RESEARCH REPORT SERIES
(Statistics #2019-03)

Direct Proof of Exact Sample Allocation Optimality with Cost Constraints

Tommy Wright

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: March 6, 2019 (Revised April 1, 2019)

Direct Proof of Exact Sample Allocation Optimality with Cost Constraints

Tommy Wright
U.S. Bureau of the Census

Abstract

We provide an elementary derivation of Kadane's dynamic sampling plan by first directly finding the sample allocation that minimizes a decomposed weighted objective function. We then prove that the sample allocation also minimizes the sampling variance.

The plan is most appropriate in the context of sampling sequentially from a stratified population where sampling costs vary among the strata. It specifies from which stratum to take the next sample unit which reduces variance by the largest amount per unit cost. Whenever sampling stops, the realized allocation minimizes the sampling variance for the cost C^* at that point, as well as for any cost and allocation that costs less than C^*. It's a form of adaptive sampling, and our proof provides complete insight into why Kadane's plan works.

KEY WORDS: Exact optimal allocation; Fixed budget; Stratification.

1. INTRODUCTION

Assume a finite population of N units is stratified into H disjoint strata where N_h is the known number of units in stratum h, for $h = 1, ..., H$. Note that $N = N_1 + \cdots + N_H$.

General Setup: Let Y_{hj} be the fixed unknown value of interest for the j^{th} unit in stratum h: $\bar{Y}_h = \left(\sum_{j=1}^{N_h} Y_{hj} \right) / N_h$, and $S_h^2 = \left(\sum_{j=1}^{N_h} (Y_{hj} - \bar{Y}_h)^2 \right) / (N_h - 1)$. See the following visual.

Stratum 1	Stratum 2	$\cdots$	Stratum h	$\cdots$	Stratum H
$N_1, \bar{Y}_1, S_1^2$	$N_2, \bar{Y}_2, S_2^2$		$N_h, \bar{Y}_h, S_h^2$		$N_H, \bar{Y}_H, S_H^2$

The unknown population total T_Y is

$$T_Y = \sum_{h=1}^{H} \sum_{j=1}^{N_h} Y_{hj} = \sum_{h=1}^{H} N_h \bar{Y}_h. \tag{1}$$

To estimate T_Y under the classical design-based approach, select a stratified random sample of n units where n_h sample units provide the sample mean $\bar{y}_h$ for stratum h. Note that $n = n_1 + \cdots + n_H$ and that $n_h \geq 1$ for all h. A natural unbiased estimator for $T_Y = \sum_{h=1}^{H} N_h \bar{Y}_h$ is

$$\hat{T}_Y = \sum_{h=1}^{H} N_h \bar{y}_h \tag{2}$$

with sampling variance (*unweighted sum of stratum sampling variances*)

$$Var(\hat{T}_Y) = \sum_{h-1}^{H} Var(N_h \bar{y}_h) = \sum_{h=1}^{H} N_h^2 \left(\frac{N_h - n_h}{N_h} \right) \frac{S_h^2}{n_h}. \tag{3}$$

For a fixed cost (budget) structure,

$$C = \sum_{h=1}^{H} c_h n_h \tag{4}$$

where C is the overall fixed budget, and c_h is the measurement cost per sample unit from stratum h, we find the optimal n and the optimal sample allocation $(n_1, ..., n_H)$ of n that will minimize $Var(\hat{T}_Y)$, where $n = \sum_{h=1}^{H} n_h$.

2. OPTIMAL SAMPLE ALLOCATION WITH FIXED BUDGET

Using Lagrange multiplier to minimize $Var(\hat{T}_Y)$ in (3) subject to the constraint in (4), we can show that the optimal n is

$$n = (C)\frac{\sum_{i=1}^{H}\frac{N_i S_i}{\sqrt{c_i}}}{\sum_{h=1}^{H} N_h S_h \sqrt{c_h}} \tag{5}$$

and the optimal allocation $(n_1, ..., n_H)$ of n is

$$n_h = (C)\frac{\frac{N_h S_h}{\sqrt{c_h}}}{\sum_{h=1}^{H} N_h S_h \sqrt{c_h}} \tag{6}$$

for $h = 1, ..., H$. The known result in (6) is appealing because it is consistent with reasonable criteria. That is, n_h is directly proportional to N_h and S_h, while it is inversely proportional to $\sqrt{c_h}$. The result in (6) is also a generalization of the well-known Neyman allocation (Tschuprow, 1923; Neyman, 1934) where all c_h are equal. Generally, each n_h in (6) needs to be rounded to an integer. With this rounding, optimality is placed in doubt. In Section 4, we present an exact optimal allocation algorithm that never requires rounding.

3. KADANE'S OPTIMAL DYNAMIC SAMPLING PLAN

Kadane (2005) defines a dynamic sampling plan among strata as a "...permutation of sampled items specifying which stratum is to receive the next item to be included in the sample. An optimal plan has the property of achieving minimum variance for its cost, whenever it is truncated". Where the allocation to stratum h is increased from $m_h - 1$ to m_h, he presents the following sampling plan and shows that it is an optimal plan: "...start with the allocation of one sampled item to each stratum, then order allocations by $\frac{N_h^2 S_h^2}{(m_h - 1)m_h c_h}$, highest first (breaking ties arbitrarily)...(sampling eventually stops)." However, the origin of his sampling plan can be clarified. An elementary derivation follows: introduce an objective function (7); decompose it (8) to immediately reveal Kadane's plan; and prove its optimality in Section 6.

4. EXACT OPTIMAL ALLOCATION Algorithm V (FIXED BUDGET)

It is required that $n_h \geq 1$ for $h = 1, ..., H$. As a result, C, the overall given available budget, must be at least $\sum_{h=1}^{H} c_h$. That is, C must permit at least one sample unit from each stratum. Our overall cost constraint remains the fixed budget as given in (4).

Without loss of generality, assume that $\frac{N_1 S_1}{\sqrt{c_1}} \geq \cdots \geq \frac{N_H S_H}{\sqrt{c_H}}$. For mathematical convenience (see Remark 1), the desire is to determine the allocation $(n_1, ..., n_H)$ that minimizes the *weighted sum of stratum sampling variances*

$$Var_W(\hat{T}_Y) = \sum_{h=1}^{H} Var(N_h \bar{y}_h)\frac{1}{c_h} = \sum_{h=1}^{H} N_h^2 \Big(\frac{N_h - n_h}{N_h}\Big)\frac{S_h^2}{n_h}\frac{1}{c_h} \tag{7}$$

subject to the cost constraint (4). Note that $Var(N_h\bar{y}_h)\frac{1}{c_h}$ is the sampling variance of $N_h\bar{y}_h$ per unit cost for stratum h. Determining the allocation to minimize $Var_W(\hat{T}_Y)$ subject to (4), we note that $Var_W(\hat{T}_Y)$ can be decomposed in a way similar to $Var(\hat{T}_Y)$ in Wright (2016, 2017). This decomposition in (8) is key to the exact optimal allocation *Algorithm V*, where $\sum_{h=1}^{H} N_h(N_h-1)S_h^2\frac{1}{c_h}$ is the (cost) weighted sum of stratum sampling variances $Var_W(\hat{T}_Y)$ when $n_h = 1$ for each h.

$$
\begin{aligned}
Var_W(\hat{T}_Y) \quad = \quad & \sum_{h=1}^{H} N_h(N_h-1)S_h^2\frac{1}{c_h} \\
& -\frac{N_1^2S_1^2/c_1}{1\cdot 2} - \frac{N_1^2S_1^2/c_1}{2\cdot 3} - \frac{N_1^2S_1^2/c_1}{3\cdot 4} - \cdots - \frac{N_1^2S_1^2/c_1}{(n_1-1)(n_1)} \\
& \quad \vdots \\
& -\frac{N_h^2S_h^2/c_h}{1\cdot 2} - \frac{N_h^2S_h^2/c_h}{2\cdot 3} - \frac{N_h^2S_h^2/c_h}{3\cdot 4} - \cdots - \frac{N_h^2S_h^2/c_h}{(n_h-1)(n_h)} \\
& \quad \vdots \\
& -\frac{N_H^2S_H^2/c_H}{1\cdot 2} - \frac{N_H^2S_H^2/c_H}{2\cdot 3} - \frac{N_H^2S_H^2/c_H}{3\cdot 4} - \cdots - \frac{N_H^2S_H^2/c_H}{(n_H-1)(n_H)}
\end{aligned} \tag{8}
$$

Clearly, (8) will be minimized, (4) will be satisfied, and each stratum will have at least one sample unit if we use the following algorithm, clarified by Remark 1.

EXACT OPTIMAL ALLOCATION Algorithm V (Fixed Budget C)

Step 1: First, note 1 unit is to be selected for the sample from each stratum.
Step 2: For additional sample units, compute the array of *priority values*:

$$
\begin{array}{lcccc}
\text{Stratum } 1 & \frac{(N_1S_1/\sqrt{c_1})}{\sqrt{1\cdot 2}} & \frac{(N_1S_1/\sqrt{c_1})}{\sqrt{2\cdot 3}} & \frac{(N_1S_1/\sqrt{c_1})}{\sqrt{3\cdot 4}} & \cdots \\
 & & \vdots & & \\
\text{Stratum } h & \frac{(N_hS_h/\sqrt{c_h})}{\sqrt{1\cdot 2}} & \frac{(N_hS_h/\sqrt{c_h})}{\sqrt{2\cdot 3}} & \frac{(N_hS_h/\sqrt{c_h})}{\sqrt{3\cdot 4}} & \cdots \\
 & & \vdots & & \\
\text{Stratum } H & \frac{(N_HS_H/\sqrt{c_H})}{\sqrt{1\cdot 2}} & \frac{(N_HS_H/\sqrt{c_H})}{\sqrt{2\cdot 3}} & \frac{(N_HS_H/\sqrt{c_H})}{\sqrt{3\cdot 4}} & \cdots
\end{array} \tag{9}
$$

Step 3: Select the largest priority value from the array that has not already been picked; associate it with its stratum; and increase that stratum sample size by 1 if the new cost including this 1 new sample unit is $\sum_{h=1}^{H} c_hn_h \le C$. Otherwise (i.e., $\sum_{h=1}^{H} c_hn_h > C$), stop without increasing the sample size by 1 for the associated stratum; cost of sample when we stop is $C^* = \sum_{h=1}^{H} c_hn_h$.
Step 4: Go to Step 3.

Remark 1: As noted, $\sum_{h=1}^{H} N_h(N_h-1)S_h^2\frac{1}{c_h}$ in (8) is the value of $Var_W(\hat{T}_Y)$ when $n_h = 1$ for all h. In fact, this sum is the largest possible value for $Var_W(\hat{T}_Y)$. Each of the other rows of (8) corresponds to one of the H strata and shows how $Var_W(\hat{T}_Y)$ decreases with additional increases in overall sample size. In particular, when the sample size for the h^{th} stratum is *"increased"* from $m_h - 1$ to m_h, the associated $Var(N_h \bar{y}_h)\dfrac{1}{c_h}$ for the h^{th} stratum *"decreases"* by

$$\frac{Var(N_h \bar{y}_{m_h-1})}{c_h} - \frac{Var(N_h \bar{y}_{m_h})}{c_h} = \frac{(N_h^2 S_h^2 / c_h)}{(m_h - 1)(m_h)} = \left(\frac{(N_h S_h / \sqrt{c_h})}{\sqrt{(m_h - 1)(m_h)}} \right)^2, \qquad (10)$$

where $\bar{y}_{m_h-1}$ ($\bar{y}_{m_h}$) is a sample mean based on $m_h - 1$ (m_h) sample units from stratum h. The result in (10) is also the amount by which $Var_W(\hat{T}_Y)$ *"decreases"* (see (8)). The quantity in (10) is the reduction in $Var_W(\hat{T}_Y)$ per unit cost by picking a unit for the sample from stratum h bringing that stratum's sample size to n_h. At each step in *Algorithm V*, we pick a unit from among the strata which reduces $Var_W(\hat{T}_Y)$ by the largest amount per unit cost at that point in the sequence of determining the final sample sizes $n_1, ..., n_H$, which is very reasonable and desirable.

The objective function in (7), the decomposition in (8), and the cost constraint in (4) provide a complete mathematical framework and insight that lead directly to *Algorithm V.*

5. EXAMPLES

Example of Optimal Allocation Using Lagrange Multiplier: Assume a stratified population of $N = 149$ units with parameters:

h	N_h	S_h	c_h	$\frac{N_h S_h}{\sqrt{c_h}}$
1	61	6	\$4	183
2	41	4	\$1	164
3	47	10	\$9	$\frac{470}{3}$

From (5) with fixed budget $C = \$55$, optimal overall sample size is:

$$n = (C)\frac{\sum_{i=1}^{H} \frac{N_i S_i}{\sqrt{c_i}}}{\sum_{h=1}^{H} N_h S_h \sqrt{c_h}} = 12.0128 \approx 12;$$

optimal sample allocation from (6) is $(n_1, n_2, n_3) = (4.3647, 3.9115, 3.7366)$; associated sampling variance is $Var(\hat{T}_Y) = 89,132.8183$; and the cost is $(4)(4.3647)+(1)(3.9115)+(9)(3.7366) = 54.9997 \approx \$55$. So the optimal allocation uses the entire budget \$55.

The optimal sample allocation just obtained does not give integers, and we must round. Rounding to nearest integers $(n_1, n_2, n_3) = (4, 4, 4)$; the associated sampling variance is $Var(\hat{T}_Y) = 87,886$; but the cost $(4)(4)+(1)(4)+(9)(4) =$\$56 exceeds \$55, the fixed budget.

Example of Algorithm V: We assume the same population of $N = 149$ and parameters as given above and begin with Array 0 and obtain Array I.

h	c_h	$\frac{N_h S_h}{\sqrt{c_h}}$	$\frac{1}{\sqrt{1\cdot 2}}$	$\frac{1}{\sqrt{2\cdot 3}}$	$\frac{1}{\sqrt{3\cdot 4}}$	$\frac{1}{\sqrt{4\cdot 5}}$	$\frac{1}{\sqrt{5\cdot 6}}$	$\frac{1}{\sqrt{6\cdot 7}}$	$\frac{1}{\sqrt{7\cdot 8}}$	$\cdots$
1	\$4	183	129.40	74.71	52.83	40.92	33.41	28.24	24.45	$\cdots$
2	\$1	164	115.97	66.95	47.34	36.67	29.94	25.31	21.92	$\cdots$
3	\$9	$\frac{470}{3}$	110.78	63.96	45.23	35.03	28.60	24.17	20.94	$\cdots$

Array 0: Array of Priority Values for *Algorithm V.*

Proceeding through Array 0 according to *Algorithm V*, the amount of the budget being spent accumulates as shown in parentheses in Array *I* as we sequentially pick the largest priority values (first 129.40; next 115.97; then 110.78;...; and finally 47.34 - all given in bold), resulting in final allocation $(n_1, n_2, n_3) = (4, 4, 3)$. Note that we use only $C^* = \$47$ of the fixed budget $C = \$55$. Along the way applying *Algorithm V*, we note the $Var(\hat{T}_Y)$ for the allocations in brackets [ ] for the allocation up to that point. The double brackets [[ ]] give the weighted sum of stratum sampling variances (7) for the allocation up to that point. Note further that the final allocation gives a sampling variance of the allocation up to that point of $Var(\hat{T}_Y) = 106,294$. Specifically, we describe the first set of entries in Array *I* as follows: (*i*) 129.40 = priority value obtained by $183 \times (\frac{1}{\sqrt{1 \cdot 2}})$; (*ii*) $(2, 1, 1) = (n_1, n_2, n_3)$ gives the allocation for $n = 4$ after the first largest priority value 129.40 (stratum 1) is picked; (*iii*) \$18 = the cost of the sample allocation (2,1,1); (*iv*) $Var(\hat{T}_Y) = \sum_{h=1}^{3} Var(N_h \bar{y}_h) = 307,222$ for the sample allocation (2,1,1); and (*v*) $Var_W(\hat{T}_Y) = \sum_{h=1}^{3} Var(N_h \bar{y}_h) \frac{1}{c_h}$ = 66,458 for the sample allocation (2,1,1).

h	c_h	$\frac{N_h S_h}{\sqrt{c_h}}$	$\frac{1}{\sqrt{1 \cdot 2}}$	$\frac{1}{\sqrt{2 \cdot 3}}$	$\frac{1}{\sqrt{3 \cdot 4}}$	$\frac{1}{\sqrt{4 \cdot 5}}$	n_h
1	\$4	183	**129.40** (2, 1, 1) (\$18) [307, 222] [[66, 458]]	**74.71** (3, 2, 2) (\$32) [160, 998] [[35, 156]]	**52.83** (4, 3, 3) (\$46) [108, 535] [[23, 792]]	40.92	4
2	\$1	164	**115.97** (2, 2, 1) (\$19) [293, 774] [[53, 009]]	**66.95** (3, 3, 2) (\$33) [156, 515] [[30, 673]]	**47.34** (4, 4, 3) (\$47) [106, 294] [[21, 551]]	36.67	4
3	\$9	$\frac{470}{3}$	**110.78** (2, 2, 2) (\$28) [183, 324] [[40, 738]]	**63.96** (3, 3, 3) (\$42) [119, 698] [[26, 583]]	45.23	35.03	3

Array ***I*****:** Application of *Algorithm V*; Final $(n_1, n_2, n_3) = (4, 4, 3)$, where $n = 1$, costs $C^* = \$47$.

Remark 2: By design, the allocation $(n_1, ..., n_H)$ of $n = \sum_{h=1}^{H} n_h$ that is observed when we stop and associated with $C^* = \sum_{h=1}^{H} c_h n_h$ minimizes $Var_W(\hat{T}_Y)$ the *weighted sum of stratum sampling variances* subject to C^* which does not exceed C.

It's worth noting that *Algorithm V* is exactly the optimal dynamic sampling plan given by Kadane (2005). While Kadane's sampling plan is motivated by a practical problem in auditing different types of units where sampling costs vary, the framework from which his plan originates can be clarified. In this paper, we introduce the objective function $Var_W(\hat{T}_Y)$ and derive directly from it an allocation of the sample size n with a fixed budget constraint that minimizes $Var_W(\hat{T}_Y)$. We do this by decomposing $Var_W(\hat{T}_Y)$, and as an immediate result, we see how to sequentially allocate the sample while staying within budget C. When we reach the point where taking one more sample unit from one of the strata would cause us to exceed the budget, we stop and do not take one more unit for the sample. With the allocation when we stop with cost C^* ($\leq C$), we get exactly the dynamic sampling plan of Kadane. We prove this explicitly in our Theorem and Remark 4.

The following Theorem that is a simplified and clarified result is similar to the main result in Kadane (2005). We show that the allocation $(n_1, ..., n_H)$ of $n = \sum_{h=1}^{H} n_h$ that is observed and associated with $C^* = \sum_{h=1}^{H} c_h n_c$ also minimizes $Var(\hat{T}_Y)$ the *unweighted sum of stratum sampling variances* over all other allocations that cost less than or equal to C^*.

Remark 3: Selecting the largest priority values $\dfrac{N_h S_h/\sqrt{c_h}}{\sqrt{(m_h-1)m_h}}$ sequentially in *Algorithm V* is equivalent to picking sequentially the largest squared priority values $\dfrac{N_h^2 S_h^2/c_h}{(m_h-1)m_h}$ (see (9)). Because these largest priority values (also largest squared priority values) are all positive, there is a positive real number k such that

$$\frac{N_h^2 S_h^2/c_h}{(m_h-1)m_h} \geq k$$

for all priority values that are selected to obtain $C^* = \sum_{h=1}^{H} c_h n_h$.

6. THE MAIN RESULT

Theorem: For fixed budget C, let E_1 be the subset of largest priority values in the array of *Algorithm V* such that for some $k \in R^+$

(*i*) $\dfrac{N_h^2 S_h^2/c_h}{(m_h-1)m_h} \geq k$ for each priority value in E_1;

(*ii*) $\dfrac{N_h^2 S_h^2/c_h}{(m_h-1)m_h} < k$ for each priority value in E_1^c, the complement of E_1;

(*iii*) $\sum_{E_1} c_h + \sum_{h=1}^{H} c_h = C^*$ where C^* is as defined in *Algorithm V*.

Let E_2 be another subset of priority values in the array of *Algorithm V* such that $C_{E_2} \leq C^*$, where C_{E_2} is the cost of a sample corresponding to E_2. Keep in mind that C_{E_2} includes the total cost of one unit from each stratum as well as for the additional units included as a result of picking some priority values from the array. Then

$$\sum_{E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} \;\geq\; \sum_{E_2} \frac{N_h^2 S_h^2}{(m_h - 1)m_h}. \tag{11}$$

Proof: It is enough to show

$$\sum_{E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} \geq 0.$$

Note that $E_1 = (E_1 \cap E_2) \cup (E_1 \cap E_2^c)$ and $E_2 = (E_2 \cap E_1) \cup (E_2 \cap E_1^c)$. Now

$$\begin{aligned}
&\sum_{E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} \\
&= \sum_{E_1 \cap E_2} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} + \sum_{E_1 \cap E_2^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2 \cap E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2 \cap E_1^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} \\
&= \sum_{E_1 \cap E_2^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2 \cap E_1^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h}
\end{aligned} \tag{12}$$

By assumption (i), for all priority values in E_1 and hence for all priority values in $E_1 \cap E_2^c$,

$$\sum_{E_1 \cap E_2^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} \geq k \sum_{E_1 \cap E_2^c} c_h. \tag{13}$$

By assumption (ii), for all priority values in E_1^c and hence for all priority values in $E_2 \cap E_1^c$,

$$\sum_{E_2 \cap E_1^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} < k \sum_{E_2 \cap E_1^c} c_h \tag{14}$$

So from (12), (13), and (14), we have

$$\begin{aligned}
\sum_{E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} &= \sum_{E_1 \cap E_2^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} - \sum_{E_2 \cap E_1^c} \frac{N_h^2 S_h^2}{(m_h - 1)m_h} \\
&\geq k\Big(\sum_{E_1 \cap E_2^c} c_h - \sum_{E_2 \cap E_1^c} c_h\Big) \\
&= k\Big(\sum_{E_1 \cap E_2^c} c_h + \sum_{E_1 \cap E_2} c_h - \sum_{E_2 \cap E_1} c_h - \sum_{E_2 \cap E_1^c} c_h\Big) \\
&= k\Big(\sum_{E_1} c_h - \sum_{E_2} c_h\Big) \\
&= k\Big(\Big[\sum_{E_1} c_h + \sum_{h=1}^{H} c_h\Big] - \Big[\sum_{E_2} c_h + \sum_{h=1}^{H} c_h\Big]\Big) \\
&= k(C^* - C_{E_2}) \geq 0
\end{aligned}$$

Thus the Theorem has been shown.

Remark 4: From the decompositions (Wright, 2016, 2017) of $Var(\hat{T}_h)$, and of $Var(\hat{T}_Y)$,

$$Var(\hat{T}_Y) = \sum_{h=1}^{H} Var(\hat{T}_h) = \sum_{h=1}^{H} N_h(N_h - 1)S_h^2 - \sum_{E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h}$$

where E_1 is the set of priority values (and associated allocation $(n_1, ..., n_H)$) leading to $C^* = \sum_{h=1}^{H} c_h n_h$. Thus by showing that

$$\sum_{E_1} \frac{N_h^2 S_h^2}{(m_h - 1)m_h}$$

is maximized in (11) of the Theorem, we have that *Algorithm V* which stops at C^* with allocation $(n_1, ..., n_H)$ not only minimizes the *weighted sum of stratum sampling variances*, but it also minimizes $Var(\hat{T}_Y)$, the *unweighted sum of stratum sampling variances*, for any allocation whose associated cost does not exceed C^*.

Remark 5: Of course, one could use the c_h values to increase the overall sample size n and move C^* closer to C. For example, in Array I of Section 5, one could take 2 additional sample units from stratum 1 and increase the overall cost to $C^* = \$47 + \$8 = \$55$ $(= C)$. The action would also decrease $Var(\hat{T}_Y)$ as well as $Var_W(\hat{T}_Y)$, but we would no longer have the guarantee of optimality for the realized cost provided by the Theorem. Continuing with the example, this new allocation (6,4,3) has $Var(\hat{T}_Y) = 95,131.33$. But the allocation (4,2,4) which costs \$54 has $Var(\hat{T}_Y) = 94,610$, and the allocation (4,3,4) which costs \$55 has $Var(\hat{T}_Y) = 90,127.33$. Hence we can not use the Theorem to guarantee minimum sampling variance stopping with (6,4,3).

Acknowledgments: The views expressed are those of the author and not those of the U.S. Bureau of the Census. The author is grateful to his colleagues (Robert Ashmead, John Eltinge, and Andrew Raim) for their technical reviews and suggestions. This report is a close follow-up and update to Algorithms V-a, b, c, and d of Section 4.3 in Wright (2016).

REFERENCES

- Kadane, J. B. (2005). "Optimal Dynamic Sample Allocation Among Strata", *Journal of Official Statistics, Vol 21, No. 4*, 531-541.
- Neyman, J. (1934). "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection", *Journal of the Royal Statistical Society, Vol. 97*, 558-606.
- Tschuprow, A. A. (1923). "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations", *Metron, Vol 2*, 461-493, 646-683.
- Wright, T. (2012). "The Equivalence of Neyman Optimum Allocation for Sampling & Equal Proportions for Apportioning the U. S. House of Representatives", *The American Statistician, Vol. 66, No. 4*, 217-224.
- Wright, T. (2016). "Two Optimal Exact Sample Allocation Algorithms: Sampling Variance Decomposition Is Key", *Research Report Series (Statistics #2016-03)*, Center for Statistical Research & Methodology, U. S. Bureau of the Census, Washington, D.C.
- Wright, T. (2017). "Exact Optimal Sample Allocation: More Efficient Than Neyman", *Statistics and Probability Letters, 129*, 50-57.