Designing Samples to Satisfy Many Variance Constraints

Donna Glassbrenner
United States Department of Transportation
National Highway Traffic Safety Administration
Washington, DC 20590

Abstract: Chromy's algorithm is frequently used to design stratified simple random samples to meet several variance constraints on a Horvitz Thompson estimator. We present a new geometrically-based algorithm for "nested" constraints. This algorithm produces a list of points that we prove contains the optimal solution.

Key words: sample design, stratified samples, reliability constraints, convex programming

0. Introduction

Complex surveys often have complex requirements. For instance, the Census Bureau's annual and monthly surveys of the retail, wholesale, and service trade areas need to produce reliable estimates in a large number of individual and aggregated kinds of business.

Designing samples for such surveys is a nontrivial constrained optimization problem, and iterative methods based on the Karush-Kuhn-Tucker conditions are believed to determine a solution. For instance, Chromy's algorithm is frequently used to determine sample sizes in a stratified design that meet multiple coefficient of variation (cv) constraints. [B],[C]

However, the constraints often have some additional structure. For example, those for the above Census Bureau surveys have a "nested" structure, in that certain "detail" kinds of business are constrained, and then various aggregations of these detail are constrained. One might expect to be able to determine a solution more easily in this simpler scenario, and prove that it works.

Members of Census Bureau developed an alternative algorithm that is easier to program and appears to produce an approximately optimal solution. [K1],[K2] The Census Bureau has used their algorithm to design samples for various economic surveys, including the Monthly Retail Trade Survey.

In this paper, we present a new procedure, inspired by the Census Bureau algorithm, and prove that it solves the problem. That is, we present and prove an algorithm that finds optimal sample sizes meeting "nested" univariate cv constraints of a Horvitz-Thompson estimator under stratified simple random sampling. Our algorithm takes a geometric approach to the problem, instead of the analytic one taken by Chromy and Bethel. We produce a list of points related to the faces of the constraint region, among which is the optimal point.

1. The Problem and its Transformation

Example 1

Suppose, for example, that you have a frame with four strata and a variable of interest Y. Suppose that the stratum sizes are N_1:=10, N_2:=20, N_3:=30, and N_4:=10, the totals of Y in the strata are Y_1:=30, Y_2:=90, Y_3:=90, and Y_4:=50, and the standard deviations of Y in the strata are σ_1:=1, σ_2:=2, σ_3:=1, σ_4:=2. That is, stratum 1

Quantity to estimate	**Upper bound on the coefficient of variation (cv)**
$Y_1 + Y_2 + Y_3 + Y_4$	0.5%
$Y_1 + Y_2$	1%
$Y_3 + Y_4$	1%

You plan to use a Horvitz Thompson estimator (e.g. you will use the sample weight $10/x_1$ in stratum 1, if you choose x_1 members from stratum 1), and want to choose at least two members from each stratum. (You can think of these bounds as arising from variance constraints on the four stratum estimates.) Your goal is to determine the stratum sample sizes x_1, x_2, x_3, and x_4 that meet these constraints with $x_1+x_2+x_3+x_4$ minimal. The corresponding variance constraints can be expressed as

$$\begin{aligned}
100/x_1 + 1600/x_2 + 900/x_3 + 400/x_4 &\leq 161.69 \\
100/x_1 + 1600/x_2 \phantom{{}+ 900/x_3 + 400/x_4} &\leq 91.44 \\
900/x_3 + 400/x_4 &\leq 71.96
\end{aligned}$$

and you have the additional constraints $2 \leq x_1 \leq 10$, $2 \leq x_2 \leq 20$, $2 \leq x_3 \leq 30$, and $2 \leq x_4 \leq 10$. You wish to determine the sample sizes x_1, x_2, x_3, and x_4 that satisfy these constraints with $x_1+x_2+x_3+x_4$ minimal. We will return to this example to illustrate our result.

In general, suppose you have N strata and variance constraints var_A

Example 1 is a small instance of this problem. The sample design for the Monthly Retail Trade Survey uses approximately N=800 strata and approximately |C|=100 reliability constraints. Applying the transformation $x_i \to a_i / x_i$ for i=1, ..., N, we obtain an equivalent problem.

The Transformed Problem

Let N be a positive integer, and let $a_1,\ldots, a_N$, $c_1,\ldots, c_N$, and $C_1,\ldots, C_N$ be positive real numbers. Let $C \subseteq S(N)$, and let c_A, for $A \in C$, be positive real numbers. Minimize $a_1/x_1 + \ldots + a_N/x_N$ subject to

$$\sum_{i \in A} x_i \le c_A \quad \text{for all } A \in C\text{, and } c_i \le x_i \le C_i \text{ for all } 1 \le i \le N.$$

That is, a point $p=(p_1,\ldots, p_N)$ satisfies the Problem if and only if the point $q=(q_1,\ldots, q_N)$ defined by $q_i := a_i/p_i$ for i=1,..., N, satisfies the Transformed Problem with variable constraints $a_i/C_i \le x_i \le a_i/c_i$ for all i=1,...,N.

In Example 1, the Transformed Problem is to minimize $100/x_1 + 1600/x_2 + 900/x_3 + 400/x_4$ subject to $x_1+x_2+x_3+x_4 \le 161.69$, $x_1+x_2 \le 91.44$, $x_3+x_4 \le 71.96$, $10 \le x_1 \le 50$, $80 \le x_2 \le 800$, $30 \le x_3 \le 450$, and $40 \le x_4 \le 200$.

Note that the Transformed Problem has a solution when the constraint region is nonempty, since the objective function is continuous and the constraint region is compact. Furthermore, the Transformed Problem has a unique solution p, since the objective function is strictly convex and the constraint region is convex. Moreover the solution p is on the boundary of the constraint region, so it satisfies a subcollection of the constraints with equality. These statements hold for any collection of mutually satisfiable constraints C. Our main result (Theorem 1) describes the solution to the Transformed Problem in more detail for *nested* collections C.

Defn A collection C of sets is *nested* if C can be partitioned into subcollections C(0), ..., C(K) for some $K \ge 0$ such that the members of each C(k) are pairwise disjoint, and for k < K, each member of C(k+1) is the union of two or more members of C(k). The members of C(k) are called the *level k sets*, and we write Lev(A):=k for all $A \in C(k)$.

For instance the collection {{1,2}, {3}, {1,2,3}} is nested with C(0) = {{1,2}, {3}} and C(1) = {{1,2,3}}. We shall see in Section 3 that the decomposition of C into C(0), ..., C(K) is unique and so members of C can't be construed to have different levels.

Our solution of the Transformed Problem narrows down the search for a minimum to a relatively easily computed list of points that contains the minimum. We present our solution in Section 2. Section 3 contains preliminary results, and Section 4 contains the proof.

Notation

Throughout this paper, N denotes a positive integer, and $a_1,\ldots, a_N$, $c_1,\ldots, c_N$, and $C_1,\ldots, C_N$ are positive real numbers. Also $S(N) := \{A \subseteq \{1,\ldots, N\}: |A| > 1\}$, and $C \subseteq S(N)$ is nested, and is partitioned into subcollections C(0), ..., C(K) as in the definition of a nested collection. Also c_A, for $A \in C$, are positive real numbers. We define $V:=\{\{1\},\ldots,\{N\}\}$, the singleton sets, and for all $1 \le i \le N$, $c_{\{i\}} \in \{c_i, C_i\}$. We define the *constraint region defined by C* to be

$$CR(C) := \left\{\sum_{i \in A} x_i \le c_A : A \in C\right\} \cap \left\{c_i \le x_i \le C_i : 1 \le i \le N\right\}$$

For any subcollection D of $S(N) \cup V$, we define the *equality constraint region defined by D* to be

$$EQ(D) := \left\{\sum_{i\in A} x_i = c_A : A \in D\right\}$$

so CR(C) and EQ(D) are subsets of R^N. Note that CR(C) and EQ(D) depend on the c_A and that EQ(D) can involve the $c_{\{i\}}$ for singletons {i}.

2. Statement of the Algorithm

Defn A collection $D \subseteq S(N) \cup V$ is *complete* if for all $1 \le k \le N$, there exists $A\in D$ such that $k\in A$.

Defn Let D be a subcollection of $C\cup V$, and let $A\in D$. Define Max(A,D) := $\{B\in D$: B is a maximal proper subset of A}.

Here we mean that B is maximal with respect to containment. We call the members of Max(A,D) the *maximal subsets of A (in D)*. E.g. for N:=4 and D:={{1,2},{3}, {1,2,3,4}}, Max({1,2,3,4},D) = {{1,2},{3}}. For $D:=C \cup V$, Max(A,D) consists of the level Lev(A)-1 sets in the decomposition of A into level Lev(A)-1 sets. We shall see for a complete subcollection D of $C \cup V$, the nonempty sets $A \setminus \cup$Max(A,D), for $A \in D$, partition {1,...,N}. This needn't be true when C isn't nested, e.g. for C=D={{2},{3},{1,2},{1,3}}.

Example: For N:=4 and D:={{1,2},{3}, {1,2,3,4}}, $\{1,2\} \setminus \cup$Max({1,2},D) = {1,2}, $\{3\} \setminus \cup$Max({3},D) = {3}, and $\{1,2,3,4\} \setminus \cup$Max({1,2,3,4},D) = {4}.

Defn Let D be a subcollection of $C\cup V$. We define D to be *union-free* if there do not exist n and A, $A_1,\dots, A_n \in D$ with $A = A_1 \cup \dots \cup A_n$.

For example, {{1},{1,2,3}} is union-free, while {{1},{2,3},{1,2,3}} is not. Note that every complete collection D has a union-free complete subcollection E. (You can create E by iteratively omitting sets A from D that are unions of other sets in D.) We now state the main theorem, which presents the algorithm for solving the Transformed Problem for nested constraints. The algorithm presents a list containing the minimum.

Theorem 1 Let $C \subseteq \{A \subseteq \{1,..., N\}: |A| > 1\}$ be nested, and suppose

$$CR(C) := \left\{\sum_{i\in A} x_i \le c_A : A \in C\right\} \cap \left\{c_i \le x_i \le C_i : 1 \le i \le N\right\}$$

is nonempty. Let $p=(p_1,\dots,p_N)\in R^N$ be the unique minimum of $\sum_{i=1}^N a_i / x_i$ on CR(C). Then there are a union-free complete subcollection $D \subseteq C\cup V$ and constants $c_{\{i\}}\in\{c_i, C_i\}$ for all singletons $\{i\}\in D$, such that for all $A\in D$ and for each $i \in A \setminus \cup$Max(A,D),

$$p_i = \frac{c_A - \sum_{B\in Max(A,D)} c_B}{\sum_{j\in A \setminus \cup Max(A,D)} \sqrt{a_j}} \sqrt{a_i}$$

As we will see this means that the constrained minimum is obtained by hierarchically satisfying a complete subcollection of the constraints with equality.

Example: The minimum of $1/x_1 + 1/x_2$ subject to $1/2 \le x_1 \le 1$, $1 \le x_2 \le 4$ and $x_1+x_2 \le 4$ occurs at (1,3). Here D={{1},{1,2}}.

Example: The minimum of $1/x_1 + 1/x_2 + 1/x_3$ subject to $x_3 \le 5$, $x_1+x_2 \le 2$, and $x_1+x_2+x_3 \le 6$ occurs at

(1,1,4). Here D = {{1,2}, {1,2,3}}.

Theorem 1 solves the Transformed Problem. The (original) Problem is solved by the point $q_i := a_i / p_i$, for $1 \le i \le N$. We note that since the bounds on redundant constraints can be made unachievable without changing the polytope, we may assume that the D in Theorem 1 consists of irredundant constraints.

Example 1: Five of these constraints are redundant. There are 27 union-free complete subcollections D of irredundant constraints, associated to each of which is a candidate solution pD from the conclusion of Theorem 1. For example, D={{1},{4},{1,2},{1,2,3,4}} is a complete union-free subcollection with associated point pD=(10, 81.44, 30.25, 40). This point is feasible, i.e. lies in the constraint region, and has objective function value 69.3984. Among these 27 points, the feasible point with minimal objective function is the point (10.4225, 80, 31.2675, 40), with function value 68.3785. This point arises from the subcollection {{2},{4},{1,2,3,4}}. Corresponding to this solution is the solution to the (original) Problem (9.5946, 20, 28.7839, 10). We would take stratum sample sizes (10, 20, 29, 10) (or play around with different roundings).

We note how Theorem 1 could be implemented and contrast it to Chromy's algorithm. To implement our algorithm, we could write a computer program to remove as many redundant constraints as possible, compute all complete union-free subcollections of irredundant constraints and their corresponding points, and determine the feasible point among these with lowest objective function value. (The algorithm will produce a feasible point if the constraint region is nonempty.) In contrast, Chromy's algorithm takes with a user-specified level of tolerance ε and either doesn't converge or iterates eventually toward a solution within ε. It won't in general produce the exact solution, might not converge, and might take longer or shorter to produce a point within ε tolerance depending on the size of ε and the choice of initial point.

3. Preliminaries

Prop 1 Let $0 \le r < s \le K$ and let $A \in C(s)$. Then there exists a positive integer t and unique $A_1, \ldots, A_t \in C(r)$ such that $A = A_1 \cup \ldots \cup A_t$.

Proof Existence is obvious. Suppose $A_1 \cup \ldots \cup A_t = B_1 \cup \ldots \cup B_u$ for some $B_1, \ldots, B_u \in C(k)$. Since distinct level r sets are pairwise disjoint, for all $1 \le i \le t$, there exists $1 \le j \le r$ such that $A_i = B_j$. □

Cor C(0),…, C(K) are pairwise disjoint.

Cor If $A,B \in C$ and $B \subseteq A$, then $Lev(B) \le Lev(A)$, and B is one of the sets in the decomposition of A into level Lev(B) sets.

Proof Suppose Lev(B) > Lev(A). Decomposing B into distinct level Lev(A) sets $B_1, \ldots, B_t$, since $B \subseteq A$, we see that t=1 and $A=B_1=B$. Contradiction. Now decompose A into distinct level Lev(B) sets $A_1, \ldots, A_t$. Since $B \subseteq A$, we see that t=1 and $B=A_i$ for some $1 \le i \le t$. □

Prop 2 $C(0) = \{A \in C: Max(A,C) = \varnothing\}$ and for $0 < k \le K$, $C(k) = \{A \in C \setminus (C(0) \cup \ldots \cup C(k-1)) : Max(A, C \setminus (C(0) \cup \ldots \cup C(k-1)) = \varnothing\}$.

Proof Let $D(k) := \{A \in C \setminus (C(0) \cup \ldots \cup C(k-1)): Max(A, C \setminus (C(0) \cup \ldots \cup C(k-1)) = \varnothing\}$ for $0 \le k \le K$. We are to show C(k)=D(k) for all k. Note that $C(k) \subseteq D(k)$ follows from the corollaries. Fix $0 \le k \le K$.

Let $A \in D(k)$. Suppose $A \in C(j)$ for some j > k. Decomposing A into distinct level k sets $A_1, \ldots, A_t$, since

Max(A, C \ (C(0) $\cup$...$\cup$ C(k-1)) = $\emptyset$, we see that t=1 and A=$A_1 \in$C(k). ☐

Defn A collection of sets D is *weakly nested* if for all A,B$\in$D, A$\cap$B=$\emptyset$, B$\subseteq$A, or A$\subseteq$B.

That is, every pair of members of D is disjoint or satisfies a containment relation. Every subcollection of a nested collection is weakly nested. In particular, the members of Max(A,D) for A in a complete subcollection D of C are pairwise disjoint. However, for example, {{1},{1,2}} is weakly nested but not nested. Note that in the next proposition, some of the members A \ $\cup$Max(A,D) of this partition might be empty, e.g. for D={{1},{2},{1,2}}.

Prop 3 Let D be a complete subcollection of C$\cup$V. Then the sets A \ $\cup$Max(A,D) for A$\in$D partition {1,...,N}.

Proof Let $1 \le i \le N$. Let A$\in$D be minimal with respect to containment such that A contains i. Then i $\in$ A \ $\cup$Max(A,D). Suppose i $\in$ T \ $\cup$Max(T,D) for some T $\in$ D. Then A and T meet, but we can't have A $\subset$ T or T $\subset$ A, so A=T.☐

We next describe in detail the solution to one equality constraint.

Prop 4 Let c be a nonzero real number. The unique minimum of $f(x_1,...,x_N) := \sum_{i=1}^N a_i/x_i$ subject to $\sum_{i=1}^N x_i = c$ and $x_1>0,\ldots, x_N>0$ occurs at the point $p := \frac{c}{\sqrt{a_1}+...+\sqrt{a_N}}\left(\sqrt{a_1},...,\sqrt{a_N}\right)$. This is the point of the first octant that satisfies the equations $\frac{a_1}{x_1^2} = ... = \frac{a_N}{x_N^2}$ and $\sum x_i = c$. Furthermore $f(p) = \frac{(\sqrt{a_1}+...+\sqrt{a_N})^2}{c}$.

Proof Straightforward. ☐

For the next proposition, we adopt the standard convention that the empty sum is zero.

Prop 5 Let D be a complete subcollection of C$\cup$V, and let $c_{\{i\}} \in \{c_i, C_i\}$ for all singletons {i}$\in$D. Then

$$EQ(D) = \left\{\sum_{i\in A\setminus\cup Max(A,D)} x_i = c_A - \sum_{B\in Max(A,D)} c_B : A\in D\right\}.$$

In particular, EQ(D)=$\emptyset$ if and only if there exists A$\in$D with A \ $\cup$Max(A,D)= $\emptyset$ and $c_A \neq \sum_{B\in Max(A,D)} c_B$.

That is, equality constraints on a complete subcollection can be rewritten as equality constraints in disjoint variables. E.g. $\{x_1=c_1, x_2=C_2, x_1+x_2+x_3=c_{\{1,2,3\}}\}=\{x_1=c_1, x_2=C_2, x_3=c_{\{1,2,3\}}-c_1-C_2\}$. Note that we need to include the empty sums in Proposition 5. For example, let N:=2, C:={{1,2}}, and D:= C $\cup$V. If $c_{\{1,2\}} = c_{\{1\}}+c_{\{2\}}$ then the set in Proposition 5 consists of the point ($c_{\{1\}}$, $c_{\{2\}}$), while if $c_{\{1,2\}} \neq c_{\{1\}}+c_{\{2\}}$, then the set in Proposition 5 is the empty set. The proposition would have failed if we had excluded the empty sum in the latter case.

Proof We first note that for all q$\in R^N$ and A$\in$D, $\left\{\sum_{i\in\cup Max(A,D)} q_i = \sum_{B\in Max(A,D)} \sum_{i\in B} q_i\right\}$.
since the members of Max(A,D) are pairwise disjoint. Let q$\in R^N$ and let A$\in$D. Suppose $\sum_{i\in T} q_i = c_T$ for all T$\in$D. Then $\sum_{i\in A\setminus\cup Max(A,D)} q_i = \sum_{i\in A} q_i - \sum_{B\in Max(A,D)} \sum_{i\in B} q_i = c_A - \sum_{B\in Max(A,D)} c_B$.
Suppose, on the other hand, $\sum_{i\in T\setminus\cup Max(T,D)} q_i = c_T - \sum_{B\in Max(T,D)} c_B$ for all T$\in$D. Then

$\sum_{i\in A} q_i = \sum_{i\in A\setminus\cup Max(A,D)} q_i + \sum_{B\in Max(A,D)} \sum_{i\in B} q_i$. By induction on Lev(A), $\sum_{i\in B} q_i = c_B$ for all $B\in$ Max(A,D). So $\sum_{i\in A} q_i = c_A - \sum_{B\in Max(A,D)} c_B + \sum_{B\in Max(A,D)} c_B = c_A$.□

<u>Notation</u>: For a complete subcollection D of C∪V and constants $c_{\{i\}}\in\{c_i, C_i\}$ for all singletons $\{i\}\in$ D, let pD denote the point defined by D in Theorem 1. That is

$$pD_i = \frac{c_A - \sum_{B\in Max(A,D)} c_B}{\sum_{j\in A\setminus\cup Max(A,D)} \sqrt{a_j}} \sqrt{a_i} \quad \text{for all } i \in A \setminus \cup Max(A,D) \text{ with } A\in D.$$

Note that some coordinates of pD might be zero or negative.

<u>Cor</u> Let D be a complete subcollection D of C∪V such that EQ(D)≠∅, and let $c_{\{i\}}\in\{c_i, C_i\}$ for all singletons $\{i\}\in$D. Suppose that all coordinates of pD are nonzero. Then pD minimizes $\Sigma\, a_i/x_i$ on EQ(D).

So pD is on the boundary of CR(C) and, if its coordinates are nonzero, is determined by hierarchically solving a subset of the constraints with equality and minimizing the objective function on each of the corresponding equality constraints in disjoint variables.

<u>Proof</u> By Proposition 5, $\left\{\sum_{i\in A} x_i = c_A : A\in D\right\} = \left\{\sum_{i\in A\setminus\cup Max(A,D)} x_i = c_A - \sum_{B\in Max(A,D)} c_B : A\in D\right\}$
Since the latter constraints are in disjoint variables, the result follows from Proposition 4. □

<u>Defn</u> Let D be a subcollection of C∪V. We define the *core* of D to be the set Core(D):={A∈D: A≠∪Max(A,D)}.

Core(D) is the union-free subcollection we get by throwing out unions. For D:={{1},{2},{1,2}}, Core(D) = {{1},{2}}, while {{1,2}} is another union-free subcollection. If D is complete, so is its core, and D and Core(D) contain the same singletons. Core(D) defines the same equality constraints as D, while an arbitrary union-free subcollection needn't.

<u>Prop 6</u> Let D be a complete subcollection of C∪V such that EQ(D)≠∅, and let $c_{\{i\}}\in\{c_i, C_i\}$ for all singletons $\{i\}\in$D. Then EQ(D)= EQ(Core(D)).

That is, nontrivial equality constraints on a complete subcollection can be rewritten as equality constraints on their union-free cores. For example, D:={{1},{2},{1,2}} and E:={{1},{2}} define the same equality constraints if those defined by D have a solution (but they don't if those defined by D have no solution).

<u>Proof</u> For k ≥ 0, define CD(k):= Core(D) ∪{A∈D: Lev(A) ≤ k}. We prove by induction on k that EQ(Core(D)) = EQ(CD(k)). For k=0, CD(0) = Core(D). Suppose the claim is true for some k ≥ 0. Let A ∈ CD(k+1) \ CD(k), and let p∈EQ(Core(D)). Then A = ∪Max(A,D), and, since EQ(D)≠∅, $c_A = \sum_{B\in Max(A,D)} c_B$. By induction, p∈EQ(E(k)) , so $\sum_{i\in B} p_i = c_B$ for all B∈Max(A,D), so $\sum_{i\in A} p_i = \sum_{B\in Max(A,D)} c_B = c_A$. □

<u>Cor</u> Let D be a complete subcollection of C∪V such that EQ(D)≠∅, let $c_{\{i\}}\in\{c_i, C_i\}$ for all singletons $\{i\}\in$D, and let E:=Core(D). Then pD = pE.

<u>Proof</u> By Proposition 6, EQ(D) = EQ(Core(D)). By the Corollary to Proposition 5, pD is the minimum of f on

EQ(D) and pE is the minimum of f on EQ(Core(D)), so we are done since the constrained minima are unique. □

Next, we specify an algorithm to generate all complete union-free subcollections D of $C \cup V$. To generate those with, say k singletons, start with a collection D of k singletons. If D is not complete (i.e. if $k < N$), add to D a nonsingleton member A of C that contains an integer in $\{1,\ldots,N\} \setminus \cup D$, i.e. that contains some "new" integer. If the resulting collection D is not complete, add to D a member B of C that contains a member of the new $\{1,\ldots,N\} \setminus \cup D$. Continue until the resulting collection D is complete.

Such a collection D is union-free since each set we added to D contains a new integer. Now fix an arbitrary union-free complete subcollection D, with, say, k singletons and n nonsingleton sets. Let $A_1,\ldots, A_n$ be its nonsingleton sets, listed in order of nondecreasing level. The D is the result of the algorithm that starts with the k singletons and adds the sets $A_1,\ldots, A_n$ in order. (Suppose that for some k, $A_{k+1} \subseteq S \cup A_1 \cup \ldots \cup A_k$, where S is the union of the singletons in D. Then by weak nesting and our ordering of the A_i's, each of $A_1,\ldots, A_k$ that meets A_{k+1} is contained in A_{k+1}. So A_{k+1} is the union of the singletons from $S \cap A_{k+1}$ and the $A_1,\ldots, A_k$ that meet A_{k+1}.) The next proposition improves on the earlier assertion that the constrained minimum is on the boundary of the constraint region.

Prop 7 Let p minimize $\Sigma\, a_i / x_i$ on CR(C). There are a complete subcollection D of $C \cup V$ and constants $c_{\{i\}} := C_i$ for all singletons $\{i\} \in D$, such that $p \in EQ(D)$.

Proof Suppose not. Let $1 \le i \le N$ be such that $c_i \le p_i < C_i$ and $\sum_{i \in A} p_i < c_A$ for all $A \in C$ such that A contains i. We may assume i=1. Since all coordinates of the gradient of f are negative at p, the directional derivative of f in the unit direction u:=(1,0,...,0) is negative at p. Then $f(p+tu) < f(p)$ for all sufficiently small $t > 0$, and we may choose t small enough so that $p+tu \in CR(C)$. Contradiction. □

Example For $f(x_1, x_2) := 1/x_1 + 1/x_2$ on $CR(C) := \{x_1+x_2 \le 2,\ 3/2 \le x_1 \le 2,\ ½ \le x_2 \le 2\} = \{(3/2,1/2)\}$, p=(3/2,1/2) $\in$ EQ(D) for D={{1,2}}. However, $p \ne pD$ for this D, but rather for D={{1},{2}}, {{1},{1,2}}, or {{2},{1,2}} with $c_{\{1\}} := 3/2$ and $c_{\{2\}} := 1/2$.

Prop 8 Let $J \subseteq V$, and $I := \cup J$. Let D be a complete subcollection of $C \cup V$. Suppose that the nonempty sets $A \setminus I$ for $A \in D$ are distinct. Let $F := \{A \setminus I : A \in D, A \setminus I \ne \emptyset\}$, and let $A \in D$ be such that $A \setminus I \ne \emptyset$. Then

(1) $\{B \in Max(A,D) : B \setminus I \ne \emptyset\} = \{B \in D : B \setminus I \in Max(A \setminus I, F)\}$.

Suppose also that $\{B \in D : B \setminus I = \emptyset\} = \{\{i\} : i \in I\}$. Then

(2) $\{B \in Max(A,D) : B \setminus I = \emptyset\} = \{\{i\} : i \in (A \cap I) \setminus \cup\{B \in Max(A,D) : B \setminus I \ne \emptyset\}\}$, and

(3) $A \setminus \cup Max(A,D) = (A \setminus I) \setminus \cup Max(A \setminus I, F)$.

Proof Note that for all $T \in F$, there is a unique $B \in D$ such that $T = B \setminus I$. Note also that by weak nesting, for all $B \in D$ such that $B \setminus I \ne \emptyset$, $A \supseteq B$ if and only if $A \setminus I \supseteq B \setminus I$. (For the direction $\Leftarrow$, note that if $A \subseteq B$, then $A \setminus I = B \setminus I$ implies A=B by hypothesis.) Part (1) now follows easily.

(2): Let $B \in Max(A,D)$ with $B \setminus I = \emptyset$. By hypothesis, $B = \{i\}$ for some $i \in A \cap I$. Suppose $i \in T$ for some $T \in Max(A,D)$ with $T \setminus I \ne \emptyset$. Then $\{i\} \subseteq T$, so by maximality $\{i\} = T$, contradicting $T \setminus I \ne \emptyset$.

Now let $i \in (A \cap I) \setminus \cup\{B \in Max(A,D) : B \setminus I \ne \emptyset\}$ and write $\{i\} \subseteq T \subseteq A$ for some $T \in Max(A,D)$. Then $T \setminus I = \emptyset$, so $|T| = 1$, so $\{i\} = T \in Max(A,D)$.

(3): By (1) and (2),

$\cup Max(A,D) = (\cup\{B \in Max(A,D) : B \setminus I \ne \emptyset\}) \cup ((A \cap I) \setminus \cup\{B \in Max(A,D) : B \setminus I \ne \emptyset\})$

$= (\cup\{B\in \mathrm{Max}(A,D): B\backslash I \neq \emptyset\}) \cup (A\cap I) = (\cup\{B\in D: B\backslash I\in \mathrm{Max}(A\backslash I,F)\}) \cup (A\cap I)$
$= (\cup\{B\backslash I: B\in D, B\backslash I\in \mathrm{Max}(A\backslash I,F)\}) \cup (A\cap I) = (\cup \mathrm{Max}(A\backslash I,F)) \cup (A\cap I)$.□

4. Proof of the Algorithm

We now prove the main theorem, which solves the Transformed (and the original) Problem for nested constraints. By the Corollary to Proposition 5, Theorem 1 says that the constrained minimum is determined by hierarchically solving a subset of the constraints with equality and minimizing the objective function on each of the corresponding equality constraints in disjoint variables.

Theorem 1 Let $C \subseteq \{A \subseteq \{1,..., N\}: |A| > 1\}$ be nested, and suppose

$$CR(C) := \left\{\sum_{i\in A} x_i \le c_A : A \in C\right\} \cap \left\{c_i \le x_i \le C_i : 1 \le i \le N\right\}$$

is nonempty. Let $p=(p_1,\dots,p_N)\in R^N$ be the unique minimum of $f(x_1,...,x_N) := \sum_{i=1}^{N} a_i / x_i$ on $CR(C)$. Then there are a union-free complete subcollection $D \subseteq C\cup V$ and constants $c_{\{i\}}\in\{c_i ,C_i\}$ for all singletons $\{i\}\in D$, such that for all $A\in D$ and for each $i \in A \setminus \cup \mathrm{Max}(A,D)$,

$$p_i = \frac{c_A - \sum_{B\in \mathrm{Max}(A,D)} c_B}{\sum_{j\in A\backslash\cup \mathrm{Max}(A,D)} \sqrt{a_j}} \sqrt{a_i}$$

Proof We prove this by induction on N. The case N=1 is obvious.

Suppose p satisfies some of the variable constraints with equality. Let $J:=\{\{i\}: p_i\in\{c_i, C_i\}\}$, and let $I:= \cup J$. Define $E:=\{A\backslash I: A\in C, A\backslash I\neq\emptyset\}$,

$$d_B := \min\left\{c_A - \sum_{i\in A\cap I} p_i : B = A \setminus I, A \in C\right\} \text{ for all } B\in E$$

and $q\in R^{N-|I|}$ by $q_i := p_i$ for $i\notin I$. Then E is nested and q minimizes $\sum_{i\notin I} a_i / x_i$ on the constraint region $CR(E) := \left\{\sum_{i\in B} x_i \le d_B : B\in E\right\} \cap \left\{c_i \le x_i \le C_i : i \notin I\right\}$. By induction, there is a complete subcollection F of $E \cup\{\{i\}: i \notin I\}$ and constants $d_{\{j\}}\in\{c_j ,C_j\}$ for all $j\notin I$ such that for all $A\in F$, and for each $j \in A \setminus \cup \mathrm{Max}(A,F)$,

$$p_j = \frac{d_A - \sum_{B\in \mathrm{Max}(A,F)} d_B}{\sum_{k\in A\backslash\cup \mathrm{Max}(A,F)} \sqrt{a_k}} \sqrt{a_j}$$

Let $D:=\{A\in C: A\backslash I\in F,\ d_{A\backslash I} = c_A - \sum_{i\in A\cap I} p_i\ \} \cup \{\{i\}: i\in I\}$, and set $c_{\{i\}}:= p_i \in \{c_i, C_i\}$ for $i\in I$. Throw out from D any sets A with duplicate $A\backslash I$ (which would require duplicate $c_A - \sum_{i\in A\cap I} p_i$) so that for each $B\in F$ there is a unique $A\in D$ such that $B=A\backslash I$. D is complete since F is. We claim that p=pD. Note that by our choice of $c_{\{i\}}$'s, $p_i = pD_i$ for all $i\in I$.

Fix $j\notin I$, and let $A\in D$ be such that $j \in A \setminus \cup \mathrm{Max}(A,D)$. Then $A\backslash I$ is nonempty and so by Proposition 8, $(A\backslash I) \setminus \cup \mathrm{Max}(A\backslash I, F) = A \setminus \cup \mathrm{Max}(A,D)$. So

$$\sum_{\in \quad \in} \qquad \sum_{\in \ \cap} \in \iota A - \sum_{\in} \qquad \neq \qquad - \sum_{\in \ \cap}$$

So we may assume that for all $1 \le i \le N$, $c_i<p_i<C_i$. Let D be a complete subcollection of C such that $p \in EQ(D)$ and suppose $p_i \neq pD_i$ for some $i \in A \setminus \cup Max(A,D)$ with $A \in D$. By Proposition 5, $| A \setminus \cup Max(A,D) | > 1$. By propositions 4 and 5, there exist $i,j \in A$ such that $a_i/p_i^2 > a_j/p_j^2$. The directional derivative of f in the direction v defined by

$$v_k := \begin{cases} 1, & \text{if } k = i \\ -1, & \text{if } k = j \\ 0, & \text{otherwise} \end{cases}$$

is negative at p, so for all sufficiently small t>0, f(p+tv) < f(p), and so p+tv is not in the constraint region. So $p_i=C_i$ or $p_j=c_j$. Contradiction. Finally, we may assume D is union-free by replacing it with its core and applying the corollary to Proposition 6.□

Remark on the proof: From Proposition 7, we know that the constrained minimum p is in EQ(D) for some complete subcollection D. If we could prove more easily that pD was in CR(C), we could use an easier argument in the proof of Theorem 1. However, as we know from Example 1, many D produce unfeasible points pD.

Conjecture Suppose that for all complete union-free subcollections D of $C \cup V$ that contain no singletons, $pD_1 < c_1$. Let E be a complete subcollection for which pE is the constrained minimum in Theorem 1. Then E contains the singleton {1}.

References

[B] Bethel, J. (1989), "Sample Allocation in Multivariate Surveys", *Survey Methodology*, **15**, pp. 45-57.

[C] Chromy, J. (1987), "Design Optimization with Multiple Objectives", *Proceedings of the Survey Research Methods Section*, American Statistical Association.

[K1] Kinyon, D., Glassbrenner, D., Black, J., and Detlefsen, R. (2000), "Designing Retail, Wholesale, and Services Samples used for Surveys Conducted by the U.S. Census Bureau", *Proceedings of the International Conference on Establishment Surveys –II*.

[K2] Kinyon, D., Glassbrenner, D., Black, J., and Detlefsen, R. (2000), "Improving the Design of Samples Used for Business Surveys Conducted by the United States Bureau of the Census", *Joint Statistical Meetings of the American Statistical Association.*