

EFFICIENT PARETO STRATIFICATION

BY

PATRICIA GUNNING, JANE M. HORGAN*

and

GARY KEOGH

School of Computing, Dublin City University

[Accepted 12 July 2005. Published 28 August 2006.]

ABSTRACT

If stratification boundaries are taken in geometric progression, then the Pareto distributions give equal coefficients of variation on strata. This stratification method does not, however, satisfy exactly the Dalenius condition for minimising the variance of the sample mean, but it compares favourably with the commonly used cumulative root frequency approximation of Dalenius and Hodges in terms of the precision of the stratified mean.

1. Introduction

For a continuous density $f(x)$ in the range (β, γ) , where $-\infty \leq \beta < \gamma \leq \infty$, with distribution

$$F(x) = \int_{\beta}^x f(t)dt,$$

we have the mean

$$\mu = \int_{\beta}^{\gamma} tf(t)dt = \int_{\beta}^{\gamma} t dF(t)$$

and variance

$$\sigma^2 = \int_{\beta}^{\gamma} (t - \mu)^2 f(t)dt = \int_{\beta}^{\gamma} (t - \mu)^2 dF(t).$$

To stratify a population into L strata is to subdivide the range (β, γ) into subintervals $[k_{h-1}, k_h]$, $\beta = k_0 < k_1 < k_2 < \dots < k_{L-1} < k_L = \gamma$. The conditional mean of the h^{th} stratum is

$$\mu_h = \int_{k_{h-1}}^{k_h} t \left(\frac{f(t)}{\int_{k_{h-1}}^{k_h} f(s)ds} \right) dt = \frac{1}{W_h} \int_{k_{h-1}}^{k_h} tf(t)dt,$$

*Corresponding author, e-mail: jhorgan@computing.dcu.ie

where

$$W_h = \int_{k_{h-1}}^{k_h} f(t) dt.$$

Similarly, the conditional variance of the h^{th} stratum

$$\sigma_h^2 = \int_{k_{h-1}}^{k_h} (t - \mu_h)^2 \left(\frac{f(t)}{\int_{k_{h-1}}^{k_h} f(s) ds} \right) dt = \frac{1}{W_h} \int_{k_{h-1}}^{k_h} t^2 f(t) dt - \mu_h^2.$$

If a sample of size n is allocated among the strata so that $n_h \geq 1$ is selected from stratum $h, 1 \leq h \leq L$, then

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$$

is an unbiased estimator of μ_h with variance σ_h^2/n_h . Here, x_{hi} is the i^{th} sample element in the h^{th} stratum.

Also, the stratified mean, defined as

$$\bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h, \quad (1.1)$$

is an unbiased estimator of the population mean μ with variance

$$V(\bar{x}_{st}) = \sum_{h=1}^L W_h^2 \frac{\sigma_h^2}{n_h}. \quad (1.2)$$

The aim is to choose optimum breaks and optimum allocation of the sample elements among the strata: i.e. those which minimise (1.2).

With respect to sample allocation, Neyman [6] showed that (1.2) is minimised for fixed n when the n_h are allocated among the strata as follows:

$$n_h = \left(\frac{W_h \sigma_h}{\sum_{i=1}^L W_i \sigma_i} \right) n. \quad (1.3)$$

Substituting (1.3) into (1.2) gives

$$V_{ney}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L W_h \sigma_h \right)^2,$$

which can be written as

$$V_{ney}(\bar{x}_{st}) = \frac{1}{n} \left(\sum_{h=1}^L \sigma_h \int_{k_{h-1}}^{k_h} f(t) dt \right)^2. \quad (1.4)$$

Now we need to choose the break points in such a way as to minimise (1.4). Dalenius [2] showed that this is achieved when the k_h satisfy

$$\frac{\sigma_h^2 + (k_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (k_h - \mu_{h+1})^2}{\sigma_{h+1}}. \quad (1.5)$$

However, these equations are ill adapted to practical computations because μ_h and σ_h depend on k_h . While many approximations have been given to (1.5), no exact solution exists. The aim of this paper is to examine this problem again, and to propose an approximation specifically for skewed populations. The proposal is based on the suggestion by numerous researchers in the field (see, for instance Cochran [1], Dalenius [2], Dalenius and Hodges [3], and Lavallée and Hidiroglou [7]) that it is desirable, when stratifying skewed populations, to arrange for equal coefficients of variation (CVs) in each subinterval.

In Section 2, we show that, for Pareto distributions, equality of CVs can be achieved by the simple ploy of subdividing the range of the variable in geometric progression. We go on in Section 3 to show that geometric stratification of the Pareto distribution does not satisfy (1.5) exactly, but we illustrate in Section 4 that it outperforms the commonly used cumulative root frequency method of stratum construction in terms of the precision of the stratified mean (1.1), when applied to some real skewed populations.

2. Geometric stratification of the Pareto distribution

The Pareto distributions [4] are given by taking

$$f(x) = \begin{cases} \lambda\beta^\lambda x^{-\lambda-1} & x \geq \beta \\ 0 & x < \beta \end{cases}, \quad (2.1)$$

with $\lambda > 2$ and $\beta \geq 1$; they are named after the nineteenth century Italian economist Vilfredo Pareto, who used them to model populations with considerable skewness in the distribution of wealth. If appropriate values of λ and β are chosen in (2.1), then the distribution of wealth will obey the '80-20 rule', in which 20% of the

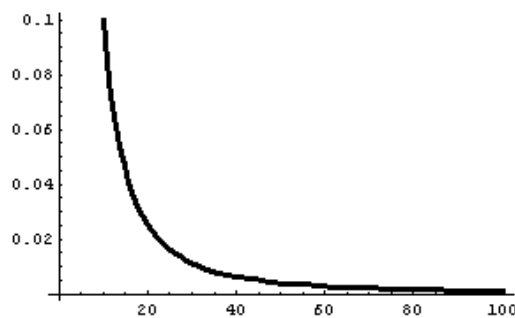


FIG. 1—A Pareto Distribution: $\lambda = 5$, $\beta = 20$

population owns 80% of the wealth: this was Pareto's empirical observation in Italy at the time. Also known as the 'power law', the Pareto distributions are highly skewed. An example of a Pareto distribution is given in Figure 1, with $\lambda = 5$ and $\beta = 20$.

Stratification of skewed distributions such as these often leads to significant improvements in the precision of the mean estimate (1.1).

Proposition 1. *If the break points $\beta = k_0 < k_1 < \dots < k_L = \gamma$ for a Pareto distribution are taken in geometric progression, then successive CVs are equal.*

PROOF. Suppose $f(\cdot)$ is a density function, with corresponding distribution function $F(x) = \int_{-\infty}^x f(t)dt$: we wish to compare the CVs on an arbitrary pair of adjacent subintervals $[a, b] = [sb, b]$ and $[b, c] = [b, rb]$, where $0 < s < 1 < r$.

Define

$$F_b(r) = F(rb) - F(b) = \int_{x=b}^{rb} f(x)dx$$

$$F_b^{(1)}(r) = \int_{x=b}^{rb} xf(x)dx = \mu_b(r)F_b(r) ,$$

$$F_b^{(2)}(r) = \int_{x=b}^{rb} x^2f(x)dx = \mu_b^{(2)}(r)F_b(r) ,$$

so that μ_b and $\mu_b^{(2)}$ are successive 'moments' restricted to subintervals $[b, rb]$, and

$$CV_b(r)^2 = \frac{F_b^{(2)}(r)F_b(r) - F_b^{(1)}(r)^2}{F_b^{(1)}(r)^2} = \frac{\sigma_b(r)^2}{\mu_b(r)^2} \quad (2.2)$$

for the associated coefficient of variation. If, in particular, f is a Pareto density, given by (2.1), then

$$F_b(r) = \beta^\lambda b^{-\lambda}(1 - r^{-\lambda}) , \quad (2.3)$$

$$F_b^{(1)}(r) = \frac{\lambda}{\lambda - 1} \beta^\lambda b^{1-\lambda}(1 - r^{1-\lambda}) ,$$

$$F_b^{(2)}(r) = \frac{\lambda}{\lambda - 2} \beta^\lambda b^{2-\lambda}(1 - r^{2-\lambda}) ,$$

so that, with $\lambda = \ell + 2$, equality of CVs,

$$CV_b(r) = CV_b(s) \quad (2.4)$$

reduces to

$$s^2 r^2 (1 - s^\ell)(1 - s^2 s^\ell)(1 - r r^\ell)^2 = s^2 r^2 (1 - r^\ell)(1 - r^2 r^\ell)(1 - s s^\ell)^2 . \quad (2.5)$$

The endpoints $a = sb, b, rb = c$ of successive intervals $[a, b], [b, c]$ form a geometric

progression provided $rs = 1$. The assumption $rs = 1$ reduces the left hand side of (2.5) to

$$\begin{aligned} (1 - s^\ell)(1 - rr^\ell)(1 - s^2s^\ell)(1 - rr^\ell) &= (1 - s^\ell - rr^\ell + r)(1 - s^2s^\ell - rr^\ell + s) \\ &= 2 + 2(s + r) - (s^\ell + r^\ell) - (rr^\ell + ss^\ell) - (s^2s^\ell + r^2r^\ell) + (s^2s^{2\ell} + r^2r^{2\ell}), \end{aligned}$$

repeatedly using the assumption $rs = 1$. Since the left hand side of (2.5) is now symmetric in r and s , it must be the same as the right hand side. This gives equality (2.4). Now apply this repeatedly with

$$(a, b, c) = (k_{h-1}, k_h, k_{h+1}).$$

■

3. Is geometric stratification optimum?

Rewriting Dalenius's optimum conditions for minimum variance given in (1.5), for successive intervals $[sb, b]$, $[b, rb]$, we have

$$\sigma_b(r) + \frac{(b - \mu_b(r))^2}{\sigma_b(r)} = \sigma_b(s) + \frac{(b - \mu_b(s))^2}{\sigma_b(s)}. \quad (3.1)$$

We show here that, for the Pareto distributions (2.1), equality (3.1) is liable to fail: the proof leans on the equality (2.4) of the CVs.

Proposition 2. *If f is a Pareto distribution and $rs = 1$, then equality (3.1) is liable to fail.*

PROOF. If f is given by (2.1) then equality (2.4) reduces (3.1) to

$$\frac{F_b^{(1)}(r)}{F_b(r)} CV_b^2(r) + \frac{(bF_b(r) - F_b^{(1)}(r))^2}{F_b^{(1)}(r)F_b^{(-1)}(r)} = \frac{F_b^{(1)}(s)}{F_b(s)} CV_b^2(s) + \frac{(bF_b(s) - F_b^{(1)}(s))^2}{F_b^{(1)}(s)F_b^{(1)}(s)};$$

equivalently, using (2.2),

$$\frac{F_b^{(2)}(r) - 2bF_b^{(1)}(r) + b^2F_b(r)}{F_b^{(1)}(r)} = \frac{F_b^{(2)}(s) - 2bF_b^{(1)}(s) + b^2F_b(s)}{F_b^{(1)}(s)}. \quad (3.2)$$

If $rs = 1$ then (3.2) reduces to

$$= \frac{s(ss^\ell - 1)((\ell + 1)(\ell + 2)r^2(r^\ell - 1) - 2\ell(\ell + 2)r(rr^\ell - 1) + \ell(\ell + 1)(r^2r^\ell - 1))}{r(rr^\ell - 1)((\ell + 1)(\ell + 2)s^2(s^\ell - 1) - 2\ell(\ell + 2)s(ss^\ell - 1) + \ell(\ell + 1)(s^2s^\ell - 1))}.$$

With, in particular, $\ell = 1$ ($\lambda = 3$), this reduces to

$$s(s + 1)(r - 1)^2 = r(r + 1)(s - 1)^2;$$

for this to hold when $rs = 1$ we need, for arbitrary $r > 0$,

$$\frac{(r-1)^2}{r(r+1)} = \frac{(1-r)^2}{1+r}.$$

Since this is untrue, (3.2), and hence (3.1), fail when $\lambda = 3$. ■

Thus, geometrical stratification fails to exactly minimise the variance. This result is not surprising; the real question is whether it is near the optimum or, more importantly, whether this simple and direct method of assignment of boundaries is more efficient than the commonly-used, but cumbersome, cumulative root method of Dalenius and Hodges [3] when stratifying skewed populations. We examine this question in what follows.

4. A comparison with the cumulative root frequency method of stratum construction

What has become known as the cumulative root frequency method of stratum construction, written $cum\sqrt{f}$, was derived by Dalenius and Hodges [3], who showed that approximations to the points which minimise (1.4) are obtained when the stratum breaks k_h are chosen so that

$$\int_{\beta}^{k_1} \sqrt{f(t)} dt = \int_{k_1}^{k_2} \sqrt{f(t)} dt = \dots = \int_{k_{L-1}}^{\gamma} \sqrt{f(t)} dt ;$$

with $H = \int_{\beta}^{\gamma} \sqrt{f(t)} dt$ the breaks will occur corresponding to the cumulative densities $\frac{H}{L}, \frac{2H}{L}, \dots, \frac{(L-1)H}{L}$. Thus, k_h is the value of x so that

$$\int_{\beta}^x \sqrt{f(t)} dt = \frac{hH}{L}.$$

Cochran [1] applied this algorithm to real populations and illustrated how, for finite data, the approximation is obtained by first dividing the sorted frame into a fairly large number of classes M , counting the number f_j of units within the interval j , $j = 1, 2, \dots, M$. Then, one calculates $\sqrt{f_j}$ and forms strata by joining the adjacent intervals into L groups (strata) in which the $\sum \sqrt{f_j}$ are to be equal or near equal. The main problem with this method is the arbitrariness in deciding the value of M . Cochran [1] cautions that it is advisable to have a substantial number of classes in the original frequency, otherwise the true optimum stratification may be missed and the calculation of the within-stratum boundaries becomes affected by grouping errors. Hedlin [5] notes that the final stratum boundaries depend on the initial choice of the number of classes M , and there is no theory which gives the best number of classes.

Clearly, our geometric method is simpler to use than $cum\sqrt{f}$. Unlike the $cum\sqrt{f}$, this method is definitive and objective, and it does not involve any arbitrary decisions about the initial classes. Of course, we have only demonstrated Proposition 1 for the specific family of Pareto distributions, whereas the cumulative root method

is not so restricted. However, for a more general skewed distribution we can look for a close approximation by one of the Pareto distributions.

We examine the effectiveness of geometric stratification by implementing it on three populations obtained from the set of skewed populations given in Cochran[1], and summarised in Figure 2:

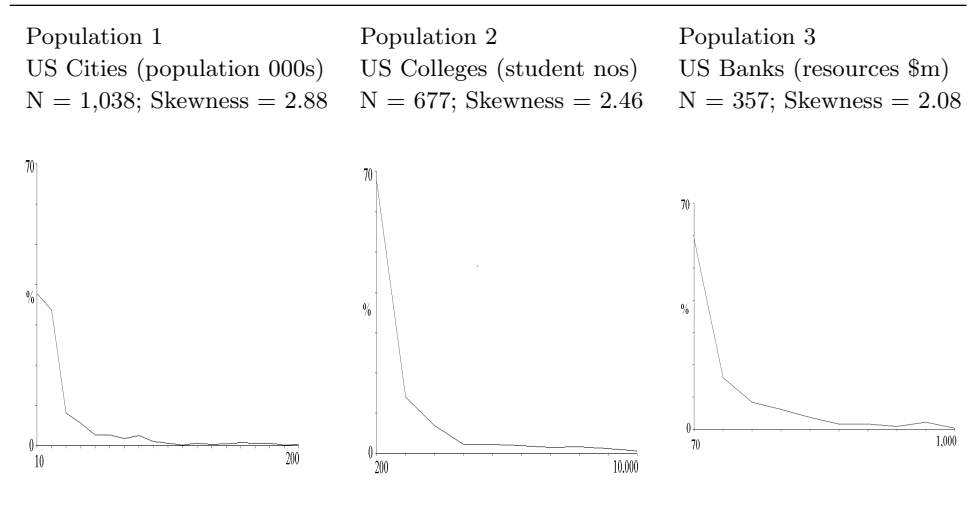


FIG. 2—Some skewed populations

We compare the geometric and cumulative root frequency methods of stratum construction in terms of the relative efficiency or variance ratio:

$$eff = \frac{V_{geom}(\bar{x}_{st})}{V_{cum}(\bar{x}_{st})}, \tag{4.1}$$

where $V_{geom}(\bar{x}_{st})$ and $V_{cum}(\bar{x}_{st})$ are the variances of the mean, respectively, with the geometric and the cumulative root frequency method. The results are given in Table 1 for $L = 4$, and 5 strata.

Table 1—Relative efficiencies

Population	<i>Eff</i>	
	4 Strata	5 Strata
1	0.84	0.59
2	0.86	0.75
3	0.96	0.85

From Table 1 we see that the variance ratio is less than one in all cases, indicating improved efficiency with the geometric stratification method. The efficiency

gains increase with skewness and the number of strata: the greatest gains in efficiency occur with 5 strata in Population 1, the most skewed of the populations; here $eff = 0.59$, which means that, with 5 strata and optimum allocation, geometric stratification achieves the same precision with only 59% of the sample size needed with the cumulative root frequency method.

ACKNOWLEDGEMENT

This work was supported by a grant from the Irish Research Council for Science, Engineering and Technology.

REFERENCES

- [1] W.G. Cochran, Comparison of methods for determining stratum boundaries, *Bulletin of the International Statistical Institute* **37** (1961), 345–56.
- [2] T. Dalenius, The problem of optimum stratification, *Skandinavisk Aktuarietidskrift* **3-4** (1950), 203–13.
- [3] T. Dalenius and J. Hodges, Minimum variance stratification, *Journal of the American Statistical Association* **54** (1959), 88–101.
- [4] M. Evans, N.A.J. Hastings and J.B. Peacock, *Statistical distributions*, 3rd edn, John Wiley & Sons, Chichester, 2000.
- [5] D. Hedlin, A procedure for stratification by an extended Ekman rule, *Journal of Official Statistics* **16** (2000), 15–29.
- [6] J. Neyman, On two different aspects of the representative method: The method of stratified sampling and the method of positive selection, *Journal of the Royal Statistical Society* **97** (1934), 558–606.
- [7] J. Lavallée and M.A. Hidioglou, On the stratification of skewed populations, *Survey Methodology* **14** (1988), 33–43.