

Stratum Boundaries at Equal Steps of Cumulative Root Frequency Come Close to Optimal, While Equal Counts Do Worst

— dividing a skewed timber-volume distribution into three strata with allocation matched to variability, boundaries equalising counts give more error than boundaries equalising value ranges, and boundaries equalising the square root of frequency, in between, nearly reach the searched optimum —the separator is whether boundaries equalise value ranges, counts, or something in between — [Paper 957]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

YOI: <https://www.yamagishiyuuki.com/yoi/957>

— *From thinking to syncing.* —

Abstract

When a forest is divided into strata for sampling, where the stratum boundaries are drawn changes the estimation error. This paper divides a skewed volume distribution, with few large trees and many small ones, into three strata, and compares the variance of the estimated mean with 30 plots allocated to match stratum variability, for boundaries equalising value ranges, boundaries equalising counts, boundaries at equal steps of cumulative root frequency, and searched optimal boundaries. No new theorem or law is claimed. First, boundaries at equal steps of cumulative root frequency give almost the same error as the searched optimum: the variance of the estimate was **1.002608** times the optimum, and 0.141324 times simple random sampling (the optimum being 0.140956 times) (Section 2). Second, boundaries equalising counts do worst: drawing boundaries so each stratum holds the same number of trees gave a variance **1.680702** times the optimum, larger than the 1.454160 times of boundaries equalising value ranges (Section 2). Third, the boundary positions are also close to optimal: boundaries at equal steps of cumulative root frequency were 103.181 and 267.212, the searched optimum 108.272 and 268.497, while boundaries equalising counts were 63.779 and 127.050, shifted to the small side (Section 2). Fourth, stratifying gains a lot with any boundaries: against a simple random sampling variance of 457.395297, even the worst boundaries, equalising counts, brought it down to 108.359509 (0.236906 times) (Section 2). Fifth, the separator is whether boundaries equalise value ranges, counts, or something in between: in a skewed distribution, equalising counts shifted the boundaries to the small side and equalising value ranges shifted them to the large side. Equalising the square root of frequency gave boundaries in between, close to the searched optimum (Section 3). **Keywords:** stratified sampling, stratum boundaries, Dalenius-Hodges method, Neyman allocation, variance of the estimate, skewed distribution.

Scope of this paper (scope note): No new theorem or law is claimed: stratified sampling and Neyman allocation, dividing stratum boundaries at equal steps of cumulative root frequency (Dalenius and Hodges), and that this method gives boundaries close to optimal are all known in sampling theory. The distribution is set: volume is set as a lognormal distribution with log-mean 4.5 and log-standard deviation 0.8. It corresponds to no forest or species, and the order for other skewed distributions is not claimed. The stratifying value and the measured value are taken as the same: in real surveys strata are formed from another guide before volume is measured, but here strata are formed by volume itself. Costs and two-stage sampling are not treated: differing measurement costs by stratum, subsampling within plots, and numbers of strata other than three are not treated. Relation to earlier papers: Paper 938, in its scope note, did not treat how strata are drawn or costs, leaving out how to draw stratum boundaries, strata with different measurement costs, and two-stage sampling within plots; this paper sits in the seat of how to draw stratum boundaries (no earlier occupant). Paper 938 showed that stratifying a forest removes between-stratum differences from sampling error, and allocating more plots to variable strata reduces it further. Here it is counted how much further the error changes with the drawing of boundaries under the same allocation. What is added: comparing the variance of the estimate for four ways of drawing boundaries, showing that cumulative root frequency boundaries reach 1.002608 times the searched optimum, showing that equalising counts does worst in a skewed distribution, confirming the optimum by refining beyond the grid, and placing the separator on “value ranges, counts, or in between”.

1. Notation and the Question

Distribution: volume is set as a lognormal distribution (log-mean 4.5, log-standard deviation 0.8). The mean is 123.965091 and the standard deviation 117.373463, with few large trees and many small ones. The distribution was represented by 20000 equal-probability points.

Variance of the estimate: dividing into three strata with boundaries $b_1 < b_2$, with stratum shares W_h and within-stratum standard deviations S_h , the variance of the estimated mean when $n = 30$ plots are allocated by Neyman allocation is:

$$V = \frac{(\sum_h W_h S_h)^2}{n}$$

Ways of drawing boundaries: (1) equalise value ranges: divide the interval from the 1 per cent point 13.998 to the 99 per cent point 578.870 into three. (2) equalise counts: draw so that each stratum holds the same share. (3) cumulative root frequency: add up the square root of frequency along the values and draw at the points dividing the total into three (Dalenius and Hodges). (4) searched optimum: choose two of 180 quantile candidates exhaustively, then refine beyond the grid by the Nelder-Mead method.

Question: in a skewed distribution, how should stratum boundaries be drawn to reduce estimation error?

2. Four Ways of Drawing Boundaries

Computation 2.1: boundaries and variance of the estimate under Neyman allocation.

method	boundaries	variance of estimate	ratio to simple random	ratio to optimum
equal value ranges	202.289, 390.580	93.753723	0.204973	1.454160
equal counts	63.779, 127.050	108.359509	0.236906	1.680702
cumulative root frequency	103.181, 267.212	64.640924	0.141324	1.002608
searched optimum	108.272, 268.497	64.472760	0.140956	1.000000

Simple random sampling (30 plots) had variance 457.395297. The exhaustive optimum 64.474556 (boundaries 108.418 and 267.458) was refined to 64.472760 by Nelder-Mead.

Observation 2.2

cumulative root frequency nearly reaches the optimum, and equalising counts does worst. Boundaries from cumulative root frequency fell near the searched optimal boundaries, with variance 1.002608 times the optimum: boundaries equalising counts shifted to the small side, with variance 1.680702 times the optimum, larger than the 1.454160 times of equal value ranges. Even so it was 0.236906 times simple random sampling, so the gain from stratifying itself is large.

3. The Separator

Observation 3.1

the separator is whether boundaries equalise value ranges, counts, or something in between. Equalising counts shifted the boundaries toward the many small trees (63.779 and 127.050): equalising value ranges shifted the boundaries to the large side (202.289 and 390.580). Cumulative root frequency lies between frequency itself (counts) and value ranges, which ignore frequency, and its boundaries were close to the searched optimum.

Observation 3.2

what is not said. The order for other skewed distributions, numbers of strata other than three, strata formed from another guide, stratum costs, two-stage sampling and real forests are not treated. What is shown is that in a set skewed distribution, the drawing of boundaries changes the Neyman allocation error from 1.002608 to 1.680702 times the optimum.

4. Roll Call and Closing

1. Neyman variance and four ways of drawing (Section 1): value ranges, counts, root frequency, optimum.
2. **1.002608** times the optimum by cumulative root frequency (Computation 2.1): nearly reaches it.
3. **1.680702** times with equal counts (Computation 2.1): worst.
4. **1.454160** times with equal value ranges (Computation 2.1): in between.
5. all at or below **0.236906** of simple random (Computation 2.1): the gain from stratifying is large.
6. value ranges, counts, or in between (Observation 3.1): the separator; the scope note of Paper 938.

Closing: Stratum boundaries at equal steps of cumulative root frequency come close to optimal, while equal counts do worst. Dividing a skewed volume distribution into three strata with 30 plots by Neyman allocation, boundaries from cumulative root frequency gave a variance 1.002608 times the searched optimum: boundaries equalising counts gave 1.680702 times and those equalising value ranges 1.454160 times, all far below simple random sampling. The separator is whether boundaries equalise value ranges, counts, or something in between. Placed among the earlier papers: to the reduction of error by allocation that Paper 938 counted, the drawing of boundaries that Paper 938 did not treat was added. To be honest: how strata should be formed in real surveys is not claimed. In a set distribution, how much the drawing of boundaries changes the error, and no further.

References

1. Cochran, W. G. (1977). *Sampling Techniques*. Wiley.
2. Dalenius, T., & Hodges, J. L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*.
3. Yamagishi, Y. (2026). *Stratifying a Forest Removes Between-Stratum Differences from Sampling Error, and Allocating More Plots to Variable Strata Reduces It Further* [Paper 938]. YOI: <https://www.yamagishiyuuki.com/yoi/938>

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)