https://doi.org/10.56093/JISAS.V79I2.10
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 79(2) 2025 149–153

Generalisation of Das & Mohanty Sampling Scheme

Padam Singh
Medanta Institute of Education and Research, Gurugram

Received 16 September 2025; Revised 18 November 2025; Accepted 02 December 2025

SUMMARY

This paper revisits and extends the Das & Mohanty sampling scheme, which integrates Probability Proportional to Size (PPS) systematic sampling with Simple Random Sampling Without Replacement (SRSWOR). After outlining the foundations of PPS sampling by Hansen and Hurwitz (1943) and the Horvitz–Thompson (1952) framework, the study proposes a generalised scheme based on a Bernoulli trial with weight *w*, allowing controlled mixing of PPS and SRSWOR. Analytical derivations of inclusion probabilities and variances are presented, with empirical illustrations across three test populations. The findings indicate that the proposed scheme approximates πPS sampling for small *w*, yielding reduced variance in certain cases (notably for populations B and C), though slightly less efficient for population A. This highlights the adaptability of the generalised approach and its potential to outperform classical estimators under specific population structures where auxiliary variables are strongly correlated with the study variable.

Keywords: Probability proportional to size (PPS); Horvitz–Thompson estimator; Simple random sampling (SRS); PPS systematic sampling,; Das & Mohanty scheme; Inclusion probability; Sampling variance; Generalised sampling scheme.

1. INTRODUCTION

The concept of sampling with probabilities proportional to size (PPS) was first introduced by **Hansen and Hurwitz (1943)**. In this approach, the chance of selecting a unit depends on a pre-assigned *size measure* (e.g., Xi for the i^{th} unit).

The selection of sample by PPS Sampling is based on Cumulative Total Method as under:

For a finite population of N units with size measures $x_1, x_2, \ldots\ldots\ldots x_n$ and total

$$C_N = \sum_{i=1}^{N} X_i,$$

a PPS sample is selected through the following steps:

1. Compute cumulative totals for all population units.
2. Generate a random number $r \in [1, C_N]$
3. Select unit i if $C_{i-1} < r \le C_i$, where C_i is the cumulative total up to unit i.

Estimation under Sampling with Replacement

If sampling is performed with replacement, the probability of selecting unit I in each draw is

$$p_i = \frac{X_i}{C_N}.$$

The Hansen–Hurwitz estimator of the population total Y is given by:

$$\hat{Y}_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i},$$

with variance

$$V(\hat{Y}_{HH}) = \frac{1}{n}\left(\sum_{i=1}^{N} p_i\left(\frac{y_i}{p_i}\right)^2 - Y^2\right).$$

E-mail address: dr.padamsingh2013@gmail.com

2. HORVITZ–THOMPSON GENERALISATION (1952)

Horvitz and Thompson extended the estimation framework to *unequal probability sampling without replacement*. For a sample of size n:

- Π_i = first-order inclusion probability of unit i.
- Π_{ij} = joint inclusion probability of units i and j.

The **Horvitz–Thompson estimator** of the population total is:

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i},$$

with variance of the estimator:

$$V(\hat{Y}_{HT}) = \sum_{i<j} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_i}{\pi_j} \right)^2$$

3. PPS SYSTEMATIC SAMPLING

In practice, **PPS systematic sampling** is widely adopted for *without-replacement* designs. For a required sample size n, the procedure is:

1. Arrange the population units with cumulative totals of their size measures.
2. Define the sampling interval as
   $$k = \frac{C_N}{n}.$$
3. Select a random start $r \in [1,k]$.
4. The selected positions are:
   r, r+k, r+2k, …, r+(n−1)k.
5. Units whose cumulative totals cover these positions are included in the sample.

The joint inclusion probability for all pairs of units is not non zero which is the necessary condition for unbiased estimation of sampling variance hence the limitation.

4. DAS & MOHANTY SAMPLING SCHEME AND ITS GENERALISATION

The **Das & Mohanty sampling scheme** represents a judicious combination of two approaches:

- **PPS (Probability Proportional to Size) systematic sampling**, and
- **SRSWOR (Simple Random Sampling Without Replacement).**

It merges the sample spaces of both methods. For instance, if the sample size is 2 from a population of 4, the sample spaces are:

- **Under SRSWOR:** (1,2), (1,3), (1,4), (2,3), (2,4), (3,4)
- **Under PPS systematic sampling:** (1,3), (2,4), (2,4), (3,4), (3,4)

The frequency of occurrence of individual and pair wise units under the sample space are summarised as under:

Frequency of occurrence	In SRSWOR	In PPS System	Total
1	3	1	4
2	3	2	5
3	3	3	6
4	3	4	7
Combinations, 1 with 2	1	0	1
Combinations, 1 with 3	1	1	2
Combinations, 1 with 4	1	0	1
Combinations, 2 with 3	1	0	1
Combinations, 2 with 4	1	2	3
Combinations, 3 with 4	1	2	3

In general, suppose the sample space under SRS is b_1 and that under PRS systematic sampling as b_2. Then sampling scheme by Das & Mohanty is as under:

Select a random number between 1 and $(b_1 + b_2)$.

If random number selected r is less than or equal to b_1, the sample to be selected by SRS or else by PRS Systematic sampling.

They derived the expressions for inclusion probabilities for individual units and pair wise units under the proposed scheme. Importantly, they suggested computation of revised sizes for ensuring inclusion probabilities proportional to size. This contribution is considered as a landmark in developing the theory for sampling with unequal probabilities without replacement.

5. PROPOSED GENERALISED SAMPLING SCHEME

Das & Mohanty's scheme can be viewed as a **special case of a more general approach**. In the generalised scheme:

1. Perform a **Bernoulli trial** with success probability **w**.

- If success: draw the sample by **SRSWOR**.
- If failure: draw the sample by **PPS systematic sampling**.

2. The **inclusion probability** for unit *i* becomes:

$$\pi_i = w \cdot \frac{n}{N} + (1 - w) \cdot \frac{nx_i}{T}$$

Where

- n = sample size
- N = population size
- x_i = measure of size for unit 'i'
- $T=\sum x_i$

3. To align with a **πPS scheme**, revised sizes can be computed as:

$$x_i = T \cdot \frac{\pi_i - \frac{w}{N}}{1 - w}$$

For small values of w, these revised x_i remain non-negative.

Note: Das & Mohanty's scheme corresponds to the particular case where

$w = \frac{b_2}{b_1 + b_2}$ with b_1 and $b_2$representing the respective sample space sizes.

Illustration with Example

Computation of inclusion probabilities (individual and joint) as well as variance of the estimate are illustrated as under:

Three test populations (A, B, C) are often used to compare efficiencies. For w = 0.05, the scheme is dominated by PPS (95%), with only 5% weight to SRS.

The three populations which have generally considered for comparison of efficiencies of different estimates is as under:

Table 1. Populations under study

Sr. No.	Size X	Study variable Y		
		A	B	C
1	1	0.5	0.8	0.2
2	2	1.2	1.4	0.6
3	3	2.1	1.8	0.9
4	4	3.2	2	0.8

For these populations the correlation Coefficient along with regression equation are as under:

Population	Regression Equation	R^2	Correlation (r)
A	Y = -0.5 + 0.90X	**0.9902**	**0.9951**
B	Y = 0.5 + 0.40X	**0.9524**	**0.9759**
C	Y = 0.1 + 0.21X	**0.7670**	**0.8758**

The value of inclusion probability for different units is given in the table below:

Table 2. Inclusion probabilities of proposed scheme

Sr. No.	SRS		PPS		Overall inclusion probability
	w	Inclusion probability	1-w	Inclusion probability	
1	0.05	0.5	0.95	0.2	0.215
2	0.05	0.5	0.95	0.4	0.405
3	0.05	0.5	0.95	0.6	0.595
4	0.05	0.5	0.95	0.8	0.785

Thus, the scheme is nearly πps for smaller values of w. The computation of πij's is illustrated in table 3.

Table 3. Computation of πij's

Pair of units	πij's under SRS	πij's under PPS	w	1-w	w*πij's under SRS	(1-w) *πij's under PPS	Overall πij's w*πij's under SRS + (1-w)*πij's under PPS
(1,2)	0.167	0	0.05	0.95	0.00835	0	0.00835
(1,3)	0.167	0.2	0.05	0.95	0.00835	0.19	0.19835
(1,4)	0.167	0	0.05	0.95	0.00835	0	0.00835
(2,3)	0.167	0	0.05	0.95	0.00835	0	0.00835
(2,4)	0.167	0.4	0.05	0.95	0.00835	0.38	0.38835
(3,4)	0.167	0.4	0.05	0.95	0.00835	0.38	0.38835

Table 4 below presents the comparison of variance of proposed scheme with others

Table 4. Comparison of different sampling schemes

Sampling Scheme	Variance of estimate		
	Population A	Population B	Population C
Sampling with replacement	0.5	0.5	0.125
Des raj order estimate	0.365	0.365	0.088
Murthy Unordered estimate	0.312	0.312	0.07
Hartley and Rao method of sampling	0.367	0.367	0.033
Rao, Hartley and Cochran method of sampling	0.333	0.333	0.083
Das and Mohanty sampling scheme	0.365	0.365	0.033
Proposed sampling scheme	0.432	0.195	0.005

Note: the figures for other sampling schemes are taken from Das & Mohanty

Interpretation

- In the paper by **Das & Mohanty scheme**, variances for Populations A and B are identical for schemes considered here.
- The **proposed generalisation**, with small w, behaves nearly like πPS sampling.
- Empirical findings suggest:
    - Performance is slightly inferior for **Population A**,
    - But **superior for Populations B and C**, where the variance is substantially reduced.

Hence, while no sweeping generalisation can be made, the proposed scheme demonstrates **improved efficiency in certain population structures**.

Choice of weights:

The optimal weight (**w**) cannot be derived mathematically. For the proposed scheme, the joint inclusion probability is expressed as:

$$\pi_{ij} = w\pi_{ij}^{(\mathrm{PPS})} + (1-w)\pi_{ij}^{(\mathrm{SRS})}$$

where

$$\pi_{ij}^{(\mathrm{PPS})} = \begin{cases} \dfrac{n}{T}, & \text{for units separated by a distance of } T/n, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\pi_{ij}^{(\mathrm{SRS})} = \frac{n(n-1)}{N(N-1)}.$$

Therefore, in the exactπPS scheme with the revised population size, the weight **w** may be chosen as large as allowed within the permissible range. Intuitively, selecting a larger **w** increases the joint inclusion probabilities for those pairs of units that would otherwise have zero probability under PPS systematic sampling.

REFERENCES

Hansen M.H., and Hurwitz W.N. On the theory of sampling from finite populations. *Annals of Mathematical Statistics*. 1943; **14**(4):333-62.

Horvitz D.G., and Thompson D.J. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*. 1952; **47**(260):663-85.

Des Raj. A representation of the sampling variance of an ordered estimator. *Sankhyā: The Indian Journal of Statistics, Series A*. 1956; **17**(1):63-72.

Murthy M.N. Ordered and unordered estimators in sampling without replacement. *Sankhyā: The Indian Journal of Statistics, Series A*. 1957; **18**(3):379-90.

Hartley H.O. and Rao J.N.K. Sampling with unequal probabilities and without replacement. *Annals of Mathematical Statistics*. 1962; **33**(2):350-74.

Rao J.N.K, Hartley H.O., Cochran W.G. On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*. 1962; **24**(2):482-91.

Das M.N., Mohanty S. On PPS Sampling without replacement ensuring selection probabilities exactly proportional to sizes.*Australian Journal of Statistics*. 1973; **15**(2):87-94.

APPENDIX

The computation of variance under proposed sampling scheme is summarised in tables A1-A3 below for different populations:

Table A1. Computation of variance for population A

Yi	πi	Yi/πi
0.5	0.215	2.325581395
1.2	0.405	2.962962963
2.1	0.595	3.529411765
3.2	0.785	4.076433121

Table A1 (Contd). Computation of variance for population A

Pair of units	Yi/πi-Yj/πj	(Yi/πi-Yj/πj)^2	πiπj	πiπj - πij	(πiπj - πij) * ((Yi/πi-Yj/πj)^2
(1,2)	-0.637381568	0.406255263	0.087075	0.078725	0.031982446
(1,3)	-1.203830369	1.449207558	0.127925	-0.070425	-0.102060442
(1,4)	-1.750851726	3.065481765	0.168775	0.160425	0.491779912
(2,3)	-0.566448802	0.320864245	0.240975	0.232625	0.074641045
(2,4)	-1.113470158	1.239815793	0.317925	-0.070425	-0.087314027
(3,4)	-0.547021356	0.299232364	0.467075	0.078725	0.023557068
Total					0.432586001

Table A2. Computation of variance for population B

Yi	πi	Yi/πi
0.8	0.215	3.720930233
1.4	0.405	3.456790123
1.8	0.595	3.025210084
2	0.785	2.547770701

Table A2 (Contd). Computation of variance for population B

Pair of units	Yi/πi-Yj/πj	$(Yi/πi-Yj/πj)^2$	πiπj	πiπj - πij	$(πiπj - πij) * ((Yi/πi-Yj/πj)^2$
(1,2)					
(1,3)	0.695720149	0.484026525	0.127925	-0.070425	-0.034087568
(1,4)	1.173159532	1.376303287	0.168775	0.160425	0.220793455
(2,3)	0.431580039	0.18626133	0.240975	0.232625	0.043329042
(2,4)	0.909019423	0.826316311	0.317925	-0.070425	-0.058193326
(3,4)	0.477439383	0.227948365	0.467075	0.078725	0.017945235
					0.195279481

Table A3. Computation of variance for population C

Yi	πi	Yi/πi
0.2	0.215	0.930232558
0.6	0.405	1.481481481
0.9	0.595	1.512605042
0.8	0.785	1.01910828

Table A3 (Contd). Computation of variance for population C

Pair of units	Yi/πi-Yj/πj	$(Yi/πi-Yj/πj)^2$	πiπj	πiπj - πij	$(πiπj - πij) * ((Yi/πi-Yj/πj)^2$
(1,2)					
(1,3)	-0.582372484	0.33915771	0.127925	-0.070425	-0.023885182
(1,4)	-0.088875722	0.007898894	0.168775	0.160425	0.00126718
(2,3)	-0.031123561	0.000968676	0.240975	0.232625	0.000225338
(2,4)	0.462373201	0.213788977	0.317925	-0.070425	-0.015056089
(3,4)	0.493496762	0.243539054	0.467075	0.078725	0.019172612
					0.005646449