is the same for the X_i as for the X_i'. In particular, the two infima are the same and are attained (if at all) at the same values of S_i. In other words, the pairwise independent random variables X_i may be assumed to be mutually independent without affecting the analysis. Note, however, that it is not enough to assume merely that the X_i are pairwise *uncorrelated*: we need that all the variables $X_i^{(S_i)}$ are pairwise uncorrelated.

Lemma 2.1 also provides us with a method of approaching the minimization problem. The essence is as follows.

COROLLARY 2.2. *If* $S_1,\dots,S_n,S_1',\dots,S_k'$ $(k < n)$ *are subsets of* $\mathbf{R}$, *then*

$$\Delta_T(S_1,\dots,S_k,S_{k+1},\dots,S_n) \le \Delta_T(S_1',\dots,S_k',S_{k+1},\dots,S_n)$$

iff $\Delta_{T'}(S_1,\dots,S_k) \le \Delta_{T'}(S_1',\dots,S_k')$, *where*

$$T' = T - \sum_{j=k+1}^{n} \mathbf{E}X_j^{(S_j)}.$$

PROOF. Immediate from Lemma 2.1. □

A consequence is that if we have an optimal solution vector, then each subvector also solves a reduced minimization problem. This will be particularly useful to us in the case $k = 1$, where we conclude that any set which occurs in an optimal solution vector must solve a minimization problem for the case $n = 1$ and some target T'.

Let us then consider the case $n = 1$. We are looking for the set S minimizing $\mathbf{E}(T - X^{(S)})^2$. Then X must "join" if it is closer to T than 0 is, but not if it is farther; hence the minimum is attained by setting $S = \{x \in \mathbf{R}: |T - x| \le |T|\}$. Thus if $T \ge 0$, we take $S = [0, 2T]$ (or $[0, 2T)$), whereas if T is negative, we set $S = [2T, 0]$ (or $(2T, 0]$).

Thus we have the following theorem:

THEOREM 2.3. *For any n-tuple of sets* $(S_1,\dots,S_n)$, *there is an n-tuple* $(I_1,\dots,I_n)$ *of closed intervals, each having one endpoint* 0, *such that* $\Delta_T(I_1,\dots,I_n) \le \Delta_T(S_1,\dots,S_n)$.

For real θ, let $S(\theta)$ denote the closed interval between 0 and θ, and for real values $\theta_1,\dots,\theta_n$, let $X^{(\theta)} = X^{(S(\theta))}$ and $\Delta_T(\theta_1,\dots,\theta_n) = \Delta_T(S(\theta_1),\dots,S(\theta_n))$. Our problem is now to minimize $\Delta_T(\theta_1,\dots,\theta_n)$.

We are now in a position to prove our claim that there is an optimum strategy. This can be shown using techniques of weak convergence, but we prefer to proceed a little more directly. However, it is noteworthy that both of the proofs depend crucially on Theorem 2.3.

THEOREM 2.4. *For all pairwise independent random variables* $X_1, \dots, X_n$ *and all real* T, *there exist values* $\theta_1, \dots, \theta_n$ *such that* $\Delta_T(\theta_1, \dots, \theta_n) \leq \Delta_T(\phi_1, \dots, \phi_n)$ *for all real* $\phi_1, \dots, \phi_n$.

PROOF. It is convenient to extend the class of random variables $X_i^{(\theta_i)}$ slightly. We wish to allow mixed (i.e., randomized) instructions of the form: "join if your value is in the interval $[0, \theta_i)$; if it is equal to θ_i, join with probability p_i; do not join if it lies outside the interval $[0, \theta_i]$." Given a value $q \in [0, 1]$ and a real-valued random variable X, we define a random variable $X(q)$ as follows. Let $\theta = \inf\{x \in \mathbf{R}: \Pr(X \leq x) \geq q\}$. If $\Pr(X = \theta) = 0$, then $X(q) = X^{(\theta)}$. Otherwise, let $q_1 = \Pr(X < \theta)$ and $q_2 = \Pr(X \leq \theta)$, so $q_1 \leq q \leq q_2$, and set $p = (q - q_1)/(q_2 - q_1)$. Now, if $\theta \geq 0$, set

$$X(q) = \begin{cases} X, & X \in [0, \theta), \\ 0, & X \notin [0, \theta], \\ X, & \text{if } X = \theta, \text{ with probability } p, \\ 0, & \text{if } X = \theta, \text{ with probability } 1 - p, \end{cases}$$

and similarly if $\theta < 0$. Clearly the random variables $X(q)$ include the previously considered random variables $X^{(\theta)}$.

Now, the function $\mathbf{E}(T - \Sigma X_i(q_i))^2$ is continuous in the variables q_i and defined on the closed unit n-cube; therefore, it must have an absolute minimum. However, suppose the minimum is attained at some mixed strategy and let i be the least index for which X_i is used with probability p when $X_i = \theta_i$ for some $p \in (0, 1)$. From Corollary 2.2 and Theorem 2.3 we know that replacing p by 0 or 1 leaves the value of the objective function unchanged. It follows by induction that the minimum is also obtained at a pure (deterministic) strategy. □

The problem remains of finding the values $\theta_1, \dots, \theta_n$ which minimize $\Delta_T(\theta_1, \dots, \theta_n)$. We know from Corollary 2.2 that these satisfy

$$\theta_i = 2\left(T - \sum_{j \neq i} \mathbf{E} X_j^{(\theta_j)}\right) \tag{2.1}$$

for all i. We call the equations (2.1) the *fundamental equations*.

Now let $C = 2(T - \Sigma_{j=1}^n \mathbf{E} X_j^{(\theta_j)})$. The fundamental equations tell us that, at the optimum, each θ_i satisfies $\theta_i - 2\mathbf{E} X_i^{(\theta_i)} = C$.

THEOREM 2.5. *Suppose that each* X_i *has a density function* f_i *such that* $|x| f_i(x) \leq 1/2$ *for every* i *and* x, *and there is no interval on which* $|x| f_i(x) \equiv 1/2$ *for some* i. *Then the fundamental equations have a unique solution,* *giving the values of the θ_i minimizing $\Delta_T(\theta_1, \dots, \theta_n)$.*

PROOF. Suppose the X_i are as given. Then

$$\mathbf{E}X_i^{(\theta)} = \int_0^\theta |x| f_i(x)\, dx$$

(for negative as well as nonnegative θ) and so

$$\theta - 2\mathbf{E}X_i^{(\theta)} = \int_0^\theta (1 - 2|x| f_i(x))\, dx.$$

By assumption, the integrand is nonnegative and not equal to zero over any interval, so $\theta - 2\mathbf{E}X_i^{(\theta)}$ is a continuously increasing function of θ. Hence each threshold θ_i is determined uniquely by C, $\theta_i = \theta_i(C)$. Also, each term $\mathbf{E}X_j^{(\theta_j(C))}$ is continuously increasing in C, so the equation

$$C = 2\Big(T - \sum \mathbf{E}X_j^{(\theta_j(C))}\Big)$$

has a unique solution C for each value of T. □

3. Identical distributions. In this section, we shall assume that the X_i's are identically distributed, as X. We first note the immediate consequence of Theorem 2.5 for this case.

THEOREM 3.1. *Suppose that the X_i have a density function f such that $|x| f(x) \le 1/2$ for every x and there is no interval on which $|x| f(x) \equiv 1/2$. Then the optimum is obtained by taking all θ_i equal to the unique solution of*

$$\theta = 2(T - (n-1)\mathbf{E}X^{(\theta)}).$$

The conditions of Theorem 3.1 are satisfied for many commonly occurring distributions, but by no means for all. The more interesting phenomena occur when these conditions are not satisfied: we shall see that this essentially implies that the optimum will not always have the θ_i equal.

If the distribution X is not continuous and $\Pr(X = \theta) > 0$ for some θ, then $\frac{1}{2}\theta + (n-1)\mathbf{E}X^{(\theta)}$ is discontinuous at that value of θ. Therefore, there is a value T_0 not assumed by this function and for $T = T_0$ there can be no solution to the fundamental equations with all the θ_i equal. The other case we consider is where X has a density f, but we do not have $|x| f(x) \le 1/2$ everywhere.

THEOREM 3.2. *Suppose that $X_1, \dots, X_n$ are i.i.d. random variables with a density function $f(x)$. Suppose also that $\theta_0 f(\theta_0) > 1/2$ for some real θ_0. Then there is a value of T such that Δ_T is minimized at a point where not all the θ_i are equal.*

PROOF. We employ some elementary calculus to show that the point $(\theta_0, \dots, \theta_0)$ is not a local minimum for the function Δ_T, where $T = \frac{1}{2}\theta_0 +$

$(n-1)\mathbf{E}X^{(\theta_0)}$. This will imply that the optimum is attained at some other solution to the fundamental equations, which will not have all θ_i equal.

Note first that

$$\begin{aligned}\Delta_T(\theta_1,\dots,\theta_n) &= \Big(T - \sum_i \mathbf{E}(X^{(\theta_i)})\Big)^2 + \sum_i \sigma^2(X^{(\theta_i)})\\ &= \Big(T - \sum_i \int_0^{\theta_i} xf(x)\,dx\Big)^2 + \sum_i \int_0^{\theta_i} x^2 f(x)\,dx\\ &\quad - \sum_i \Big(\int_0^{\theta_i} xf(x)\,dx\Big)^2.\end{aligned}$$

Differentiating this expression with respect to θ_i yields

$$\begin{aligned}&-2\theta_i f(\theta_i)\Big(T - \sum_{j=1}^n \int_0^{\theta_j} xf(x)\,dx\Big) + \theta_i^2 f(\theta_i) - 2\theta_i f(\theta_i)\int_0^{\theta_i} xf(x)\,dx\\ &\quad = \theta_i f(\theta_i)\Big[-2T + 2\sum_{j\neq i}\int_0^{\theta_j} xf(x)\,dx + \theta_i\Big].\end{aligned}$$

This is 0 at a solution of the fundamental equations. We next compute the matrix of second derivatives at the point $(\theta_0,\dots,\theta_0)$. We have

$$\frac{\partial^2\Delta_T}{\partial\theta_i^2} = \theta_i f(\theta_i),$$

$$\frac{\theta^2\Delta_T}{\partial\theta_i\,\partial\theta_j} = 2\theta_i f(\theta_i)\theta_j f(\theta_j),$$

for $i \neq j$.

Thus the matrix of partial derivatives fails to be nonnegative definite if the submatrix

$$\begin{pmatrix} \theta_0 f(\theta_0) & 2\theta_0^2 f(\theta_0)^2 \\ 2\theta_0^2 f(\theta_0)^2 & \theta_0 f(\theta_0)\end{pmatrix}$$

has negative determinant, which is the case since $\theta_0 f(\theta_0) > 1/2$ by assumption. □

THEOREM 3.3. *Suppose that $X_1,\dots,X_n$ are i.i.d. random variables with density function $f(x)$. The following statements are equivalent*:

(i) $|x|f(x) \leq 1/2$ *for all* x.

(ii) *For every real T, $\Delta_T(\theta_1,\dots,\theta_n)$ is minimized by taking all θ_i's equal to the root of*

$$\theta = 2(T - (n-1)\mathbf{E}(X^{(\theta)})).$$

PROOF. That (ii) implies (i) follows from Theorem 3.2. The proof of Theorem 2.5 almost suffices to show that (i) implies (ii), the missing case being where $|x|f(x) = 1/2$ over an interval.

We know that, at a solution $(\theta_1, \dots, \theta_n)$ of the fundamental equations, all the θ_i have the same value of $\theta_i - 2\int_0^{\theta_i} |x| f(x)\, dx$. If $|x| f(x) \le 1/2$ for all x, then this is a nondecreasing function of θ_i, so all the θ_i lie in some interval $I = [a, b]$ over which $f(x) = 1/2|x|$. Since f is a density function, I cannot contain the origin; it is convenient to invoke symmetry and assume $0 < a < b$.

We set $A = \int_0^a x f(x)\, dx$ and $B = \int_0^a x^2 f(x)\, dx$, and calculate $\Delta_T(\theta_1, \dots, \theta_n)$ in the case where all the θ_i's lie in the interval I:

$$\begin{aligned}\Delta_T(\theta_1, \dots, \theta_n) &= \left(T - nA - \sum_i \int_a^{\theta_i} \frac{1}{2}\, dx\right)^2 \\ &\quad + nB + \sum_i \int_a^{\theta_i} \frac{x}{2}\, dx - \sum_i \left(A + \int_a^{\theta_i} \frac{1}{2}\, dx\right)^2 \\ &= \left(T - nA + \frac{1}{2} na - \frac{1}{2} \sum_i \theta_i\right)^2 + nB - \frac{na^2}{4} + \sum_i \frac{\theta_i^2}{4} \\ &\quad - n\left(A - \frac{a}{2}\right)^2 - \left(A - \frac{a}{2}\right) \sum_i \theta_i - \frac{1}{4} \sum_i \theta_i^2.\end{aligned}$$

The terms involving $\sum_i \theta_i^2$ cancel out and the remaining expression depends only on $\sum_i \theta_i$. Hence Δ_T is also minimized at $(\theta_0, \dots, \theta_0)$, where $\theta_0 = (1/n)\sum_i \theta_i$. □

If T is nonnegative, it might at first sight seem likely that all the thresholds θ_i can be taken nonnegative too. A simple example showing that this is not always so comes from considering two random variables X_1, X_2 so that $\Pr(X_i = 1) = 0.99$, $\Pr(X_i = -0.1) = 0.01$ and target 0.9. It is evident that the correct strategy is to set $\theta_1 = 1$ and $\theta_2 = -0.1$. The same phenomenon may occur when the X_i are normally distributed with mean 1 and sufficiently small variance. We now give a simple condition which prevents such pathological behavior.

PROPOSITION 3.4. *Suppose that $X_1, \dots, X_n$ are real-valued random variables such that $2|\mathbf{E} X_i^{(\theta)}| < |\theta|$ for all i and all real $\theta \ne 0$. Then, for all positive T, $\Delta_T(\theta_1, \dots, \theta_n)$ is minimized at a point where all θ_i are strictly positive.*

PROOF. Recall that, at the optimum, each θ_i satisfies $\theta_i - 2\mathbf{E} X_i^{(\theta_i)} = C$ for some constant C. The condition implies that this function of θ_i is positive if θ_i is positive, negative if it is negative and 0 if $\theta_i = 0$. Therefore, all the θ_i have the same sign as C. However, we also have $C = 2(T - \sum_i \mathbf{E}(X_i^{(\theta_i)}))$ and so, since T is positive, so are all the θ_i. □

If we weaken the conditions of Proposition 3.4 slightly to allow $\theta = 2\mathbf{E}\mathbf{X}_i^{(\theta)}$ for positive values of θ only, the conclusion is that all θ_i are nonnegative at the optimum.

Our final result of this section shows that, in normal circumstances, if the target is small, the optimum is obtained by taking all the θ_i equal.

THEOREM 3.5. *Suppose that $X_1, \dots, X_n$ are i.i.d. random variables such that $2|\mathbf{E}X_i^{(\theta)}| \le |\theta|$, with strict inequality if θ is negative. Suppose also that, in the range $(0, a]$, X_i has a density function $f(x)$ such that $xf(x) \le 1/2$. Then, for all $T \le a/2$, $\Delta_T(\theta_1, \dots, \theta_n)$ is minimized by taking all the θ_i equal.*

PROOF. We have that, at the optimum, all the θ_i will be nonnegative. Therefore, each θ_i will be at most $2T \le a$. In this range, we know that the optimum is attained with all θ_i equal. □

Theorem 3.5 is quite weak: we shall see shortly that often much more is true. However, it is in one sense best possible. If the X_i are distributed so that $\Pr(X_i = a)$ is 1, then for targets below $a/2$ it is correct to reject all X_i's, while for targets just above $a/2$, it is correct to accept one X_i and reject the rest.

4. The uniform case. In this section, we shall restrict attention to a specific case, namely, the one in which all the X_i are distributed as uniform $[0, 1]$ variables. This example seems to be fairly typical of the general behavior in this problem. We shall obtain the complete solution for the problem in this case.

The first thing to observe is that all thresholds $\theta \ge 1$ are equivalent, so that if the fundamental equations call for a value of θ larger than 1, this can be interpreted as the instruction to accept any value of X_i, that is, $X_i^{(\theta)} = X_i$.

Next, observe that the random variables have a density

$$f(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & \text{otherwise}, \end{cases}$$

so $xf(x) > 1/2$ in the range $(1/2, 1]$.

On the other hand, $2|\mathbf{E}(X_i^{(\theta)})| \le |\theta|$ for all θ, so the general theory tells us that, when T is positive, all the thresholds will be nonnegative. Also, for $T \le 1/4$, all the thresholds will be equal, but at some larger values this will cease to be the case. The intuition is that at small values of the target we should accept all suitably small values of X_i, whereas at larger values we should accept a certain number of the X_i regardless ($\theta_i = 1$) and tend not to take the rest.

Next we perform a few simple calculations. The expectation $\mathbf{E}X^{(\theta)}$ is given by $\int_0^\theta x\,dx = \theta^2/2$, for $0 \le \theta \le 1$. The variance $\sigma^2(X^{(\theta)})$ is equal to

$$\mathbf{E}(X^{(\theta)})^2 - (\mathbf{E}X^{(\theta)})^2 = \int_0^\theta x^2\,dx - \theta^4/4 = \theta^3/3 - \theta^4/4,$$

for $0 \le \theta \le 1$. For $\theta \ge 1$, the mean is $1/2$ and the variance is $1/12$. The form of the objective function thus depends on the number of θ_i which lie strictly in the interval $[0, 1)$. We suppose without loss of generality that $0 \le \theta_1 \le$

$\theta_2 \le \cdots \le \theta_m < 1 \le \theta_{m+1} \le \cdots \le \theta_n$. The objective function can then be expressed as

$$\Delta_T(\theta_1, \dots, \theta_n) = \left(T - \frac{1}{2}\sum_{i=1}^{m} \theta_i^2 - \frac{n-m}{2}\right) + \frac{1}{3}\sum_{i=1}^{m}\theta_i^2 - \frac{1}{4}\sum_{i=1}^{m}\theta_i^4 + \frac{n-m}{12},$$

and the fundamental equations say that, for $i \le m$,

$$\theta_i = 2\left(T - \frac{1}{2}\sum_{\substack{j=1\\ j\ne i}}^{m} \theta_j^2 - \frac{n-m}{2}\right). \tag{4.1}$$

Therefore, the value of $\theta_i - \theta_i^2 = \theta_i(1-\theta_i)$ is the same for each $i = 1, \dots, m$, and so the θ_i take on at most two distinct values below 1 and these two values sum to 1. Also, we have

$$1 \le 2\left(T - \frac{1}{2}\sum_{j=1}^{m}\theta_j^2 - \frac{n-m-1}{2}\right), \tag{4.2}$$

if $m < n$, since this is the fundamental equation for θ_n.

Let us next consider the case $n = 2$. This is of interest in its own right, but our primary purpose is to use Corollary 2.2 to deduce information about the general case. We consider the possibilities $m = 0, 1, 2$ in turn.

If $m = 0$, condition (4.2) gives $1 \le 2T - 1$ or $T \ge 1$.

If $m = 1$, condition (4.1) gives $\theta = 2T - 1$, which is possible if $1/2 \le T < 1$ [and satisfies condition (4.2)].

The interesting case is $m = 2$. Here we have the coupled equations

$$\theta_1 = 2T - \theta_2^2, \qquad \theta_2 = 2T - \theta_1^2.$$

One solution to this system is obtained by setting $\theta_1 = \theta_2$: this gives

$$\theta_1 = \theta_2 = \frac{\sqrt{1+8T} - 1}{2}.$$

Recall from Theorem 3.2 that such a solution only gives a local minimum when $\theta_i f(\theta_i) \le 1/2$, which in this case implies that $T \le 3/8$.

Solutions to the coupled equations with $\theta_1 \ne \theta_2$ satisfy $\theta_1 = 1 - \theta_2$, which gives $\theta_1 = 2T - (1-\theta_1)^2$, and so

$$\theta_1 = \frac{1 - \sqrt{8T-3}}{2}, \qquad \theta_2 = \frac{1 + \sqrt{8T-3}}{2}.$$

This solution is valid for $3/8 \le T < 1/2$.

Summarizing, we have the following picture of the optimal solution in the case $n = 2$. For values of T up to $3/8$, we should take $\theta_1 = \theta_2$. At $T = 3/8$, both values of θ are $1/2$. Between $T = 3/8$ and $T = 1/2$, θ_1 drops to 0 while θ_2 rises to 1. Between $T = 1/2$ and $T = 1$, θ_2 remains at 1 while θ_1 rises linearly from 0 to 1. Beyond $T = 1$, we do best to take $\theta_1 = \theta_2 = 1$.

We now move on to the general case: $n \ge 3$. If $m \le 2$, Corollary 2.2 tells us that the solution is as for $n = 2$, with a reduced target $T - (n-2)/2$.

If $m \geq 3$, we have two different types of solution. First, all the θ_i below 1 could take the same value; second, we could have one θ_i above 1/2 and the remaining θ_i below, with the two values summing to 1. (We cannot have more than one θ_i above 1/2 since by Corollary 2.2 they would solve a reduced problem with $n = 2$.)

Set $T^* = T - (n - m)/2$. For the first solution, we have

$$\theta_i = 2T^* - (m - 1)\theta_i^2$$

for each $i \leq m$, yielding

$$\theta_i = \frac{\sqrt{1 + 8T^*(m - 1)} - 1}{2(m - 1)}.$$

This gives a local minimum provided $T^* \leq (m + 1)/8$.

For the second type of solution, with $\theta_1 = \theta_2 = \cdots = \theta_{m-1} < 1/2$ and $\theta_m = 1 - \theta_1$, the fundamental equations give

$$\theta_1 = 2T^* - (m - 2)\theta_1^2 - (1 - \theta_1)^2,$$

$$\theta_1 = \frac{1 \pm \sqrt{1 + 4(m - 1)(2T^* - 1)}}{2(m - 1)}$$

The square root is defined when $T^* \geq 1/2 - 1/(8m - 8)$, and the negative root also requires $T^* \leq 1/2$.

The next step is to show that the positive root above does not correspond to a local minimum. As before, we form the matrix of partial derivatives. Recall that, at a stationary point,

$$\frac{\partial^2 \Delta_T}{\partial \theta_i^2} = \theta_i, \qquad \frac{\partial^2 \Delta_T}{\partial \theta_i \, \partial \theta_j} = 2\theta_i \theta_j,$$

so the matrix is

$$\begin{pmatrix} \theta_1 & 2\theta_1^2 & \cdots & 2\theta_1^2 & 2\theta_1\theta_m \\ 2\theta_1^2 & \theta_1 & \cdots & 2\theta_1^2 & 2\theta_1\theta_m \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2\theta_1^2 & 2\theta_1^2 & \cdots & \theta_1 & 2\theta_1\theta_m \\ 2\theta_1\theta_m & 2\theta_1\theta_m & \cdots & 2\theta_1\theta_m & \theta_m \end{pmatrix}.$$

This matrix has $m - 2$ eigenvectors of the form $(1, 0, \dots, 0, -1, 0, \dots, 0)^t$, with the -1 occurring in each of the rows $2, \dots, m - 1$. The remaining two eigenvectors are of the form $(1, 1, \dots, 1, x)^t$, with eigenvalue $\lambda(x)$. Such a vector is an eigenvector iff $(1, x)^t$ is an eigenvector of the matrix

$$\begin{pmatrix} \theta_1 + 2(m - 2)\theta_1^2 & 2\theta_1\theta_m \\ 2(m - 1)\theta_1\theta_m & \theta_m \end{pmatrix}$$

with eigenvalue $\lambda(x)$. Therefore the original matrix is nonnegative definite iff this reduced matrix has nonnegative determinant. The determinant in question is

$$\theta_1\theta_m(1 + 2(m-2)\theta_1 - 4(m-1)\theta_1\theta_m)$$
$$= \theta_1(1-\theta_1)(1-2\theta_1)(1-2(m-1)\theta_1),$$

on setting $\theta_m = 1 - \theta_1$. Since $0 \le \theta_1 \le 1/2$, the first three terms above are nonnegative. Thus a solution with $\theta_1 > 1/(2m-2)$ will not be a local minimum. This rules out the solution

$$\theta_1 = \frac{1 + \sqrt{1 + 4(m-1)(2T^* - 1)}}{2(m-1)}.$$

For each m, we are left with two candidate solutions:

(A) $$\theta_i = \frac{\sqrt{1 + 8T^*(m-1)} - 1}{2(m-1)} \quad \forall\, i, \text{ valid for } 0 \le T^* \le \frac{m+1}{8};$$

(B) $$\theta_1 = \cdots = \theta_{m-1} = \frac{1 - \sqrt{1 + 4(m-1)(2T^* - 1)}}{2(m-1)}, \qquad \theta_m = 1 - \theta_1,$$

$$\text{valid for } \frac{1}{2} - \frac{1}{8(m-1)} \le T^* \le \frac{1}{2}.$$

For each fixed m, we now compare solutions (A) and (B) in the range where (B) is valid. The reader may wish to skip the proof of Lemma 4.1 on first reading and proceed to Theorem 4.2.

LEMMA 4.1. *Suppose $m \ge 3$ and $1/2 - 1/(8m-8) \le T^* \le 1/2$. Solution* (A) *has a lower value of the objective function than solution* (B) *iff it has a higher value of*

$$\left| \frac{1}{2}\sum_{i=1}^{m} \theta_i^2 - T^*\left(\frac{m-2}{m-1}\right) - \frac{1}{4(m-1)} \right|.$$

PROOF. In both solutions (A) and (B), we have $\theta_i^2 = \theta_i - 2(T^* - U)$, for every i, where $U = \frac{1}{2}\sum_i \theta_i^2$. We shall use this identity to reduce the objective function Δ_{T^*} to a function of T^* and U.

Multiplying through by θ_i^2 and summing over i, we obtain

$$\tfrac{1}{4}\sum_{i=1}^{m} \theta_i^4 = \tfrac{1}{4}\sum_{i=1}^{m} \theta_i^3 - (T^* - U)U.$$

Therefore,

$$\Delta_{T^*}(\theta_1, \dots, \theta_m) = (T^* - U)^2 + \tfrac{1}{12}\sum_{i=1}^{m} \theta_i^3 + (T^* - U)U.$$

Similarly,

$$\frac{1}{12}\sum_{i=1}^{m}\theta_i^3 = \frac{U}{6} - \left(\frac{T^* - U}{6}\right)\sum_{i=1}^{m}\theta_i$$

and $\sum_{i=1}^{m}\theta_i = 2U + 2m(T^* - U)$. Combining all the above, we have

$$\begin{aligned}\Delta_{T^*}(\theta_1,\ldots,\theta_m) &= (T^* - U)^2 + \frac{U}{6} - \left(\frac{T^* - U}{3}\right)(U + m(T^* - U)) \\ &\quad + (T^* - U)U \\ &= \left(1 - \frac{m}{3}\right)(T^* - U)^2 + \frac{2}{3}U(T^* - U) + \frac{U}{6}.\end{aligned}$$

For fixed T^*, this function, $g(U)$, say, is quadratic in U, with leading term $(1 - m)U^2/3$. It is maximized when $g'(U) = 0$, which is when

$$U = T^*\left(\frac{m-2}{m-1}\right) + \frac{1}{4(m-1)},$$

and the result follows. □

For solution (A), the quantity mentioned in Lemma 4.1 turns out to be

$$\frac{1}{4(m-1)^2}\sqrt{1 + 8T^*(m-1)}\left(m - \sqrt{1 + 8T^*(m-1)}\right).$$

and for solution (B) it is

$$\frac{1}{4(m-1)^2}\sqrt{1 + 4(2T^* - 1)(m-1)}\left(\sqrt{1 + 4(2T^* - 1)(m-1)} + m - 2\right).$$

At the bottom of the range, $T^* = 1/2 - 1/(8m - 8)$, the first expression is positive and the second zero, so solution (A) is better. However, as T^* increases through the range, the second expression grows faster than the first. The conclusion is that if solution (B) is ever optimal, it is optimal at $T^* = 1/2$.

At the point $T^* = 1/2$, the first expression above is

$$\frac{\sqrt{4m-3}\left(m - \sqrt{4m-3}\right)}{4(m-1)^2},$$

while the second is just $1/(4m - 4)$. The second is larger iff $m \le 5$.

Therefore, for $m \ge 6$, solution (B) is never optimal, while for $3 \le m \le 5$, solution (B) beats solution (A) in some range $1/2 - \varepsilon < T^* \le 1/2$, where $\varepsilon < 1/(8m - 8)$.

We now have to compare solutions with different values of m. For the moment, let us just consider the solutions (A). Set $D(m, T)$ equal to the value of the objective function when $n - m$ of the θ_i are set to 1 and the remainder are set as in solution (A):

$$D(m,T) = \Delta_{T^*}(\phi_m, \dots, \phi_m) + \frac{n-m}{12},$$

$$\phi_m = \frac{\sqrt{1+8T^*(m-1)}-1}{2(m-1)}, \qquad T^* = T - \frac{n-m}{2}.$$

We shall regard $D(m,T)$ as a function of a real variable m, for $m > 1$, and take its derivative.

To begin with, ϕ_m satisfies $(m-1)\phi_m^2 + \phi_m = 2T - n + m$, so, differentiating throughout by m, we have

$$\phi_m^2 + (2(m-1)\phi_m + 1)\frac{d\phi_m}{dm} = 1.$$

Now $D(m,T)$ can be written in the form

$$D(m,T) = \left(\frac{\phi_m(1-\phi_m)}{2}\right)^2 + m\left(\frac{\phi_m^3}{3} - \frac{\phi_m^4}{4}\right) + \frac{n-m}{12},$$

and so

$$\frac{\partial D(m,T)}{\partial \phi_m} = \frac{1}{2}\phi_m(1-\phi_m)(1-2\phi_m) + m\phi_m^2(1-\phi_m)$$
$$= \frac{1}{2}\phi_m(1-\phi_m)(2(m-1)\phi_m + 1).$$

Therefore,

$$\frac{\partial D(m,T)}{\partial m} = \left(\frac{\phi_m^3}{3} - \frac{\phi_m^4}{4} - \frac{1}{12}\right) + \frac{\partial D(m,T)}{\partial \phi_m}\frac{d\phi_m}{dm}$$
$$= -\frac{(1-\phi_m)^2(3\phi_m^2 + 2\phi_m + 1)}{12} + \frac{1}{2}\phi_m(1-\phi_m)(1-\phi_m^2)$$
$$= \frac{(1-\phi_m)^2}{12}(-3\phi_m^2 - 2\phi_m - 1 + 6\phi_m + 6\phi_m^2)$$
$$= \frac{(1-\phi_m)^2}{12}(3\phi_m^2 + 4\phi_m - 1).$$

Now ϕ_m increases with m, so $D(m,T)$ decreases with m until $\phi_m = (\sqrt{7} - 2)/3 \simeq 0.23$, and increases thereafter.

Translating back, our conclusion is that to optimize over all solutions of type (A), we should take m to be one of the two integers either side of the value m_0 making ϕ_{m_0} equal to $(\sqrt{7} - 2)/3$. This value m_0 turns out to be

$$(n-2T)\frac{2\sqrt{7}+1}{6} + \frac{3-\sqrt{7}}{2}.$$

For each integer $m = 5, 6, \dots, n-1$, the target T can be chosen so that $m_0 = m$: for this value of the target, the optimal solution is a type (A) solution with exactly m thresholds less than 1. It remains to be shown that solutions of type (B) are optimal for some values of T. Recall that, for $m = 3, 4, 5$, the type (B) solution beats the type (A) solution in a range $1/2 - \varepsilon \leq T - (n - m)/2 \leq 1/2$, with $\varepsilon < 1/(8m - 8)$. In the range $(n - m + 1)/2 - 1/(8m - 8) \leq T \leq (n - m + 1)/2$, the value $m_0(T)$ lies between $m - 1$ and m. However, the type (A) solution with $m - 1$ non-1 θ_i's does not in fact satisfy the fundamental equations (T^* is less than 0, so the equations call for negative values of θ_i). Thus the only candidate for a type (A) solution in this range is that with just m non-1 θ_i's, and we know this is beaten by the type (B) solution.

Thus we have a complete picture of the optimal solution for every value of the target T. One other quantity of interest is the value of T at which we first get $m \neq n$: this is

$$n\left(\frac{11 - 4\sqrt{7}}{18}\right) + O(1).$$

We summarize the results in the following theorem and accompanying Figure 1.

THEOREM 4.2. *Let $X_1, \ldots, X_n$ be pairwise independent random variables with the uniform distribution on* $[0, 1]$ *and let T be a positive real target. Values of $\theta_1, \ldots, \theta_n$ minimizing $\Delta_T(\theta_1, \ldots, \theta_n)$ satisfy the following statements*:

(i) *For* $T \leq T_0(n) = n((11 - 4\sqrt{7})/18) + O(1)$, *all the θ_i are equal to* $(\sqrt{1 + 8T(n-1)} - 1)/(2(n-1))$.

(ii) *For each* $m = 5, 6, \ldots, n - 1$, *there is a range of T around*

$$T = \frac{9n - (4\sqrt{7} - 2)m + 7\sqrt{7} - 17}{18},$$

where $n - m$ of the θ_i's are 1 *and the rest are equal to* $(\sqrt{1 + 8T^*(m-1)} - 1)/(2(m-1))$, *where* $T^* = T - (n - m)/2$.

(iii) *For* $m = 5, 4, 3$, *there is an* $\varepsilon(m) < 1/(8m - 8)$ *such that, in the range* $-\varepsilon(m) \leq T - (n - m + 1)/2 \leq 0$, *we have $n - m$ of the θ_i's equal to* 1, $m - 1$ *of them equal to*

$$\theta_1 = \frac{1 - \sqrt{1 + 4(m-1)(2T^* - 1)}}{2(m-1)}$$

and the remaining one equal to $1 - \theta_1$.

(iv) *For* $m = 4, 3$, *in the range* $(n - m)/2 \leq T \leq (n - m + 1)/2 - \varepsilon(m)$, $(n - m)$ *of the θ_i's are* 1 *and the remainder are all equal, with value as given in* (ii) *above*.

(v) *For* $(n - 2)/2 \leq T \leq (4n - 5)/8$, $n - 2$ *of the θ_i's are* 1 *and the other two are equal to* $\frac{1}{2}(\sqrt{1 + 8(T - (n-2)/2)} - 1)$.

(vi) *For* $(4n - 5)/8 \leq T \leq (n - 1)/2$, $n - 2$ *of the θ_i's are* 1 *and the other two are given by* $\frac{1}{2}(1 \pm \sqrt{8(T - (n-2)/2) - 3})$.

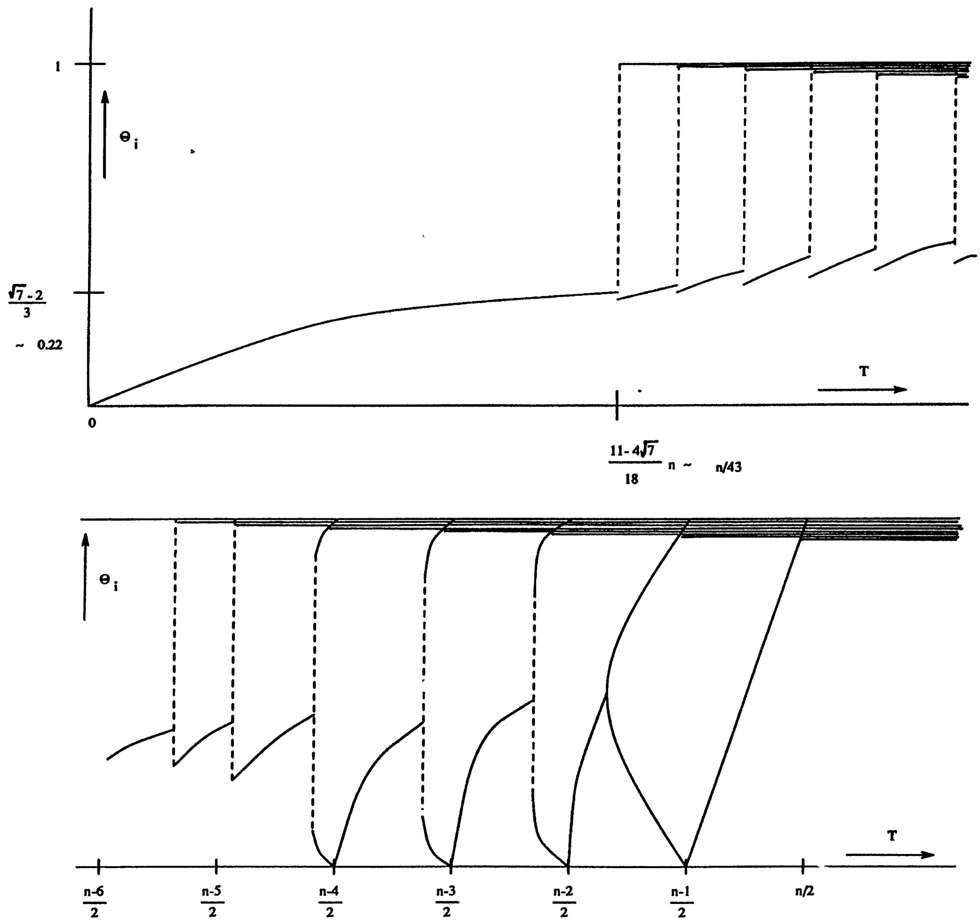

FIG. 1. *Behavior of the θ_i's as a function of T.*

(vii) *For* $(n-1)/2 \le T \le n/2$, *all but one of the* θ_i*'s are* 1 *and the remaining one is equal to* $2T-(n-1)$.

(viii) *For* $T \ge n/2$, *all the* θ_i*'s are equal to* 1.

5. The partial contribution case. We return finally to the general case of pairwise independent and *nonnegative* random variables $X_1, \dots, X_n$ with positive target T. Now, however, we consider measurable functions g_i from the range of X_i to the nonnegative reals such that $g(x) \le x$ for all x, and we define

$$\Delta_T(g_1, \dots, g_n) = \mathbf{E}\left(T - \sum_{i=1}^{n} g_i(X_i)\right)^2.$$

We seek to minimize $\Delta_T(g_1, \dots, g_n)$.

Thus, we allow in effect any random variable to contribute itself or any part of itself to the sum. Surprisingly, this greater freedom brings better behavior.

Note that if the sum of the X_i's happens to be (with probability 1) bounded below by some $B > T$, then there are multiple solutions with expected error 0. We call this the "overkill" case.

THEOREM 5.1. *In the partial contribution case, there is an optimal strategy in which every g_i is of the form $g_i(x) = \min(x, \gamma_i)$; in other words, each variable X_i is instructed to contribute all of itself up to γ_i. Moreover, this strategy is unique up to measure 0 except in the overkill case, and the γ_i can be taken to be the unique solutions to the equations*

$$\gamma_i = T - \sum_{j \neq i} \mathbf{E}\left(\min\{X_j, \gamma_j\}\right).$$

PROOF. Proofs of the analogous forms of Lemmas 2.1 and 2.2, with sets S_i replaced by functions g_i, go through without a hitch; thus we may again reduce to the $n = 1$ case. With one random variable X and target T, however, it is immediate that $\Delta_T(g)$ is uniquely minimized by the function $g(x) = \min(x, T)$. Since this function is given (possibly in more than one way) by a threshold γ, we conclude that all the g_i's may be defined by γ_i's as claimed.

Moreover, if T_i is the target for X_i with the other programs already chosen, then $T_i = T - \sum_{j \neq i} \mathbf{E}(\min\{X_j, \gamma_j\})$. We may as well take $\gamma_i = T_i$ satisfying the equations of the theorem.

To show uniqueness, choose an optimal solution and define the γ_i's as above. Then $K \equiv \gamma_i - \mathbf{E}g_i(X_i)$ is independent of i. This function is strictly increasing in γ_i, so the value of K determines all the $\gamma_i = \gamma_i(K)$. It remains only to note that except in the overkill case, the equations $K = \gamma_i - \mathbf{E}(\min\{x, \gamma_i(K)\})$ admit only a single solution. □

THEOREM 5.2. *When the variables are identically distributed, the cutoffs γ_i of the optimal partial contribution strategy may be taken to be equal, with value given by the unique root of the equation*

$$\gamma = T - (n-1)\mathbf{E}\left(\min\{x, \gamma\}\right). \qquad \square$$

6. Load balancing. The original target shooting problem generalizes to what we call load balancing in the following manner. Suppose that there are k buckets B_j, each with a target T_j and a nonnegative weight w_j which reflects the importance we attach to hitting that bucket's target. Each random variable X_i must assign itself to one bucket, according to its own value; thus we may encode the ith set of instructions as a (measurable) function h_i from the range of X_i to the set $\{1, 2, \ldots, k\}$ of bucket indices.

Our penalty will now be the expected value of the weighted sum of squared errors, namely,

$$\Delta(h_1,\dots,h_n) = \mathbf{E}\left(\sum_{j=1}^{k} w_j \left(T_j - \sum_{h_i(X_i)=j} X_i \right)^2 \right).$$

When $k = 2$, $w_1 > 0$, $w_2 = 0$ and $T_1 = T$, we are back to target shooting, in which case we know from Theorem 2.3 that there is an optimal solution in which each h_i is defined to be 1 on an interval $[0, \theta_i]$ (or $[\theta_i, 0]$) and 2 elsewhere. The load balancing version is a generalization of this.

THEOREM 6.1. *In the load balancing case, there is always an optimal solution in which each h_i can be described as follows: There is a partition of the real line into intervals $I_1, I_2, \dots, I_k$, some of which may be empty or unbounded, such that $h_i(X_i) = j$ iff X_i lies in I_j.*

PROOF. It is easily seen that the analogue of Corollary 2.2 carries through for our new penalty function, enabling us once again to reduce to the $n = 1$ case. From this reduction the above statement (and more) can be derived in straightforward manner. □

Of particular interest, naturally, is the equal weight case. Let us note first that when $w_1 = w_2 = \cdots = w_k$, any constant additive shift of the targets has no effect on the problem. For example, setting all targets equal to 0 (or to any other constant) is tantamount to trying to balance the loads equally.

Now Theorem 6.1 simplifies drastically.

THEOREM 6.2. *In load balancing with equal weights, there is always an optimal solution in which for each variable X_i there are two buckets B_{i^+} and B_{i^-} such that X_i goes into B_{i^+} if it is nonnegative and into B_{i^-} otherwise.*

It follows, of course, that when the random variables have nonnegative ranges to begin with, they can be assigned optimally to buckets *without looking*—by no means an intuitively obvious result, at least to us. This theorem suggests, even more than earlier results, that our no-cooperation, no-feedback constraints are very severe indeed.

The possible unification of Sections 5 and 6 seems to call for a final remark. Suppose each variable may partition itself at will, sending a (possibly empty) part to each bucket. Then again the usual reduction to the $n = 1$ case can be effected and one can verify that there is a well-behaved optimal solution, in which the partition of X_i varies continuously with X_i. We leave further analysis of this case to the future.

REFERENCES

[1] COFFMAN, E. G., GAREY, M. R. and JOHNSON, D. S. (1984). Approximation algorithms for bin packing—an updated survey. In *Algorithm Design for Computer System Design* (G. Ausiello, M. Lucertini and P. Serafini, eds.) 49–106. Springer, New York.

[2] GAREY, M. R. and JOHNSON, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, New York.

[3] PAPADIMITRIOU, C. and YANNAKAKIS, M. (1991). On the value of information in distributed decisionmaking. In *Proceedings of the 10th ACM Symposium on Principles of Distributed Computing* 61–64.

G. BRIGHTWELL
LONDON SCHOOL OF ECONOMICS
AND POLITICAL SCIENCE
HOUGHTON STREET
LONDON
UNITED KINGDOM

T. OTT
BELLCORE
445 SOUTH STREET
MORRISTOWN, NEW JERSEY 07962

P. WINKLER
AT & T BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974