Target Shooting With Programmed Random Variables

Graham Brightwell
London School of Economics
Houghton Street, London
U.K.

Teun Ott
Bellcore
445 South St., Morristown
New Jersey

Peter Winkler
Bellcore
445 South St., Morristown
New Jersey

ABSTRACT.

Let $X_1, \ldots, X_n$ be pairwise independent random variables of known (but not necessarily identical) distribution; we wish to select a subset of these whose sum will be as close as possible to some known target value T. Conditions described below force the selections to be made by a primitive distributed system (similar to one considered by Papadimitriou and Yannakakis [2] in PODC '91); here we are able to obtain a surprising amount of information about optimal solutions.

The conditions are that each variable must be "programmed" in advance, joining the selected set according to its own value. Thus, for example, one variable might be programmed to join just if its value lies between α and β, while another is told to join regardless of its value. Our object is to find a strategy, that is, a collection of programs, which minimizes the mean square error in approximating T. Typical applications involve producing a steady flow of some commodity when supply is controlled at a multiplicity of random sources.

It turns out that there is always an optimal strategy in which each X_i is programmed to join if its value is between 0 and θ_i, for appropriate choice of thresholds θ_i. When the variables are identically distributed, we examine conditions under which the θ_i's must be equal. The case of uniform distributions on $[0, 1]$, for which the above conditions are *not* satisfied, 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.

Next, we analyze 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 6 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 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 are two examples of what we call "target shooting." A set of random variables, pairwise independent and of known distribution, is given; 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 certain target; in the ballon case 1500 pounds, in the oil case 280 barrels per minute. The *error* or *discrepancy* in target shooting is the absolute 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 selection process are imaginable. Here are four possibilities, beginning with the ideal case where all the values of the random variables are known in advance. In each case, the performance estimates are based on the assumption that the distributions are identical, fixed, continuous and non-negative, while the number n of random variables is increasing, and the target value remains roughly proportional to n.

(1) *Offline*

The selection is made with knowledge of the values of each X_i. Although optimal selection can be NP-hard [1], the results even with a primitive algorithm can be expected to be excellent. Under the above conditions, the expected error will be exponentially small. 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 weighed 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 it is easily seen that one may expect a discrepancy on the order of c/n.

(3) *Online initial segment*

Variables $X_1, \ldots, X_k$ 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 discrepancy will be asymptotically 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 (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 [2] who study an intriguing bin-packing problem under these (among other) conditions.

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 non-negative, 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 4.1 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 all 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 non-negative and an instruction becomes a function g_i from the range of X_i to the non-negative reals such that $g(x) \leq x$ for all x.

This makes no sense in the balloon scenario, but it is reasonable that the oil wells might be equipped with valves 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 case not because it is less plausible but because the results are neater.

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 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 non-negative, 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 $X_1, \ldots, X_n$ 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 $S_1, \ldots, S_n$ subsets of $\mathbf{R}$, define

$$\Delta_T(S_1, \ldots, S_n) = \mathbf{E}\left(T - \sum_{i=1}^{n} X_i^{(S_i)}\right)^2.$$

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

Lemma 2.1.

$$\Delta_T(S_1, \ldots, S_n) = \left(T - \sum_{i=1}^{n} \mathbf{E}X_i^{(S_i)}\right)^2 + \sum_{i=1}^{n} \sigma^2(X_i^{(S_i)}).$$

Proof. Routine expansion. □

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 $(X_1, \ldots, X_n)$ and $(X_1', \ldots, X_n')$ 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 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 $S_1, \ldots, S_n, S_1', \ldots, S_k'$ $(k < n)$ are subsets of $\mathbf{R}$, then*

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

iff $\Delta_{T'}(S_1, \ldots, S_k) \leq \Delta_{T'}(S_1', \ldots, S_k')$, 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 $k = 1$, which tells us 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 then 0 is, but not if it is farther; hence the minimum is attained by setting $S = \{x \in \mathbf{R} : |T - x| \leq |T|\}$. Thus if $T \geq 0$, 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 2.3. *For any n-tuple of sets $(S_1, \ldots, S_n)$, there is an n-tuple $(I_1, \ldots I_n)$ of closed intervals, each having one endpoint 0, such that $\Delta_T(I_1, \ldots, I_n) \leq \Delta_T(S_1, \ldots, S_n)$.* □

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

We are now finally 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 $X_1, \ldots, X_n$ and all real T, there exist values $\theta_1, \ldots, \theta_n$ such that $\Delta_T(\theta_1, \ldots, \theta_n) \leq \Delta_T(\phi_1, \ldots, \phi_n)$ for all real $\phi_1, \ldots, \phi_n$.*

Proof. It is convenient to extend the class of random variables $X_i^{(\theta_i)}$ slightly. We wish to allow mixed (that is, 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 $\mathbf{E}(T - \sum X_i(q_i))^2$ is continuous in the variables q_i and defined on the closed unit n-cube, therefore must have an absolute minimum. But, if the minimum is attained at any mixed strategy, say with X_i being used with probability p when $X_i = \theta_i$, then the objective function must remain the same if we replace p by either 0 or 1. Hence the minimum is also obtained at a pure (deterministic) strategy. □

The problem remains of finding the values $\theta_1, \ldots, \theta_n$ which minimize $\Delta_T(\theta_1, \ldots, \theta_n)$. We know from Corollary 2.2, and the discussion of the case $n = 1$, that these satisfy:

$$\theta_i = 2\left(T - \sum_{j \neq i} \mathbf{E}X_j^{(\theta_j)}\right) \quad (1)$$

for all i. We call the equations (1) the *fundamental equations.*

Now let $C = 2\left(T - \sum_{j=1}^n \mathbf{E}X_j^{(\theta_j)}\right)$. 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 values of the θ_i minimizing $\Delta_T(\theta_1, \ldots, \theta_n)$.*

Proof. Suppose the X_i are as given. Then

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

and so

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

By assumption, the integrand is non-negative, 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 continuous and non-decreasing in C, so the equation

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

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 $|x|f(x) \leq 1/2$ 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*

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

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 satisified: 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 $|x|f(x) \leq 1/2$ everywhere.

Theorem 3.2. *Suppose that $X_1, \ldots, X_n$ 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 $(\theta_0, \ldots, \theta_0)$ 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

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

Differentiating this expression with respect to θ_i yields

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

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

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

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

for $i \neq j$.

Thus the matrix of partial derivatives fails to be non-negative 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 $X_1, \ldots, X_n$ are i.i.d. random variables with density function $f(x)$. The following are equivalent.*

(1) $|x|f(x) \le 1/2$ for all x.

(2) For every real T, $\Delta_T(\theta_1, \ldots, \theta_n)$ is minimized by taking all θ_is equal to the root of

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

Proof. Most of this is contained in Theorems 3.1 and 3.2. The one remaining case is where $|x|f(x) \le 1/2$ for all x, but there is an interval over which $|x|f(x) = 1/2$. In this case, it turns out that, for some values of T, there is more than one setting of the θ_is attaining the optimum, including that given above with all the θ_i equal. We omit the details. □

If T is positive, it might at first sight seem likely that all the thresholds θ_i can be taken non-negative too. A simple example showing that this is not always so comes from considering two random variables X_1, X_2 so that $\Pr(X_i = 1) = .99$, $\Pr(X_i = -.1) = .01$, and target $.9$. It is evident that the correct strategy is to set $\theta_1 = 1, \theta_2 = -.1$. This is not just a pathological example: it is easy to see that the same phenomenon occurs when the X_i are normally distributed with mean 1 and sufficiently small variance, and the target is just below some natural number. However, we can give a simple necessary condition for this behavior to occur.

Proposition 3.4. *Suppose that $X_1, \ldots, X_n$ 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, $\Delta_T(\theta_1, \ldots, \theta_n)$ 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. But 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 $\theta = 2\mathbf{E}X_i^{(\theta)}$ for positive values of θ only, the conclusion is that all θ_i are non-negative 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 $X_1, \ldots, X_n$ 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 $T \le a/2$, $\Delta_T(\theta_1, \ldots, \theta_n)$ is minimized by taking all the θ_i equal.*

Proof. We have that, at the optimum, all the θ_i will be non-negative. 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 general 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 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 obtain the complete solution for the problem in this case, but the analysis is lengthy and cannot be included here.

The results are summarized in the following theorem and illustrated in the accompanying diagram. Roughly speaking, as the target value runs from 0 up to $n/2$, the θ_i's begin rising slowly together; then they start jumping to 1, one at a time. When only five θ_i's are left below 1 their behavior changes, with one at a time jumping partway to 1 and then rising to 1 as the others descend (!) to 0. A curious consequence is that a variable can be told *never* to join only when T has one of the five values 0, $n-2$, $n-3/2$, $n-1$ or $n-1/2$.

Theorem 4.1. *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.*

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

$$\frac{\sqrt{1 + 8T(n-1)} - 1}{2(n-1)}.$$

(2) For each $m = 5, 6, \ldots, n-1$, there is an interval for T including

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

where $n - m$ of the θ_i's are 1, and the rest are equal to

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

where $T^* = T - (n-m)/2$.

(3) For $m = 5, 4, 3$, there is an $\epsilon(m) < 1/(8m - 8)$ such that, in the range $-\epsilon(m) \le T - (n - m + 1)/2 \le 0$, 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 $1 - \theta_1$.

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

(5) For $(n-2)/2 \le T \le (4n-5)/8$, $n-2$ of the θ_i's are 1, and the other two are equal to

$$\frac{\sqrt{1 + 8(T - (n-2)/2)} - 1}{2}.$$

(6) For $(4n-5)/8 \le T \le (n-1)/2$, $n-2$ of the θ_i's are 1, and the other two are given by

$$\frac{1 \pm \sqrt{8(T-(n-2)/2)-3}}{2}.$$

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

(8) 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 *non-negative* random variables $X_1, \ldots, X_n$ with positive target T. Now, however, we consider measurable functions g_i from the range of X_i to the non-negative reals such that $g(x) \le x$ for all x, and we define

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

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

Thus, we allow in effect any random variable to contribute itself or any part of itself to the sum. As sometimes happens with children, 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, with the γ_i being the unique solutions to the equations*

$$\gamma_i = T - \sum_{j \ne i} \mathbf{E}(\min\{x, \gamma_j\}).$$

Proof. The proofs of these assertions are similar to those in Section 2: the details are omitted, and we merely sketch the argument showing that the optimal strategy is unique.

We observe (by considering the $n = 1$ case) that optimal settings of the γ_i do indeed satisfy the given equations. Thus at the optimum, $K \equiv \gamma_i - \mathbf{E}g_i(X_i)$ is independent of i. It is easy to see that this function is strictly increasing in γ_i, so the value of K determines all the $\gamma_i = \gamma_i(K)$. Further, one can check 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*

$$\gamma = T - (n-1)\mathbf{E}(\min\{x, \gamma\}).$$ □

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 non-negative 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:

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

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 $I_1, I_2, \ldots, I_k$, 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 to reduce to the $n = 1$ case as before. From this reduction the above statement (and more) can be derived. □

Of particular interest, naturally, is the equal weight case. Let us note first that when $w_1 = w_2 = \ldots = w_k$, 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 non-negative and into B_i^- otherwise.* □

It follows, of course, that when the random variables have non-negative ranges to begin with, they can be assigned optimally to buckets *without looking*—by no means an intuitively obvious result, at least to us. Of course, "optimally" means here according to our no-feedback constraints, which are therefore seen to be 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 the details, for now, to the reader.

REFERENCES

[1] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, New York (1979).

[2] C. Papadimitriou and M. Yannakakis, On the value of information in distributed decision-making, *Proc. 10th ACM Symposium on Principles of Distributed Computing* (1991) 61–64.

This research was carried out while the first author was visiting Bellcore under the auspices of DIMACS.