OPTIMUM STRATUM BOUNDARIES FOR SAMPLING FROM A RIGHT TRIANGULAR DISTRIBUTION OR FROM AN EXPONENTIAL DISTRIBUTION

by

A. R. Manson and C. H. Proctor

Institute of Statistics
Mimeograph Series # 1184
Raleigh - July 1978

Abstract

The shape reproducibility (under left truncation) property of the right triangular probability distribution and of the exponential probability distribution generates recursive relationships among the minimum variance stratum boundaries. These recursive relationships simplify the problem of determining the optimum stratum boundaries from that of a multi-dimensional minimization problem to one having an explicit solution. Solutions are obtained for Neyman allocation, proportional allocation, and equal allocation of sampling effort for stratified random sampling from these two distributions.

KEY WORDS: Stratified random sampling; Right triangular probability distribution; Exponential distribution; Neyman allocation; Proportional allocation; Equal allocation.

* A. R. Manson and C. H. Proctor are with the Department of Statistics, North Carolina State University, Raleigh, N.C. 27650.
The authors gratefully acknowledge the work of Dr. James R. Harvey for his discovery of an iterative, graphical solution for the right triangular distribution, proportional allocation case and of Mr. Ismail Erdum for determining the optimum value of $X_{2,1} = 3 - \sqrt{7}$ for the right triangular distribution, Neyman allocation case, while each was a graduate student in the sample survey course.

Optimum Stratum Boundaries for Sampling from a Right Triangular Distribution or from an Exponential Distribution

A. R. MANSON AND C. H. PROCTOR*

1. INTRODUCTION

Using inexpensively available auxiliary information on the elements of a population to form strata, then allocating optimum sampling effort to each stratum is, perhaps along with clustering, among the most widely used features of sample design. The present problem is concerned with how best to mark off stratum boundaries when the auxiliary information is a single numerical variable, highly correlated with the variable of interest, whose distribution conforms more or less either to a right triangular shape or to that of an exponential distribution.

A variable having a right triangular distribution may occur, for example, in a survey to estimate the average interpoint travel time or fuel usage for all pairs of exits on an interstate highway. Careful measurement of the variable of interest requires actually traveling by car or truck between the sampled pairs of points under a variety of conditions, while interpoint distances are easily obtainable for use as an auxiliary variable. This auxiliary variable is similar in distribution to that of the distance between two independently drawn random variables from a uniform distribution, and would be likely to be close to a right triangular distribution.

In sampling practice, the distribution of an auxiliary variable may be close enough in shape to that of a right triangular distribution or to an exponential distribution to make boundaries which would be optimal for these distributions, of interest as a quick approximate basis for stratification. Sethi (1963) has given such results for the gamma distribution.

Aside from its practical importance, the construction of optimum stratum boundaries has provided rich opportunities for the exercise of elegant mathematics. Witness the paper of Dalenius and Hodges (1959) and others stimulated by this pioneering work that are discussed by Cochran (1977) in his textbook. In this paper, we derive recursive schemes for determining the optimum stratum boundaries for the right triangular and the exponential distributions. These recursive schemes are effective for Neyman, proportional, and for equal allocation of sampling effort. The recursive relationships eliminate the need for multivariate interative solutions in that they provide explicit solutions for the right triangular distribution and a one-dimensional minimization problem for the exponential distribution.

2. THE RIGHT TRIANGULAR DISTRIBUTION

Consider an effectively infinite population having the frequency function:

$$f(x) = 2(1 - x) \quad \text{for } 0 \le x \le 1 \qquad (2.1)$$
$$= 0 \quad \text{elsewhere.}$$

The problem of subdividing the range of X into a total of L strata for the purpose of taking a stratified random sample from the population defined by (2.1), was considered as one of the examples in Dalenius and Hodges (1959) and in Ekman (1959). We shall use standard notation (Cochran 1977, Chapter 5) for the most part, wherein y_{hi} is the value obtained for the i^{th} sampled unit in the h^{th} stratum, n_h is the number of sample observations allocated to the h^{th} stratum, $x_{L,h-1}$ is the lower boundary of the h^{th} stratum, and $x_{L,h}$ is the upper boundary of the h^{th} stratum. We will at times set $x_{L,h-1} = a$ and $x_{L,h} = b$ as shorthand notation. Three basic quantities are expressed as

$$W_{L,h} = \int_a^b f(x)dx, \quad \bar{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h, \quad \text{and} \quad \bar{y}_{ST} = \sum_{h=1}^{L} W_{L,h}\bar{y}_h .$$

In developing the optimum stratum boundaries for Neyman allocation ($n_h \propto W_{L,h}S_{L,h}$), for proportional allocation ($n_h \propto W_{L,h}$), and for equal allocation ($n_h = n/L$), we shall take advantage of recursive relationships existing among these stratum boundaries that arise from a reproducibility property (under left truncation) of f(x). This property may be referred to as "shape reproducibility." Once the upper boundary of the first stratum ($x_{L,1}$) is obtained, that portion of f(x) having $x_{L,1} \le X \le 1$, is optimally stratified for all three types of allocation considered, by mimicking the optimum stratification obtained for L-1 strata, with appropriate rescaling with respect to the altered base of the truncated distribution.

The values of the weighting function, and the variance and mean of the h^{th} stratum are

$$W_{L,h} = \int_a^b 2(1-x)dx = (x_{L,h} - x_{L,h-1})(2-x_{L,h-1} - x_{L,h}) \qquad (2.2)$$

$$\mu_{L,h} = E(X|a=x_{L,h-1} \le X \le x_{L,h}=b) = W_{L,h}^{-1} \int_a^b 2(1-x)dx$$

$$= [3(x_{L,h-1} + x_{L,h}) - 2(x_{L,h-1}^2 + x_{L,h-1}x_{L,h} + x_{L,h}^2)]$$

$$/3(2-x_{L,h-1} - x_{L,h}) \ , \qquad (2.3)$$

and

$$S_{L,h}^2 = E[(X-\mu_{L,h})^2 | a=x_{L,h-1} \le X \le x_{L,h} = b]$$

$$= W_{L,h}^{-1} \int_a^b 2(x-\mu_{L,h})^2 (1-x)dx$$

$$= (x_{L,h} - x_{L,h-1})^2 (6-6x_{L,h-1} - 6x_{L,h} + x_{L,h-1}^2 + x_{L,h}^2 + 4x_{L,h-1}x_{L,h})$$

$$/18(2 - x_{L,h-1} - x_{L,h})^2 \ . \qquad (2.4)$$

If we define

$$Q_L = \sum_{h=1}^{L} W_{L,h} S_{L,h} \ , \qquad (2.5)$$

$$R_L = \sum_{h=1}^{L} W_{L,h} S_{L,h}^2 \ , \qquad (2.6)$$

and

$$T_L = \sum_{h=1}^{L} W_{L,h}^2 S_{L,h}^2 \qquad (2.7)$$

then the variance of $\bar{y}_{ST}$ for Neyman (optimum), proportional, and equal allocation can be written as

$$V_{OPT} = n^{-1} Q_L^2 \ , \qquad (2.8)$$

$$V_{PROP} = n^{-1}R_L , \tag{2.9}$$

and

$$V_{EQ} = n^{-1}LT_L , \tag{2.10}$$

respectively.

The stratum boundaries which minimize the variance of $\bar{y}_{ST}$ are given by repeated applications of Theorem 2.1.

Theorem 2.1 For any arbitrary value of $x_{L,1}(0 < x_{L,1} < 1)$, the variance of $\bar{y}_{ST}$ obtained from subdividing a right triangular distribution into a total of L strata, is minimized for Neyman, proportional, and equal allocation when the stratum boundaries satisfy:

$$x_{L,h} = x_{L,1} + (1- x_{L,1})\, m_{L-1,\ h-1} \text{ for } h = 2,3,\ldots,L, \tag{2.11}$$

where the $m_{L-1,h-1}$ are the stratum boundaries that minimize $\text{Var}(\bar{y}_{ST})$ for L-1 strata for Neyman, proportional, and equal allocation, respectively

Proof: Minimization of $\text{Var}(\bar{y}_{ST})$ is equivalent to minimization of Q_L, R_L or T_L for Neyman, proportional, or equal allocation, respectively. Set the arbitrary value of $x_{L,1}$ temporarily equal to c, where $0 < c < 1$. Then simple substitution gives

$$Q_L = c^2[(6-6c + c^2)/18]^{\frac{1}{2}} + \sum_{h=2}^{L} W_{L,h}S_{L,h} ,$$

$$R_L = [c^3(6-6c + c^2)/18(2-c)] + \sum_{h=2}^{L} W_{L,h}S^2_{L,h} ,$$

and

$$T_L = [c^4(6-6c + c^2)/18] + \sum_{h=2}^{L} W^2_{L,h}S^2_{L,h} .$$

Now establish a correspondence between two sets of boundaries as:

$$x_{L,h} = c + (1-c)z_{L-1,h-1} \text{ for } h = 2,3,\ldots,L. \tag{2.12}$$

One may consider the quantities $W_{L,h}$ and $S_{L,h}$ explicity as functions of these sets of boundaries by writing $W_{L,h}(x) = W_{L,h}$ and $S_{L,h}(x) = S_{L,h}$ or in terms of $z_{L-1,h-1}$ as $W_{L-1,h-1}(z)$ and $S_{L-1,h-1}(z)$. The quantities $Q_L(x)$, $Q_{L-1}(z)$, $R_L(x)$, $R_{L-1}(z)$, $T_L(x)$, and $T_{L-1}(z)$ become defined similarly.

$$W_{L,h}(x) = (1-c)^2(z_{L-1,h-1} - z_{L-1,h-2})(2-z_{L-1,h-2} - z_{L-1,h-1})$$

$$= (1-c)^2 W_{L-1,h-1}(z) \tag{2.13}$$

and

$$S_{L,h}(x) = (1-c)(z_{L-1,h-1} - z_{L-1,h-2})[(6-6z_{L-1,h-2} - 6z_{L-1,h-1} + z^2_{L-1,h-2}$$

$$+ z^2_{L-1,h-1} + z_{L-1,h-2}\ z_{L-1,h-1})/18(2-z_{L-1,h-2} - z_{L-1,h-1})^2]^{\frac{1}{2}}$$

$$= (1-c)S_{L-1,h-1}(z). \tag{2.14}$$

Note that both (2.13) and (2.14) represent L-1 equations which when substituted into Q_L, R_L, and T_L give

$$Q_L(x) = c^2[(6-6c + c^2)/18]^{\frac{1}{2}} + (1-c)^3 \sum_{h=2}^{L} W_{L-1,h-1}(z)S_{L-1,h-1}(z), \tag{2.15}$$

$$R_L(x) = [c^3(6-6c + c^2)/18(2-c)] + (1-c)^4 \sum_{h=2}^{L} W_{L-1,h-1}(z)S^2_{L-1,h-1}(z), \tag{2.16}$$

and

$$T_L(x) = [c^4(6-6c + c^2)/18] + (1-c)^6 \sum_{h=2}^{L} W^2_{L-1,h-1}(z)S^2_{L-1,h-1}(z). \tag{2.17}$$

In the expressions (2.15), (2.16), and (2.17) for Q_L, R_L, and T_L, respectively; one may set $j = h-1$ as a change of summation index.

Comparison of the resulting expressions with those expressions in (2.5), (2.6) and (2.7) shows that

$$Q_L(x) = c^2[(6-6c + c^2)/18]^{\frac{1}{2}} + (1-c)^3 Q_{L-1}(z), \qquad (2.18)$$

$$R_L(x) = [c^3(6-6c + c^2)/18(2-c)] + (1-c)^4 R_{L-1}(z), \qquad (2.19)$$

and

$$T_L(x) = [c^4(6-6c + c^2)/18] + (1-c)^6 T_{L-1}(z). \qquad (2.20)$$

From the expressions of (2.18), (2.19), and (2.20) it is apparent that whatever value of $x_{L,1} = c$ is specified on (0,1), that $Q_L(x)$, $R_L(x)$, and $T_L(x)$ are minimized when $Q_{L-1}(z)$, $R_{L-1}(z)$, and $T_{L-1}(z)$ are minimized. This occurs when the $z_{L-1,h-1}$ are set equal to $m_{L-1,h-1}$ say, the stratum boundaries which minimize $\text{Var}(\bar{y}_{ST})$ for a total of L-1 strata. Thus, the relationship among the optimum stratum boundaries for fixed $x_{L,1} = c$, is

$$x_{L,h} = c + (1-c)z_{L-1,h-1} = c + (1-c)z_{L-1,j} = x_{L,1} + (1-x_{L,1})m_{L-1,h-1},$$

which proves Theorem 2.1.

Because the recursive relationship of (2.11) holds for arbitrary $x_{L,1}$ it holds for $x_{L,1} = m_{L,1}$, that value which gives absolute minimum $\text{Var}(\bar{y}_{ST})$. Hence $m_{L,h} = m_{L,1} + (1-m_{L,1})m_{L-1,h-1}$ for $h = 2,3,\ldots,L$.

To minimize $\text{Var}(\bar{y}_{ST})$ over $x_{L,1} = c$ for the three types of allocation, we note that $Q_{L-1}(z)$, $R_{L-1}(z)$, and $T_{L-1}(z)$ do not contain $x_{L,1} = c$; and therefore we may solve for c by setting $\partial Q_L/\partial c = 0$, $\partial R_L/\partial c = 0$, $\partial T_L/\partial c = 0$ using ordinary calculus. These conditions yield the following quartic equations in $m_{L,1} = c$:

Neyman allocation:

$$(18\tilde{Q}^2_{L-1} - 1)c^4 - 8(18\tilde{Q}^2_{L-1} - 1)c^3 + 2(171\tilde{Q}^2_{L-1} - 8)c^2 - 324\tilde{Q}^2_{L-1}c + 108\tilde{Q}^2_{L-1} = 0 , \qquad (2.21)$$

Proportional allocation:

$$(18\tilde{R}^2_{L-1} - 1)c^4 - 6(18\tilde{R}_{L-1} - 1)c^3 + 9(26\tilde{R}_{L-1} - 1)c^2 - 216\tilde{R}_{L-1}c + 72\tilde{R}_{L-1} = 0, \qquad (2.22)$$

Equal allocation:

$$(18\tilde{T}_{L-1} + 1)c^4 - 4(18\tilde{T}_{L-1} + 1)c^3 + 108\tilde{T}_{L-1}c^2 - 72\tilde{T}_{L-1}c + 18\tilde{T}_{L-1} = 0, \qquad (2.23)$$

where $\tilde{Q}_{L-1}$, $\tilde{R}_{L-1}$, and $\tilde{T}_{L-1}$, are the minimum values of Q_{L-1}, R_{L-1}, T_{L-1}, respectively. Solutions of (2.21), (2.22), and (2.23) may be obtained by noting that $\tilde{Q}_1 = \sqrt{1/18}$ and $\tilde{R}_1 = \tilde{T}_1 = 1/18$ and then solving for $m_{2,1} = c$ in the appropriate quartic equation.

Two of the solutions are of special interest in that the quartic equations simplify to quadratic equations. For Neyman allocation with L=2, the resulting quadratic equation is $3c^2 - 18c + 6 = c^2 - 6c + 2 = 0$, which has one admissible solution at $m_{2,1} = c = 3 - \sqrt{7} \doteq 0.354248689$, giving minimum $Q_2 = \tilde{Q}_2 = [(667 - 342\sqrt{7})/18]^{\frac{1}{2}}$ and minimum$[n\cdot Var(\bar{y}_{ST})] = \tilde{Q}^2_2 = (667 - 342\sqrt{7})/18 \doteq 0.015037007$. For proportional allocation with L = 2 strata, the resulting quadratic equation is $4c^2 - 12c + 4 = c^2 - 3c + 1 = 0$ which has one admissible solution at $m_{2,1} = c = (3 - \sqrt{5})/2$, giving minimum R_2 = minimum$[n\cdot Var(\bar{y}_{ST})] = \tilde{R} = [(45-20\sqrt{5})/18] \doteq 0.0154800256$.

The method for computing the remaining stratum boundaries is to substitute $\tilde{Q}_2$, $\tilde{R}_2$ and $\tilde{T}_2$ into (2.21), (2.22), and (2.23) respectively; solve for $m_{3,1} = c$; employ the recursive relationship of (2.11) to solve for $m_{3,2}$; determine minimum Q_3, R_3, and T_3; and repeat the aforementioned process for as many strata as one desires. The recursive relationship of (2.11) may be rewritten in the modified form

$$m_{L,h} = \bigcup_{i=L-h+1}^{L} m_{i,1} = \sum_{i=L-h+1}^{L} m_{i,1} - \sum\sum_{i<j} m_{i,1} m_{j,1} + \ldots + (-1)^h \prod_{i=L-h+1}^{L} m_{i,1}, \quad (2.24)$$

showing that the optimum stratum boundaries are functions only of the upper boundaries of the first stratum, $m_{i,1}$. The results obtained for L = 2(1)6 are presented in Table 1.

(INSERT TABLE 1 HERE)

It is of interest to note that the use of equal allocation of sampling effort increases the variance of $\bar{y}_{ST}$ over the level obtained using Neyman allocation by less than 0.05% (one twentieth of 1%) which is an inconsequential increase. Furthermore as the number of strata increases, the variance of $\bar{y}_{ST}$ for equal allocation approaches the value obtained for Neyman allocation. Of equal importance is the fact that to approximately two decimal places the optimum stratum boundaries for both Neyman allocation and for equal allocation are the same. On the other hand, each of the optimum stratum boundaries obtained for proportional allocation are 5% to 10% larger than those obtained for Neyman allocation, which costs an experimenter a 3% to 5% increase in $n \cdot \text{Var}(\bar{y}_{ST})$ over the value obtained using Neyman allocation, for L = 2,3,...,6 strata. One could reduce this cost to an increase in $n \cdot \text{Var}(\bar{y}_{ST})$ of less than 2% by using Neyman allocation for those stratum boundaries which are optimum for proportional allocation, although in practice one might be quite reluctant to give up the convenience of uniform sampling fractions.

3. THE EXPONENTIAL DISTRIBUTION

Dalenius and Hodges (1959) and Ekman (1959) considered the optimum stratification of the frequency function

$$f(x) = e^{-x} \quad \text{for } x > 0, \tag{3.1}$$

by using multidimensional optimization techniques. We shall derive recursive relationships among the optimum $x_{L,h}$ for the exponential distribution of (3.1) for Neyman, proportional, and equal allocation.

Using notation identical to that used for the right triangular distribution in the previous section, we obtain

$$W_{L,h} = e^{-x_{L,h-1}} - e^{-x_{L,h}}, \tag{3.2}$$

$$\mu_{L,h} = [(x_{L,h-1} + 1)e^{-x_{L,h-1}} - (x_{L,h} + 1)e^{-x_{L,h}}]/W_{L,h}, \tag{3.3}$$

and

$$S^2_{L,h} = \{(e^{-x_{L,h-1}} - e^{-x_{L,h}})^2 - (x_{L,h} - x_{L,h-1})^2 e^{-(x_{L,h-1} + x_{L,h})}\} / W^2_{L,h}. \tag{3.4}$$

Note that $x_{L,0} = 0$ and $x_{L,L} = \infty$.

The variance of $\bar{y}_{ST}$ for the three types of allocation may be expressed in terms of Q_L, R_L, and T_L as in (2.8), (2.9), and (2.10) but with the values of $W_{L,h}$ and $S^2_{L,h}$ as given in (3.2) and (3.4). The recursive relationship of Theorem 3.1 may then be used to simplify the minimization of $\text{Var}(\bar{y}_{ST})$.

Theorem 3.1. For any arbitrary value of $x_{L,1}(x_{L,1} > 0)$, the variance of $\bar{y}_{ST}$ obtained from subdividing the exponential distribution into a total of L strata for Neyman, proportional, and equal allocation , is minimized when the stratum boundaries satisfy the recursive relationship:

$$x_{L,h} = x_{L,1} + m_{L-1,h-1}, \tag{3.5}$$

where the $m_{L-1,h-1}$ are the stratum boundaries which minimize $Var(\bar{y}_{ST})$ for a total of L-1 strata for Neyman, proportional, and equal allocation, respectively.

Proof: Because minimization of $Var(\bar{y}_{ST})$ is equivalent to minimization of Q_L, R_L, and T_L of (2.5), (2.6) and (2.7) for the three types of allocation, let the arbitrary value of $x_{L,1} = c$ $(c > 0)$. This gives

$$Q_L = [(1-e^{-c})^2 - c^2 e^{-c}]^{\frac{1}{2}} + \sum_{h=2}^{L} W_{L,h} S_{L,h}$$

$$= [(1-e^{-c})^2 - c^2 e^{-c}]^{\frac{1}{2}} + \sum_{h=2}^{L} W_{L,h} S_{L,h}(x),$$

$$R_L = \{[(1-e^{-c})^2 - c^2 e^{-c}]/(1-e^{-c})\} + \sum_{h=2}^{L} W_{L,h} S^2_{L,h} ,$$

$$= \{[(1-e^{-c})^2 - c^2 e^{-c}]/(1-e^{-c})\} + \sum_{h=2}^{L} W_{L,h}(x) S^2_{L,h}(x)$$

and

$$T_L = [(1-e^{-c})^2 - c^2 e^{-c}] + \sum_{h=2}^{L} W^2_{L,h} S^2_{L,h},$$

$$= [(1-e^{-c})^2 - c^2 e^{-c}] + \sum_{h=2}^{L} W^2_{L,h}(x) S^2_{L,h}(x).$$

The linear correspondence between sets of boundaries appropriate for the exponential distribution

$$x_{L,h} = c + z_{L-1,h-1} \quad \text{for} \quad x_{L,h} > c \text{ and } h = 2,3,\ldots,L$$

gives $z_{L-1,h-1} > 0$ and

$$W_{L,h}(x) = e^{-c}(e^{-z_{L-1,h-2}} - e^{-z_{L-1,h-1}}) = e^{-c}W_{L-1,h-1}(z) \qquad (3.7)$$

and

$$S_{L,h}(x) = e^{-c}\{(e^{-z_{L-1,h-2}} - e^{-z_{L-1,h-1}})^2$$

$$- (z_{L-1,h-1} - z_{L-1,h-2})^2 e^{-(z_{L-1,h-2} + z_{L-1,h-1})}\}^{\frac{1}{2}}/(e^{-z_{L-1,h-2}} - e^{-z_{L-1,h-1}})$$

$$= e^{-c}S_{L-1,h-1}(z). \qquad (3.8)$$

These expressions may be substituted into the equations for Q_L, R_L, and T_L to obtain

$$Q_L(x) = [(1-e^{-c})^2 - c^2 e^{-c}]^{\frac{1}{2}} + e^{-c}\sum_{h=1}^{L-1} W_{L-1,h}(z)S_{L-1,h}(z) \qquad (3.9)$$

$$R_L(x) = \{[(1-e^{-c})^2 - c^2 e^{-c}]/(1-e^{-c}\sum_{h=1}^{L-1} W_{L-1,h}(z)S^2_{L-1,h}(z), \qquad (3.10)$$

and

$$T_L(x) = [(1-e^{-c})^2 - c^2 e^{-c}] + e^{-2c}\sum_{h=2}^{L} W^2_{L-1,h}(z)S^2_{L-1,h}(z) \qquad (3.11)$$

Comparison of (3.9), (3.10) and (3.11) with the original expressions for Q_L, R_L and T_L gives

$$Q_L(x) = [(1-e^{-c})^2 - c^2 e^{-c}]^{\frac{1}{2}} + e^{-c}Q_{L-1}(z), \qquad (3.12)$$

$$R_L(x) = \{[(1-e^{-c})^2 - c^2 e^{-c}]/(1-e^{-c})\} + e^{-c}R_{L-1}(z), \qquad (3.13)$$

and

$$T_L(x) = [(1-e^{-c})^2 - c^2 e^{-c}] + e^{-2c}T_{L-1}(z). \qquad (3.14)$$

These expressions show that regardless of the arbitrary value of $x_{L,1} = c$, that Q_L, R_L, and T_L are minimized when Q_{L-1}, R_{L-1}, and T_{L-1} respectively are minimized. This occurs when the $z_{L-1,h-1}$ are set equal to the $m_{L-1,h-1}$ say, the values of the stratum boundaries which minimize $\text{Var}(\bar{y}_{ST})$ for a total of L-1 strata. Thus, the relationship among the optimum stratum boundaries for fixed $x_{L,1} = c$, is

$$x_{L,h} = c + z_{L-1,h-1} = c + z_{L-1,j} = x_{L,1} + m_{L-1,h-1} \text{ for } h = 2,3,\ldots,L.$$

Because the recursive relationship of (3.5) holds for arbitrary $x_{L,1}$, it must hold for $x_{L,1} = m_{L,1}$, the value which gives absolute minimum $\text{Var}(\bar{y}_{ST})$. Hence $m_{L,h} = m_{L,1} + m_{L-1,h-1}$.

To minimize $\text{Var}(\bar{y}_{ST})$ over $x_{L,1} = c$ for the three types of allocation, note that Q_{L-1}, R_{L-1}, and T_{L-1} do not contain $x_{L,1} = c$. Therefore setting $\partial Q_L/\partial c = 0$, $\partial R_L/\partial c = 0$, and $\partial T_L/\partial c = 0$ to find c gives:

Neyman allocation:

$$c^4 - 4c^3 + 8c^2 - 8c + 4(1-\tilde{Q}^2_{L-1}) - 4e^{-c}[(c^2+2)(1-\tilde{Q}^2_{L-1}) - 2c] + 4e^{-2c}(1-\tilde{Q}^2_{L-1}) = 0, \quad (3.15)$$

Proportional allocation:

$$c^2 - 2c + (1-\tilde{R}_{L-1}) - 2e^{-c}[(1-\tilde{R}_{L-1})-c] + (1-\tilde{R}_{L-1})e^{-2c} = 0, \quad (3.16)$$

Equal allocation:

$$c^2 - 2c + 2 - 2(1+\tilde{T}_{L-1})e^{-c} = 0, \quad (3.17)$$

where $\tilde{Q}_{L-1}$, $\tilde{R}_{L-1}$, and $\tilde{T}_{L-1}$ are written to represent the minimum values of Q_{L-1}, R_{L-1}, and T_{L-1}, respectively.

By noting that $\tilde{Q}_1 = \tilde{T}_1 = \tilde{R}_1 = 1$, (3.15), (3.16), and (3.17) may be solved for $m_{2,1} = c$. With $x_{2,1}$ known, $\tilde{Q}_2$, $\tilde{R}_2$, and $\tilde{T}_2$ may be computed, substitited into (3.15), (3.16), and (3.17) respectively; and $c = m_{3,1}$ determined. The recursive relationship of Theorem 3.1 given in (3.5) may then be used to determine $m_{3,2}$. Alternatively, one may re-write (3.5) in slightly modified form to give

$$m_{L,h} = \sum_{i=L-h+1}^{L} m_{i,1}, \quad (3.18)$$

for $h \geq 2$, which clearly indicates that once the $m_{L,1}$ are known then all $m_{L,h}$ are completely specified. The aforementioned procedure may be interated to obtain the optimum stratum boundaries for as many strata as one desires. The results obtained for L = 2(1)6 are presented in Table 2.

(INSERT TABLE 2 HERE)

Once again, the small size of the deviation of $Var(\bar{y}_{ST})$ for optimum equal allocation from that obtained for Neyman allocation is striking. The cost to an investigator (in terms of $Var(\bar{y}_{ST})$) for using equal allocation is less than 0.12%, and the optimum stratum boundaries for equal allocation are less than 3% larger than those obtained for Neyman allocation. Furthermore, as the number of strata increases the Neyman and equal allocation solutions draw even closer together. This occurs because $W_{L,h} S_{L,h}$ is nearly constant for the exponential distribution (as it was for the right triangular distribution) and under such circumstances Neyman allocation gives a constant sample size $n_h = n/L$ for all strata, see Cochran (1977, page 130).

On the other hand, the optimum stratum boundaries for proportional allocation run 26% to 41% larger than those obtained for Neyman allocation, and the cost to the experimenter is a 23% to 50% increase in $Var(\bar{y}_{ST})$, for L = 2,3,4,5 and 6 strata. In addition, as the number of strata increases, the optimum boundaries for proportional allocation get further (percentagewise) from those obtained for Neyman allocation, accompanied by a steady increase (percentagewise) in $Var(\bar{y}_{ST})$ over the level obtained for Neyman allocation.

4. CONCLUDING REMARKS

Tables 1 and 2 clearly indicate that when the distribution of the auxiliary variable is near that of a right triangular distribution or that of an exponential distribution and optimum stratum boundaries have been designated,one may simply use equal allocation of sampling effort over all strata.

There are several exact boundary points for two strata which are expressible neatly in arithmetical terms. For the right triangular distribution, the exact value of the optimum upper boundary of the first stratum out of L = 2 strata under Neyman allocation is $x_{2,1} = 3 - \sqrt{7}$, and thus joins the optimum boundary for proportional allocation $x_{2,1} = (3-\sqrt{5})/2$, which is also the optimum boundary for Ekman's rule (1959) of making $W_{L,h}(X_{L,h}-X_{L,h-1})$ constant, as values to be compared to the $1 - (4)^{1/3}$ value that Cochran (1977, page 148) gives for the cumulative $[f(x)]^{\frac{1}{2}}$ rule.

The recursive relationship (2.11) among the optimum stratum boundaries for the right triangular distribution eliminates the need for an L-1 dimensional minimization solution and gives an explicit method of solution. This recursive relationship is made possible by the shape reproducibility property of the right triangular distribution. For the exponential distribution, the recursive relationship (3.5) among the optimum boundaries reduces an L-1 dimensional minimization problem to that of solving a single exponential equation in one variable. The recursive relationship for the exponential distribution occurs because of the "time independence" property of this distribution which is also shape reproducibility. Such recursive relationships will exist among the optimum stratum boundaries for other probability distributions having shape reproducibility and thereby provide simple techniques for their solutions.

TABLE 1.

Minimum Variance Stratum Boundaries for the Right Triangular Distribution: $f(x) = 2(1-x)$ for $0<x<1$.

Number of Strata (L)	Stratum Boundaries $x_{L,1}$	$x_{L,2}$	$x_{L,3}$	$x_{L,4}$	$x_{L,5}$	n·Min Var($\bar{y}_{ST}$)
			Neyman Allocation			
2	0.3542487	-------	-------	-------	-------	0.0150372007
3	0.2297821	0.5026308	-------	-------	-------	0.0068784215
4	0.1705247	0.3611233	0.5874445	-------	-------	0.0039271382
5	0.1356650	0.2830554	0.4477965	0.6434138	-------	0.0025363347
6	0.1126701	0.2330497	0.3638336	0.5100133	0.6835904	0.0017721818
			Equal Allocation			
2	0.3591492	-------	-------	-------	-------	0.0150433505
3	0.2320589	0.5078643	-------	-------	-------	0.0068811370
4	0.1718301	0.3640143	0.5924281	-------	-------	0.0039285032
5	0.1365096	0.2848833	0.4508325	0.6480656	-------	0.0025371067
6	0.1132608	0.2343093	0.3658780	0.5130317	0.6879260	0.0017726583
			Proportional Allocation			
2	0.3819660	-------	-------	-------	-------	0.0154800256
3	0.2513940	0.5373360	-------	-------	-------	0.0071608428
4	0.1878838	0.3920449	0.6242631	-------	-------	0.0041128094
5	0.1501032	0.3097850	0.4833009	0.6806624	-------	0.0026660836
6	0.1250085	0.2563476	0.3960678	0.5478927	0.7205823	0.0018675329

TABLE 2.

Minimum Variance Stratum Boundaries for the Exponential Distribution: $f(x) = e^{-x}$ for $x>0$.

Number of Strata (L)	$x_{L,1}$	$x_{L,2}$	$x_{L,3}$	$x_{L,4}$	$x_{L,5}$	n·Min Var($\bar{y}_{ST}$)
	Stratum Boundaries					
			Neyman Allocation			
2	1.2619065	-------	-------	-------	-------	0.2854738212
3	0.7639635	2.0258700	-------	-------	-------	0.1332249230
4	0.5506425	1.3146060	2.5765125	-------	-------	0.0768680910
5	0.4310264	0.9816689	1.7456325	3.0075389	-------	0.0499687474
6	0.3542833	0.7853097	1.3359522	2.0999158	3.3618222	0.0350683714
			Equal Allocation			
2	1.3000752	-------	-------	-------	-------	0.2858096777
3	0.7790597	2.0791350	-------	-------	-------	0.1333775088
4	0.5587499	1.3378096	2.6378849	-------	-------	0.0769459747
5	0.4360882	0.9948381	1.7738978	3.0739731	-------	0.0500132227
6	0.3577447	0.7938329	1.3525828	2.1316425	3.4317178	0.0350960023
			Proportional Allocation			
2	1.5936243	-------	-------	-------	-------	0.3523897621
3	1.0175778	2.6112021	-------	-------	-------	0.1797366122
4	0.7540323	1.7716101	3.3652344	-------	-------	0.1089519758
5	0.6004295	1.3544618	2.3720396	3.9656639	-------	0.0730895386
6	0.4993198	1.0997493	1.8537816	2.8713594	4.4649837	0.0524268215

REFERENCES

Cochran, William G. (1977), Sampling Techniques, 3rd Edition, New York: John Wiley and Sons.

Dalenius, Tore (1957), Sampling in Sweden, Stockholm: Almquist and Wiksell.

______, and J. L. Hodges (1959), "Minimum Variance Stratification," Journal of the American Statistical Association, 54, 88-101.

Ekman, Gunnar (1959), "An Approximation Useful in Univariate Stratification," The Annals of Mathematical Statistics, 30, 219-229.

Sethi, V. K. (1963), "A Note on Optimum Stratification for Estimating the Population Means," Australian Journal of Statistics, 5, 20-33.

BIOMATHEMATICS TRAINING PROGRAM