Choice of Strata Boundaries for Allocation Proportional to Stratum Cluster Totals in Stratified Cluster Sampling

Bhuwaneshwar Kumar Gupt1,*, F. Lalthlamuanpuii1, Md. Irphan Ahamed2

1Department of Statistics, North-Eastern Hill University, Shillong, 793022, India
2Department of Mathematics, Umshyrpi College, Shillong, 793004, India

Received June 11, 2021; Revised July 30, 2021; Accepted August 22, 2021

Cite This Paper in the following Citation Styles
(a): [1] Bhuwaneshwar Kumar Gupt, F. Lalthlamuanpuii, Md. Irphan Ahamed, "Choice of Strata Boundaries for Allocation Proportional to Stratum Cluster Totals in Stratified Cluster Sampling," Mathematics and Statistics, Vol. 9, No. 5, pp. 697 - 710, 2021. DOI: 10.13189/ms.2021.090509.

(b): Bhuwaneshwar Kumar Gupt, F. Lalthlamuanpuii, Md. Irphan Ahamed (2021). Choice of Strata Boundaries for Allocation Proportional to Stratum Cluster Totals in Stratified Cluster Sampling. Mathematics and Statistics, 9(5), 697 - 710. DOI: 10.13189/ms.2021.090509.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In survey planning, sometimes, there arises situation to use cluster sampling because of nature of spatial relationship between elements of population or physical feature of land over which elements are dispersed or unavailability of reliable list of elements. At the same time, there requires technique and strategy for ensuring precision of the sample in representing the parent population. Although several theoretical cum practical works have been done in cluster sampling, stratified sampling and stratified cluster sampling, so far, the problem of stratified cluster sampling for a study variable based on an auxiliary variable, which is required in practice, has never been approached. For the first time, this paper deals with the problem of optimum stratification of population of clusters in cluster sampling with clusters of equal size of a characteristic $y$ under study based on highly correlated concomitant variable $x$ for allocation proportional to stratum cluster totals under a super population model. Equations giving optimum strata boundaries (OSB) for dividing population, in which sampling unit of the population is a cluster, are obtained by minimising sampling variance of the estimator of population mean. As the equations are implicit in nature, a few methods of finding approximately optimum strata boundaries (AOSB) are deduced from the equations giving OSB. In deriving the equations, mathematical tools of calculus and algebra are used in addition to statistical methods of finding conditional expectation of variance. All the proposed methods of stratification are empirically examined by illustrating in live data, population of villages in Lunglei and Serchhip districts of Mizoram State, India, and found to perform efficiently in stratifying the population. The proposed methods may provide practically feasible solution in planning socio-economic survey.

Keywords Allocation, Gamma Probability Density Function, Cluster Size, Optimum Strata Boundaries, Stratified Cluster Sampling, Stratification Variable

1. Introduction
In stratified sampling, a heterogeneous population is divided into a number of groups called strata which are within strata homogeneous and sample is selected from strata using suitable sample selection method; the method is used for administrative convenience and enhancing the precision of representation of the sample for the parent population. On the other hand, when the availability of reliable list of elements (units) of population is difficult or the elements are spatially dispersed in such a way that there requires lots of energy, time and cost while surveying the elements selected by simple random sampling, cluster sampling or area sampling is employed by grouping the contiguous elements or elements, which can be conveniently surveyed together without much extra effort, into clusters; then, the clusters are taken as
sampling units of population while selecting sample from the population. The strategy used in cluster sampling for enhancing its precision is to make the population within cluster as heterogeneous as possible and increase inter-cluster homogeneity as much as possible. Formation of cluster primarily depends on the spatial relationships between elements in terms of geographical contiguity, good connectivity and convenience in surveying together, less energy, resource and time while surveying the elements within cluster, in addition to scheming for increasing intra-cluster heterogeneity and inter-cluster homogeneity. When the clusters are considered as sampling units of population and then stratified by methods of stratified sampling, the inter cluster homogeneity is increased within strata of clusters which in turn serves the purpose of scheming in cluster sampling for enhancing the precision of representation of sample for the parent population.

In stratified sampling, ever since Dalenius [1] introduced the problem of finding optimum strata boundaries (OBS) based on Tschuprow [2] and Neyman [3] optimum allocation (TNOA) for enhancing homogeneity within strata, the vastness of research in the area has been increasing as a number of researchers, inter alia, Dalenius and Gurney [4], Mahalanobis [5], Hansen et al. [6], Dalenius and Hodges [7,8], Ekman [9], etc., embarked on the work who initially used study variable as stratification variable. As the use of study variable as stratification variable is unrealistic, many workers mostly in the later years extended the work of finding OBS and AOSB by using an auxiliary variable which is highly correlated with the study variable. Dalenius [10], Taga [11], Singh and Sukhatme [12], Singh [13-16], Singh and Prakash [17], Yadava and Singh [18], etc., to mention a few among many, worked on the problem of finding OBS and AOSB based on auxiliary variable for various allocations under different sampling designs. The problem of optimum stratification was again considered from the perspective of more than one study variable by, inter alia, Ghosh [19], Gupta and Seth [20], Rizvi et al. [21,22] etc., whereas Danish and Rizvi [23] approached the problem from the perspective of two auxiliary variables having one study variable.

It is pertinent to mention that in the direction of development of allocation of sample size to strata in stratified sampling, ever since Tschuprow [2] and Neyman [3] proposed TNOA based on study variable, it is Hanurav [24] and Rao [25] who introduced using auxiliary variable under a superpopulation model considered by them. Gupta and Rao [26] obtained allocation of sample size to strata for probability proportional to size under the superpopulation model. Gupta [27,28] modified the aforesaid superpopulation model into a more general form and hence obtained a few generalised model-based allocations; Gupta and Ahamed [29,30] obtained a few methods of stratification for some of the generalised model-based allocations under simple random sampling with and without replacement (SRSWR and SRSWOR) in the form of equations giving OSB and solutions to the equations giving AOSB. Gupt et al. [31] also obtained methods of stratification giving OSB and AOSB for auxiliary variable optimum allocation (AOSB) obtained by Hanurav [24].

In the area of stratified cluster sampling, Mehta and Mandalwara [32] considered problem of finding OBS and AOSB in stratifying population based on study variable for TNOA, proportional and equal allocation under SRSWOR design.

For the first time, we have introduced in this paper the problem of optimum stratification for a characteristic under study $y$ based on a highly correlated auxiliary variable $x$ in stratified cluster sampling with clusters of equal size under the following superpopulation model which is a modified form of the model used by Hanurav [24] and Rao [25].

$$
\begin{align*}
(i) & \xi(y|x) = \alpha + \beta x_i \\
(ii) & V(y|x) = \sigma^2 x_i \\
(iii) & \zeta(y_i, y_j|x_i, x_j) = 0
\end{align*}
$$

(1)

where $\alpha, \beta$ and $\sigma^2$ are the superpopulation parameters with $\sigma^2 > 0$ and the scripts $\xi$, $V$ and $\zeta$ denote conditional expectation, variance and covariance given $x$’s respectively.

Here in this paper, the crux of the work is to simultaneously address the inevitable conditions of spatial relationship of elements leading to the use of cluster sampling and scheming for increasing inter-cluster homogeneity and intra-cluster heterogeneity to increase precision of the sampling.

We use information on the auxiliary variable $x$ which is highly correlated with study variable $y$ to stratify population whose units are clusters whereas clusters are formed by grouping the elements in the way discussed above elaborately; the allocation and sample selection procedure used in this work are allocation proportional to stratum total and SRSWWR, which will hold true for SRSWWR too when finite population correction is neglected.

The paper has six sections. Section 2 deals with obtaining conditional expectation of population variance between cluster means. In section 3, the derivation of equations giving OBS is presented. In section 4, a few methods of finding AOSB are presented. In Section 5, empirical illustration of all the proposed methods of stratification is carried out in live data and results are discussed. Section 6 gives the conclusion.

2. Expression for Conditional Expectation of Population Variance between Clusters Means

Considering a population consists of $N$ clusters of $M$
elements each and a sample of n clusters is to be selected from N clusters by SRSWR. Let $Y_{ij}$ be the value of characteristic under study for the $j^{th}$ element in the $i^{th}$ cluster, $j = 1, 2, \ldots, M$; $i = 1, 2, \ldots, N$. Then, mean square between the cluster means, $\sigma_{by}^2 = \frac{1}{N} \sum_{i=1}^{N} (\bar{Y}_i - \bar{Y})^2$, where $\bar{Y}_i$ and $\bar{Y}$ are the means of the $i^{th}$ cluster and cluster means. $\sigma_{by}^2$ can again be expressed as $\sigma_{by}^2 = \frac{1}{M^2} \sigma_{by}^2$, where $\sigma_{by}^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ is the mean square of cluster totals and $\bar{Y}$ is the mean of cluster totals

$$\Rightarrow \sigma_{by}^2 = \frac{1}{N M^2} \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \bar{Y}_i)^2 - \frac{1}{M^2} \left( \frac{\sum_{i=1}^{N} Y_i}{N} \right)^2$$

Taking conditional expectation of (2) given $x$’s

$$\Rightarrow \sigma_{by}^2 = \frac{1}{N M^2} \left( \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{X}_i)^2 \right)$$

For stratification purpose, we divide the population of N units into L number of strata such that $\sum_{h=1}^{L} N_h = N$ and a sample of size $n_h$ is selected by SRSWR from $h^{th}$ stratum of size $N_h$ such that $\sum_{h=1}^{L} n_h = n$.

Sampling variance of the estimator of the population mean in stratified sampling for the study variable is $V(\bar{y}_n) = \sum_{h=1}^{L} \frac{w_h^2 \sigma^2_{by}[x_h]}{n_h}$, where $w_h = \frac{N_h}{N}$ is the weight of $h^{th}$ stratum.

$$\Rightarrow V(\bar{y}_n) = \sum_{h=1}^{L} \frac{w_h^2 \sigma^2_{by}[x_h]}{n_h}$$

where $\bar{y}_n = \frac{\sum_{h=1}^{L} w_h \bar{y}_h}{n}$ and $\bar{y}_h = \frac{\sum_{i=1}^{N_h} y_{ih}}{N_h}$ are the means of the $h^{th}$ and $i^{th}$ stratum respectively for cluster totals as unit of population, then we can rewrite (5) as

$$\xi(\sigma_{by}^2|x) = \frac{\beta^2}{M^2} \sigma_{bx}^2 + \frac{\sigma^2(N-1)}{NM^2} \mu_{xT}$$

3. Derivation of Methods of Finding Optimum Strata Boundaries

The conditional expectation of $\sigma_{by}^2$ given $x$ in (4) can be expressed as

$$\xi(\sigma_{by}^2|x) = \frac{\beta^2}{M^2} \sigma_{bx}^2 + \frac{\sigma^2(N-1)}{NM^2} \mu_{xT}$$

(6)

For stratification purpose, we divide the population of N units into L number of strata such that $\sum_{h=1}^{L} N_h = N$ and a sample of size $n_h$ is selected by SRSWR from $h^{th}$ stratum of size $N_h$ such that $\sum_{h=1}^{L} n_h = n$.

Sampling variance of the estimator of the population mean in stratified sampling for the study variable is $V(\bar{y}_n) = \sum_{h=1}^{L} \frac{w_h^2 \sigma^2_{by}[x_h]}{n_h}$, where $w_h = \frac{N_h}{N}$ is the weight of $h^{th}$ stratum.

$$\Rightarrow V(\bar{y}_n) = \sum_{h=1}^{L} \frac{w_h^2 \sigma^2_{by}[x_h]}{n_h}$$

where $X = (x_1, x_2, \ldots, x_N)$ and $X' = (x_{h1}, x_{h2}, \ldots, x_{Nh})$.

Since, the allocation proportional to stratum total of the auxiliary variable is considered, we have

$$n_h \propto X_h$$

$$\Rightarrow n_h = \frac{n \mu_h(x)}{X}$$

(8)

where $\mu_h(x)$ is the mean the $h^{th}$ stratum and $X$ is the population mean of the $x$ variable.

From (6), (7) and (8), we get

$$n \xi(V(y_st|x)) = \frac{\beta^2 w_h^2 \sigma_{by}^2}{M^2 \mu_h(x)} + \frac{\sigma^2(N-1)}{NM^2} \frac{\mu_{xT}}{\mu_h(x)}$$

$$\Rightarrow n \xi(V(y_st|x)) = \frac{\beta^2 w_h^2 \sigma_{by}^2}{M^2 \mu_h(x)} + \frac{\sigma^2(N-1)}{NM^2} \frac{\mu_{xT}}{\mu_h(x)}$$

(9)
If \( f(x) \) is probability density function, we can get

\[
W_h = \int_{x_{h-1}}^{x_h} f(t) \, dt
\]

\[
W_h\mu_h(x) = \int_{x_{h-1}}^{x_h} tf(t) \, dt
\]

\[
\{\sigma^2_h + \mu^2_h(x)\}W_h = \int_{x_{h-1}}^{x_h} t^2 f(t) \, dt
\]

Minimising \( n^2 \{Y_k^2 - \mu^2_k\} \) in (9) is equivalent to minimising \( \sum_{h=1}^{L} \frac{W_h \sigma^2_h}{\mu_h(x)} \).

Therefore, we take

\[
\frac{\delta}{\delta x_h} \left( \frac{W_h \sigma^2_h}{\mu_h(x)} \right) = 0.
\]

From (11), (12) and (13), we get

\[
\frac{\delta}{\delta x_h} \left( \frac{W_h \sigma^2_{h+1}}{\mu_{h+1}(x)} \right) = 0.
\]

Considering the first term of (11) and using (10)

\[
\frac{\delta}{\delta x_h} \left( \frac{W_h \sigma^2_h}{\mu_h(x)} \right) = \{x_h - \mu_h(x)\}^2 f(x_h),
\]

and

\[
\frac{\delta}{\delta x_h} \left( \frac{W_h \sigma^2_{h+1}}{\mu_{h+1}(x)} \right) = \{x_h - \mu_{h+1}(x)\} \left[ \mu_{h+1}(x) - \mu_h(x) \right] ^{-2} f(x_h).
\]

Similarly, we can get the second term of (11) as

\[
\frac{\delta}{\delta x_h} \left( \frac{W_h \sigma^2_{h+1}}{\mu_{h+1}(x)} \right) = \{x_h - \mu_{h+1}(x)\} \left[ \mu_{h+1}(x) - \mu_h(x) \right] ^{-2} f(x_h).
\]

From (11), (12) and (13), we get

\[
\{x_h - \mu_h(x)\} \{\mu_h(x) x_h - \mu^2_h(x) \} - \sigma^2_{h+1}
\]

\[
= \{x_h - \mu_{h+1}(x)\} \left[ \mu_{h+1}(x) - \mu_h(x) \right] ^{-2} \sigma^2_{h+1, \mu}.
\]

Equations (14) give OSB for the estimation variable \( y \) based on auxiliary variable \( x \). We call (14) as exact equations.

**4. Derivation of Methods of Finding AOSB Corresponding to the Exact Equations**

**4.1. Approximation Based on Series Expansion**

Since the exact equations (14) are implicit, i.e., equations consist of parameters which are the functions of OSB, it is difficult in solving for OSB in stratifying populations. Therefore, in this section, we carry out derivation for obtaining the solutions of equations (14) which give AOSB. Singh and Sukhatme [12] developed a technique to use Ekman’s [33] identity for obtaining series expansion of conditional mean and variance. Gupta and Ahamed [29,30], Gupta et al. [31] used the technique for obtaining series expansion of conditional mean and variance for some functions. Here, in this paper too, the same technique is used for which we assume the existence of continuity and first three partial derivatives of \( f(x) \) with respect to \( x \), \( \forall x \in (x_{h-1}, x_{h+1}) \) for all the values of \( h \). For expanding, right hand side of equations (14) about the point \( x_h \), we take \( k_{h+1} = x_{h+1} - x_h \) and all the derivatives are evaluated at \( x_h \). Thus series expansion of conditional mean and variance are obtained as follows:

\[
\mu_{(h+1)x} = x_h + \frac{k_{h+1}}{2} f' + \frac{f'' - f'^2}{24f^2} k_{h+1}^3 + \frac{9f'''f^2 - 25f'f'' + 15f'^3}{720f^5} k_{h+1}^5 + O(k_{h+1}^7).
\]

\[
\sigma^2_{(h+1)x} = \frac{k_{h+1}^2}{12} + \frac{2f'' - 5f'r - f''^2}{720f^2} k_{h+1}^4 + O(k_{h+1}^6).
\]

From (15), we get

\[
= x_h x_h^{-1} \left\{ \frac{k_{h+1}}{2x_h} + \frac{f'}{12fx_h} k_{h+1}^2 + \frac{f'' - f'^2}{24f^2 x_h} k_{h+1}^3 \right\}
\]

\[
= \frac{9f'''f^2 - 25f'f'' + 15f'^3}{720f^5} x_h k_{h+1}^5 + O(k_{h+1}^7).
\]

Again from (15) and (16), we get

\[
\sigma^2_{(h+1)x} = \left\{ \frac{k_{h+1}^2}{12} + \frac{2f'' - 5f'r - f''^2}{720f^2} k_{h+1}^4 + O(k_{h+1}^6) \right\}
\]

\[
= x_h x_h^{-1} \left\{ \frac{k_{h+1}}{2x_h} + \frac{f'}{12fx_h} k_{h+1}^2 + \frac{f'' - f'^2}{24f^2 x_h} k_{h+1}^3 \right\}
\]

\[
= \frac{45f'''f^2 x_h^2 - 45f'f'' + 15f'^3}{720f^5} \frac{k_{h+1}^5}{x_h^4} + O(k_{h+1}^7).
\]
Using (15), (17) and (18) in the exact equations (14), we can get as follows:

Right Hand Side of equations (14)

Using (15), (17) and (18) in the exact equations (14), we can rewrite the right hand side of equations (14) as

LHS = \frac{k^2}{4x_h} \left\{ 1 - \frac{\delta (f(x_h))}{f(x_h)} k_{h+1} + O(k_{h+1}^2) \right\}

where

\delta (f(x_h)) = \frac{\delta (f(x_h))}{f(x_h)}

Therefore, equations (14) can be reduced to

\begin{align*}
\text{RHS} &= \frac{k^2}{4x_h} \left\{ 1 + \frac{\delta (f(x_h))}{f(x_h)} k_{h+1} + O(k_{h+1}^2) \right\}
\end{align*}

Similarly, we can obtain the left hand side of equations (14) as

\begin{align*}
\text{LHS} &= \frac{k^2}{4x_h} \left\{ 1 - \frac{\delta (f(x_h))}{f(x_h)} k_{h+1} + O(k_{h+1}^2) \right\}
\end{align*}

Thus, we can rewrite the right hand side of equations (14) as

\begin{align*}
\text{RHS} &= \frac{k^2}{4x_h} \left\{ 1 + \frac{\delta (f(x_h))}{f(x_h)} k_{h+1} + O(k_{h+1}^2) \right\}
\end{align*}

The AOSB corresponding to exact equations (14) are given by the two equivalent methods (21) and (22). The values of constants \( C_1 \) and \( C_2 \) can be approximately
evaluated by $C_1 = \frac{1}{L}(b-a)^2 \int_a^b g(t)dt$ and $C_2 = \frac{1}{L \mu_a} \int_a^b \frac{1}{\sqrt{g(t)}}dt$ respectively, where we assume $b$ and $a$ are upper and lower bounds of points of stratification $x_\ell$’s, i.e., $a \leq x_\ell \leq b$. We can use (21) and (22) in finding AOSB, i.e., $x_\ell$’s by fixing lower boundary $x_{\ell-1}$ every time.

Thus, the above analytical study has led to arrive at the following theorem.

**Theorem 1:** If the function $g(x)$ is bounded and possesses first two partial derivatives for all values of $x$ in $(a, b)$, for a given number of strata, taking equal intervals on the cumulative of $\frac{3}{\sqrt{g(x)}}$ gives AOSB.

### 4.2. Other Approximations Using Assumptions on Coefficient of Variation

Again, we deduce some more methods of approximation from exact equations (14), these approximation methods are still implicit but easy to use. We proceed as follows:

Equations (14) can be rewritten as

$$\{x_h - \mu_h(x)\} \left\{ \frac{x_h}{\mu_h(x)} - 1 - c_h^2 \right\}$$

$$= \{x_h - \mu_{h+1}(x)\} \left\{ \frac{x_h}{\mu_{h+1}(x)} - 1 - c_{h+1}^2 \right\}, \quad (23)$$

where $c_{hx}^2$ and $c_{h+1x}^2$ are the square of coefficients of variation of $h$th and $(h+1)$th strata.

If we consider square of coefficients of variation are approximately equal for every two successive strata, i.e., $c_{hx}^2 \approx c_{h+1x}^2$ and approximately equal to arithmetic mean of square of coefficients of variation of the consecutive strata, then AOSB are given by

$$x_h = \sqrt{\left(1 + \frac{c_{hx}^2}{c_{h+1x}^2}\right) \frac{\mu_h(x)\mu_{h+1}(x)}{2}}, \quad (24)$$

where $c_{(h+1)x}^2 = \frac{c_{hx}^2 c_{h+1x}^2}{2}$

If we consider the square of coefficients of variation for every two successive strata are approximately equal to geometric mean of the square of coefficients of variation the two consecutive strata, AOSB are given by

$$x_h = \sqrt{\left(1 + \frac{\tilde{c}_{(h+1)x}^2}{\tilde{c}_{hx}^2}\right) \frac{\mu_h(x)\mu_{h+1}(x)}{2}}, \quad (25)$$

where $\tilde{c}_{(h+1)x}^2 = \tilde{c}_{hx}^2 c_{h+1x}$

If we consider square of coefficients of variation are negligibly small relative to unity, AOSB are given by

$$x_h = \sqrt{\mu_h(x)\mu_{h+1}(x)}, \quad (26)$$

Thus, we have obtained (24), (25) and (26) as approximations to equations (14) to give AOSB.

### 5. Empirical Illustration

The proposed exact equations (14) and methods of approximation (22), (24), (25) and (26) are illustrated in stratifying population of two districts, Lunglei and Serchhip districts, of Mizoram, a state of India, in which villages are taken as elements of cluster. The data of villages is taken from Village Profile & Development Indicators [34]. There are 193 villages in the two districts. We take number of households of a village as the study variable $y$ and population of village as the auxiliary variable $x$. The correlation between the study variable $y$ and stratification variable $x$ is sufficiently high, i.e., 0.9604. The data is shown in appendix I.

Mizoram is a hilly state of India, 88.93% of the total geographical area is covered by hilly forests; the villages and towns are spread over the hilly terrain of the state. There are rolling hills, tiny valleys, rivers and lakes in the state. Villages are connected by mostly minor and a few major hilly roads. In many cases, the geographical distance between any two villages may be short but they are separated by rivers, lakes, streams, and swamps in the thick rainforest. Therefore, the road transport connecting the two villages may be extremely long requiring lots of energy and time to travel from one to the other. Considering physical feature of the land and pattern of distribution of villages, stratified cluster sampling may be an appropriate sampling design in survey planning. Therefore, we use Google Earth pro and Geographic Information System to locate the villages, rivers, minor roads, major roads, hill ranges and altitudes while forming the clusters. The clusters are formed not only by combining the villages connected by the shortest roads but also taking in account other constraints like variation in altitude and separation by rivers, lakes and thick forest cover. The formation of clusters is shown in appendix II.

In the case of illustrating methods of approximation (21) and (22), since we have theoretically proven both the methods are equivalent, we conveniently use (22). While using (22), we require a Probability Density Function (PDF) that the auxiliary variable $x$ follows. For fitting a suitable distribution, we use the data of $x$ variable in which each value of the variable is divided by 1000. We have two sets of data, one is when each cluster is made of two villages, i.e., cluster size, $M = 2$ and the other is when each cluster is made of three villages, i.e., cluster size, $M = 3$.

We try to find the most suitable PDF that the $x$ variable of the live data follows. The fitdistplus package in R-software is used in fitting a number of known PDFs in data of $x$ variable of both the populations by using the methods - Maximum Likelihood Estimation (MLE), Moment Matching Estimation (MME) and Quantile Matching Estimation (QME) one after another.

Of all the various PDFs we have tried to fit to the data, Gamma Probability Density Function (GPDF) is found to
be fitting best to the data of both populations; the decision of best fitting is made by taking into consideration simultaneously the values of LL (log likelihood), AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria) and standard errors (s.e.) of parameters.

Thus, the PDF followed by the \( x \) variable is as follows:

\[
f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (27)
\]

where \( \alpha > 0, \beta > 0, \forall x \in (0, \infty) \).

The two populations are characterised as follows in fitting GPDF to them.

a. In the data for \( M = 2 \), shape parameter \( \alpha = 2.24436 \), rate parameter \( \beta = 1.49746 \), s.e.(\( \alpha \)) = 0.302965, s.e.(\( \beta \)) = 0.226752, \( LL = -120.354 \), \( AIC = 244.7089 \), \( BIC = 249.8376 \), which are estimated by MLE method.

b. In the data for \( M = 3 \), shape parameter \( \alpha = 3.29280 \), rate parameter \( \beta = 1.464706 \), s.e.(\( \alpha \)) = 0.551335, s.e.(\( \beta \)) = 0.266752, \( LL = -97.53638 \), \( AIC = 199.0728 \), \( BIC = 203.3905 \), which are estimated by MLE method.

We use the above PDF (27) in illustrating approximation method (22), along with the estimates of parameters, in stratifying both the populations. Numerical integration and differentiation methods are used in working out the approximation method in stratifying the populations.

We examine the performance of all the proposed methods of stratification in the stratified cluster sampling design in the two sampling frames. At first, we illustrate the methods in the population in which cluster is formed by the combination of two villages and secondly in the population in which cluster is formed by combination of three villages; the results are shown in Tables 1, 2, 3 and 4. For population of clusters of size two, we present points of stratification due to all the proposed methods of stratification in Table 1 and the variances and relative efficiencies due to the methods in Table 2. Similarly, for population of clusters of size 3, we present the said results in the same way in Tables 3 and 4. Each of the two populations is stratified into two, three, four, five, and six strata by using the all the proposed stratification methods and equal interval stratification. The efficiencies of the proposed methods are compared with that of equal interval stratification in both the populations for each considered number of strata, \( L = 2, 3, 4, 5 \) and 6.

### Table 1. Points of stratification in the population of clusters of size two, \( M = 2 \)

| \( L \) | Equal Interval Stratification | Stratification due to Exact equations (14) | Stratification due to approx. method (22) | Stratification due to approx. method (24) | Stratification due to approx. method (25) | Stratification due to approx. method (26) |
|---|---|---|---|---|---|---|
| 2  | 2606 | 1948.76 | 1670.25 | 1956.62 | 1948.17 | 1576.24 |
| 3  | 1737.33 | 1195.17 | 1054.43 | 1197.12 | 1181.24 | 1138.76 |
|  | 3474.37 | 2562.68 | 2446.52 | 2562.16 | 2551.44 | 2488.94 |
| 4  | 1303 | 1141.11 | 785.83 | 1143.14 | 1134.86 | 938.54 |
|  | 2606 | 2170.56 | 1670.23 | 2170.90 | 2170.72 | 1910.28 |
|  | 3909 | 3714.61 | 2930.7 | 3715.32 | 3714.08 | 3421.17 |
| 5  | 1042.40 | 1037.59 | 633.96 | 1039.08 | 1023.56 | 665.25 |
|  | 2084.80 | 1867.38 | 1286.33 | 1868.23 | 1859.03 | 1212.51 |
|  | 3127.20 | 2871.18 | 2110.77 | 2871.44 | 2870.71 | 2063.53 |
|  | 4169.60 | 4083.96 | 3267.01 | 4084 | 4083.99 | 3513.50 |
| 6  | 868.67 | 685.69 | 535.91 | 763.70 | 680.49 | 665.25 |
|  | 1737.33 | 1224.06 | 1054.43 | 1300.39 | 1224.04 | 1212.51 |
|  | 2606 | 1965.15 | 1670.23 | 2005.12 | 1964.47 | 1952.09 |
|  | 3474.67 | 2901.17 | 2446.52 | 2901.25 | 2901.16 | 2890.63 |
|  | 4343.33 | 4083.96 | 3516.02 | 4084 | 4083.99 | 4071.54 |
At the exact equations (14) perform with higher efficiencies at $L = 2, 5$ and much higher efficiencies at $L = 3, 4$ and 6 than that of equal interval stratification. In the same way, approximation methods (22), (24), (25) and (26) too perform when compared with equal interval stratification. Approximation methods (22) and (26) are found to be relatively better in overall performances than all the other four proposed methods of stratification.

In the population of clusters of size three, the exact equations (14), approximation methods (22), (24), (25) and (26) perform with considerably higher efficiencies than that of equal interval stratification at all the considered number of strata except when $L = 3$ at which all the proposed methods of stratification other than (22) perform with same efficiency with equal interval stratification; method (22) performs with slightly lower efficiency than that of equal interval stratification at $L = 3$. However, in all other number of strata, method (22) performs with strikingly high efficiencies; whereas all

### Table 2. Variance and Relative Efficiencies in the population of clusters of size two, $M = 2$

| $L$ | Equal Interval $nV(\bar{Y}_{st})$ | Exact equations (14) $nV(\bar{Y}_{st})$ | RE | Approx. method (22) $nV(\bar{Y}_{st})$ | RE | Approx. method (24) $nV(\bar{Y}_{st})$ | RE | Approx. method (25) $nV(\bar{Y}_{st})$ | RE | Approx. method (26) $nV(\bar{Y}_{st})$ | RE |
|-----|---------------------------------|-----------------------------------|----|-----------------------------------|----|-----------------------------------|----|-----------------------------------|----|-----------------------------------|----|
| 2   | 3837.94                        | 3481.67                          | 110.23 | 3769.89                          | 101.78 | 3481.67                          | 110.23 | 3481.67                          | 110.23 | 3777.08                          | 101.61 |
| 3   | 2657.10                        | 1747.45                          | 152.06 | 1401.19                          | 189.63 | 1747.45                          | 152.06 | 1673.32                          | 158.79 | 1542.09                          | 172.31 |
| 4   | 1929.32                        | 1303.18                          | 148.05 | 1389.96                          | 138.80 | 1303.18                          | 148.05 | 1303.18                          | 148.05 | 1327.32                          | 145.35 |
| 5   | 1177.71                        | 1110.60                          | 106.04 | 997.14                           | 119.75 | 1135.11                          | 103.75 | 1157.05                          | 101.79 | 917.41                           | 128.37 |
| 6   | 1198.35                        | 904.22                           | 132.53 | 693.87                           | 179.28 | 1030.96                          | 116.24 | 908.45                           | 131.91 | 908.45                           | 131.91 |

### Table 3. Points of stratification in the population of clusters of size two, $M = 3$

| $L$ | Equal Interval Stratification | Points of stratification |
|-----|--------------------------------|---------------------------|
| 2   | 2760.5                         | 2252.81                   |
| 3   | 1840.33                        | 1882.76                   |
| 4   | 1380.25                        | 1281.16                   |
| 5   | 1104.2                         | 1242.71                   |
| 6   | 920.17                         | 1190.62                   |

### Table 4. Variance and Relative Efficiencies in the population of clusters of size three, $M = 3$

| $L$ | Equal Interval $nV(\bar{Y}_{st})$ | Exact equations (14) $nV(\bar{Y}_{st})$ | RE | Approx. method (22) $nV(\bar{Y}_{st})$ | RE | Approx. method (24) $nV(\bar{Y}_{st})$ | RE | Approx. method (25) $nV(\bar{Y}_{st})$ | RE | Approx. method (26) $nV(\bar{Y}_{st})$ | RE |
|-----|---------------------------------|-----------------------------------|----|-----------------------------------|----|-----------------------------------|----|-----------------------------------|----|-----------------------------------|----|
| 2   | 2759.67                        | 2097.53                          | 117.27 | 2097.52                          | 117.27 | 2097.53                          | 117.27 | 2097.53                          | 117.27 | 2097.53                          | 117.27 |
| 3   | 1025.14                        | 1025.14                          | 100  | 1078.06                          | 95.09  | 1025.14                          | 100  | 1025.14                          | 100  | 1025.14                          | 100  |
| 4   | 1081.06                        | 923.48                           | 117.06 | 529.32                           | 204.23 | 923.48                           | 117.06 | 923.48                           | 117.06 | 923.48                           | 117.06 |
| 5   | 675.81                         | 580.56                           | 116.41 | 555.01                           | 121.76 | 580.56                           | 116.41 | 580.56                           | 116.41 | 633.51                           | 106.86 |
| 6   | 678.61                         | 512.77                           | 132.34 | 554.45                           | 122.39 | 533.59                           | 127.18 | 533.59                           | 127.18 | 523.16                           | 129.71 |

It is seen that in the population of clusters of size two, the exact equations (14) perform with higher efficiencies at $L = 2, 5$ and much higher efficiencies at $L = 3, 4$ and 6 than that of equal interval stratification.
other proposed methods of stratification are performing with more or less same efficiency with that of method (22).

But, although the proposed methods of stratification perform well in both the populations, it is found that the methods perform relatively better in the population of clusters of size two than in the population of clusters of size three.

### 6. Conclusions

It is seen in stratified cluster sampling with clusters of equal size, the proposed methods of stratification are performing excellently. The inevitability of the use of cluster sampling due to the nature of spatial relationship between elements of a population or unavailability of proper sampling frame and strategy for ensuring the precision of estimator of population parameter are simultaneously addressed in stratified cluster sampling design presented in this paper. The exact equations (14) and approximation methods (24), (25) and (26) are all performing with more or less same efficiencies, rather interestingly, the approximation method (26), i.e., AOSB are given by the geometric mean of means of consecutive strata, is found to be performing slightly better than other three proposed methods of stratification- (14), (24) and (25). The approximation method (22) which is in the form definite integral of a defined function according to the population used performs best in the overall. Although the methods of approximation (24), (25) and (26) are implicit, they are easy to use. Therefore, all these proposed methods of stratification may be useful in the practical application of survey planning for socio-economic survey.

### Acknowledgements

We, the authors of the article, express our sincere gratitude to the anonymous reviewer for the valuable suggestions provided to us during review.

### Appendix I

Village wise population and number of households, Lunglei and Serchhip districts, Mizoram

| Sl. No. | Name of village | Population | No. of household | Sl. No. | Name of village | Population | No. of household |
|---------|----------------|------------|------------------|---------|----------------|------------|------------------|
| 1       | Thinglian      | 504        | 82               | 62      | Sertlangpui     | 560        | 110              |
| 2       | Thentlang      | 780        | 147              | 63      | Sesawm         | 428        | 89               |
| 3       | Sialhau        | 600        | 109              | 64      | Thenhlum       | 1266       | 250              |
| 4       | Rullam         | 650        | 112              | 65      | Terabonia      | 525        | 104              |
| 5       | Khumtung       | 1340       | 254              | 66      | Tleu           | 107        | 23               |
| 6       | Khawbel        | 757        | 131              | 67      | Tuikawi        | 949        | 217              |
| 7       | Lungpho        | 1005       | 182              | 68      | Belhei         | 631        | 105              |
| 8       | Mualpui Chhingehhip | 1750       | 350              | 69      | Belpei         | 1186       | 262              |
| 9       | Vanchengte     | 126        | 27               | 70      | Balukiasury    | 193        | 37               |
| 10      | Vanchengpui    | 1000       | 170              | 71      | Balungsury     | 356        | 79               |
| 11      | Keitum         | 2150       | 435              | 72      | Bindiasora     | 490        | 115              |
| 12      | Neihloh        | 380        | 72               | 73      | Borsegojasora  | 305        | 72               |
| 13      | Ngentiang      | 676        | 130              | 74      | Bomasury       | 430        | 86               |
| 14      | Hmunzawl       | 538        | 110              | 75      | Chhumkhum      | 212        | 53               |
| 15      | Hmuntha        | 1028       | 160              | 76      | Chhuahdhum     | 291        | 51               |
| 16      | Hmawngkawn     | 165        | 40               | 77      | Chawngte L     | 1053       | 254              |
| 17      | Hualtu         | 1140       | 208              | 78      | Chawilung North| 369        | 65               |
| 18      | Hriangtlang    | 640        | 120              | 79      | Dilibagh       | 1860       | 382              |
| 19      | Bungtlang      | 2263       | 435              | 80      | Gulsil         | 227        | 54               |
| 20      | Baktawng       | 2050       | 336              | 81      | Hunthar        | 186        | 39               |
| 21      | Buangpui       | 485        | 105              | 82      | Kalapani       | 556        | 107              |
|   | Town                  | Pop'n     | Pop'n 1 |   | Pop'n 2 | Pop'n 3 | Pop'n 4 |
|---|-----------------------|-----------|---------|---|---------|---------|---------|
|22 | Chhuanthar Tlangmuam | 2420      | 280     | 83| Kauchhuah| 639     | 146     |
|23 | Chhingchhip           | 2092      | 750     | 84| Khawmawi | 1665    | 297     |
|24 | Sialsir               | 344       | 57      | 85| Khojoysuri| 410     | 82      |
|25 | Sailuak               | 964       | 192     | 86| Letisury  | 386     | 58      |
|26 | Lungchhuan            | 780       | 145     | 87| Lalnutui  | 304     | 64      |
|27 | Khawlailung           | 2672      | 562     | 88| Lamthai I | 354     | 80      |
|28 | Pler                  | 505       | 100     | 89| Lamthai II| 763     | 168     |
|29 | Leng                  | 839       | 180     | 90| Lamthai III| 281     | 51      |
|30 | Lungkawh              | 900       | 175     | 91| Lungsen I | 1255    | 251     |
|31 | Muacheng              | 1510      | 300     | 92| Lungsen II| 1716    | 322     |
|32 | Tuichang              | 27        | 8       | 93| Mariskata | 267     | 55      |
|33 | Bawkflang             | 350       | 74      | 94| Mautlang  | 263     | 56      |
|34 | Chekawn               | 350       | 54      | 95| Nunsury II| 680     | 136     |
|35 | East Lungdar          | 3700      | 735     | 96| Ngiautlang| 204     | 44      |
|36 | Bandisora             | 765       | 137     | 97| New Balukiasury | 378 | 67     |
|37 | Bungmun               | 1232      | 230     | 98| New Khojoysury| 803 | 195     |
|38 | Bungflang             | 255       | 56      | 99| New Vuakmaul| 315 | 62      |
|39 | Changpui              | 457       | 103     | 100| New Lungrang| 750 | 270     |
|40 | Dampui                | 36        | 12      | 101| Nunsury I | 852     | 186     |
|41 | Denssur               | 616       | 127     | 102| Old Khojoysury| 490 | 123     |
|42 | Devasora              | 835       | 145     | 103| Phairuangchhuah| 244 | 48      |
|43 | Hmundo                | 180       | 35      | 104| Phairuangkai| 1200 | 308     |
|44 | Kawlhawk              | 146       | 34      | 105| Putlungasih| 1758 | 250     |
|45 | Lungehmu              | 401       | 85      | 106| Rangte    | 840     | 167     |
|46 | Lokhi Sury            | 256       | 54      | 107| Rolui     | 543     | 95      |
|47 | Laisawral             | 578       | 109     | 108| Rotlang West| 750 | 154     |
|48 | Marpara South         | 2691      | 571     | 109| Rualalung | 380     | 82      |
|49 | Mauzam                | 715       | 143     | 110| S. Lungrang| 332 | 72      |
|50 | Malsury               | 900       | 170     | 111| Sailen    | 132     | 30      |
|51 | New Sachan            | 114       | 222     | 112| Samuksury | 975     | 214     |
|52 | New Khawleik          | 164       | 33      | 113| Sedailui  | 220     | 68      |
|53 | Puankhai              | 1182      | 229     | 114| Silkur    | 345     | 73      |
|54 | Phainuam              | 218       | 44      | 115| Siiphirtlang| 126 | 28      |
|55 | South Khawleik        | 140       | 30      | 116| Sugorbasora| 537 | 111     |
|56 | Sumasumui             | 487       | 103     | 117| Tipperagath I | 1171 | 202     |
|57 | Salmsur               | 272       | 55      | 118| Tipperaghat II| 636 | 122     |
|58 | Saisei                | 255       | 55      | 119| Tipperaghat III| 633 | 118     |
|59 | Sachan                | 600       | 112     | 120| Thanzamasora| 385 | 75      |
|60 | S. Lungdai           | 280       | 53      | 121| Thekaduar  | 347     | 75      |
|61 | Serte                 | 446       | 89      | 122| Tuichawngchhuah| 221 | 49      |
### Table

| Sl. No. | Name of village     | Population | No. of household | Sl. No. | Name of village     | Population | No. of household |
|---------|---------------------|------------|------------------|---------|---------------------|------------|------------------|
| 123     | Tuichawng           | 3419       | 583              | 124     | Tuisenchhuah        | 330        | 85               |
| 125     | Tuisen Bolia        | 793        | 153              | 126     | Undermanik          | 777        | 106              |
| 127     | Ugudasury           | 491        | 109              | 128     | Vuakmual            | 6          | 1                |
| 129     | Vairawkai           | 185        | 35               | 130     | Zehntet             | 596        | 118              |
| 131     | Zawlpui             | 1766       | 352              | 132     | Zohmun              | 487        | 96               |
| 133     | Serthuan            | 597        | 135              | 134     | Buknuam             | 393        | 81               |
| 135     | Buarpui             | 1450       | 360              | 136     | Bualpui V           | 720        | 112              |
| 137     | Bualte              | 446        | 96               | 138     | Chengpui            | 180        | 38               |
| 139     | Chithar             | 220        | 52               | 140     | Chhipphir           | 1411       | 287              |
| 141     | Dawn                | 430        | 72               | 142     | Haulawng            | 2245       | 527              |
| 143     | Hlumte              | 210        | 37               | 144     | Hmuntlang           | 122        | 24               |
| 145     | Kanghmun South      | 537        | 135              | 146     | Mamte               | 852        | 127              |
| 147     | Mausen              | 300        | 70               | 148     | Mualcheng South     | 960        | 167              |
| 149     | Mualthuam North     | 1519       | 330              | 150     | Lungmawi            | 272        | 52               |
| 151     | Pachang             | 39         | 13               | 152     | Phaileng South      | 330        | 63               |
| 153     | Ramlahtui           | 657        | 158              | 154     | Ralvawng            | 520        | 76               |
| 155     | Runung              | 211        | 34               | 156     | Sairep              | 264        | 52               |
| 157     | Sekhum              | 384        | 85               | 158     | Thlengang           | 240        | 43               |
Appendix II

Physical Map of Lunglei and Serchhip District, Mizoram
(Formation of clusters of size 2 and size 3)
REFERENCES

[1] T. Dalenius. The problem of optimum stratification- I. Skandinavisk Aktuarietidskrift, Vol. 33,203-213, 1950.
[2] A.A. Tschuprow. On mathematical expectation of the moments of frequency distributions in the case of correlated observations. Metron, Vol.2, 461-493, 1923.
[3] J. Neyman. On 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, No.4, 558-625, 1934.
[4] T. Dalenius, M. Gurney. The problem of optimum stratification-II. Skandinavisk Aktuarietidskrift, Vol. 34, 133-148, 1951.
[5] P.C. Mahalanobis. Some aspects of the design of sample surveys. Sankhya, Vol. 12, 1-7, 1952.
[6] M.H. Hansen, W.N. Hurwitz, W.G. Madow. Sample survey methods and theory, Vol. I and II, New York: John Wiley and Sons, Inc, 1953.
[7] T. Dalenius, J.L. Hodges. The Choice of stratification points, Skandinavisk Aktuarietidskrift, Vol. 34, 198-203, 1957.
[8] T. Dalenius, J.L. Hodges. Minimum variance stratification. Jour. Amer. Stat. Assoc., Vol. 54, 88-101, 1959.
[9] G. Ekmak. An approximation useful in univariate stratification. The Annals of Mathematical Statistics, Vol. 30, 219-229, 1959.
[10] T. Dalenius. Sampling in Sweden. Contributions to the methods and theories of sample survey practice. Almqvist and Wikells, Stockholm, 1957.
[11] Y. Taga. On optimum stratification for the objective variable based on concomittant variable. Annals of the Institute of Statistical Mathematics, Vol. 19,101-130, 1967.
[12] R. Singh, B.V. Sukhatme. Optimum stratification. Annals of the Institute of Statistical Mathematics, Vol. 21, 515-528, 1969.
[13] R. Singh. Approximately optimum stratification on the auxiliary variable. Journal of the American Statistical Association, Vol. 66, 829-833, 1971.
[14] R. Singh. An alternate method of stratification on the auxiliary variable. Sankhya, Vol. 37, 100-108, 1975.
[15] R. Singh. On optimum stratification for proportional allocation. Sankhya, Vol. 37, Pt. I, 109-115, 1975.
[16] R. Singh. A note on optimum stratification in sampling with varying probabilities. Australian Journal of Statistics, Vol. 27, No. 1, 12-21, 1975.
[17] R. Singh, D. Prakash. Optimum stratification for equal allocation. Annals of the Institute of Statistical Mathematics, Vol. 27, 273-280, 1975.
[18] S.S. Yadava, R. Singh. Optimum stratification for allocation proportional to strata totals for simple random sampling scheme. Communications in Statistics: Theory and Methods, Vol. 13, No. 22, 2793-2806, 1984.
[19] S.P. Gosh. Optimum stratification with two characters. Ann. Math. Statist., Vol. 34, 866-872, 1963.
[20] P.C. Gupta, G.R. Seth. On stratification in sampling investigation involving more than one character. J. Ind. Soc. Agril. Statist., Vol. 31, No. 2, 1-15, 1979.
[21] S.E.H. Rizvi, J.P. Gupta, R. Singh. Approximately optimum stratification for two study variables using auxiliary information. J. Ind. Soc. Agril. Statist., Vol. 53, No. 3, 287-298, 2000.
[22] S.E.H. Rizvi, J.P. Gupta, M. Bhargava. Effect of optimum stratification on sampling with varying probabilities under proportional allocation. Statistica, Vol. 64, No. 4, 721-733, 2004.
[23] F. Danish, S.E.H. Rizvi. Optimum stratification in bivariate auxiliary variables under Neyman allocation. Journal of Modern Applied Statistical Methods, Vol. 17, No. 1, 1-24, 2018. doi: 10.22237/jmsm/1529418671.
[24] T.V. Hanurav. Optimum Sampling Strategies and Some Related Problems, unpublished Ph.D thesis submitted to the Indian Statistical Institute, 1965.
[25] T.J. Rao. On the allocation of sample size in stratified sampling. Ann. Inst. Stat. Math, Vol. 20, 159-166, 1968.
[26] B.K. Gupt, T.J. Rao. Stratified PPS sampling and allocation of sample size. Jour. Ind. Soc. Agri. Statistics, Vol. 50, No.2, 199-208, 1997.
[27] B.K. Gupt. Sample size allocation for stratified sampling under a correlated superpopulation model. METRON-International Journal of Statistics, Vol. LXI, No. 1, 35-52, 2003.
[28] B.K. Gupt. Allocation of Sample Size in Stratified Sampling Under Superpopulation Models. Saarbrucken, Deutschland/Germany: LAP LAMBERT Academic Publishing Publishing AV Akademikerverlag GmbH & Co.KG, 2012.
[29] B.K. Gupt, M.I. Ahamed. Optimum stratification for a generalized auxiliary variable proportional allocation under a superpopulation model. Communications in Statistics-Theory and Methods, Published online: 25 July 2020. https://doi.org/10.1080/03610922.2020.1793203.
[30] B.K. Gupt, M.I. Ahamed. Construction of strata for a model-based allocation under a superpopulation model, Journal of Statistical Theory and Applications, Vol. 20, No. 1, 46-60, 2021.
[31] B.K. Gupt, M.I. Ahamed, M. Phukon. Optimum stratification for an auxiliary variable optimum allocation under a superpopulation model, Advances and Applications in Statistics, Vol. 67, No.1, 1-20, 2021.
[32] S. Mehta, V.L. Mandowara. An optimum stratification for stratified cluster sampling design when clusters are of varying sizes. International Journal of Scientific & Technology Research, Vol. 1, No. 9, 74-79, 2012.
[33] G. Ekman. Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution. *The Annals of Mathematical Statistics*, Vol. 30, No. 4, pp. 1131-1134, 1959b.

[34] Village profile & Development Indicators 2017-2018, Mizoram state’ published by Directorate of Economics and Statistics Government of Mizoram, Mizoram: Aizawl. https://des.mizoram.gov.in/page/village-profile-2017-2018-mizoram-state