Introduction to Sampling Theory

Lecture 25
Varying Probability Sampling

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

http://home.iitk.ac.in/~shalab/sp

Methods of Sample Selection in Varying Probability Scheme:

In pps sampling, there are two possibilities to draw the sample, i.e., with replacement and without replacement.

1. PPSWR

2. PPSWOR

Selection of Units With Replacement:

The probability of selection of a unit will not change and the probability of selecting a specified unit is the same at any stage. There is no redistribution of the probabilities after a draw.

Selection of Units Without Replacement:

The probability of selection of a unit will change at any stage and the probabilities are redistributed after each draw.

PPS without replacement (WOR) is more complex than PPS with replacement (WR) . We consider both the cases separately.

PPS sampling with replacement (PPSWR) :

First we discuss the two methods to draw a sample with PPS and WR.

1. Cumulative Total Method:

The procedure of selection of a simple random sample of size *n* consists of

- **associating the natural numbers from 1 to *N* units in the population and**
- **then selecting those *n* units whose serial numbers correspond to a set of *n* numbers where each number is less than or equal to *N* which is drawn from a random number table.**

1. Cumulative Total Method:

In selection of a sample with varying probabilities, the procedure is to associate with each unit a set of consecutive natural numbers, the size of the set being proportional to the desired probability.

If $X_1, X_2, \ldots, X_N$ **are the positive integers proportional to the probabilities assigned to the** ***N*** **units in the population, then a possible way to associate the cumulative totals of the units.**

Then the units are selected based on the values of cumulative totals.

This is illustrated in the following table:

1. Cumulative Total Method:

Units	Size	Cumulative sizes		
1	X_1	$T_1 = X_1$	**Select a random number *R* between 1 and T_N by using random number table.**	• **If $T_{i-1} \leq R \leq T_i$, then i^{th} unit is selected with probability $\frac{X_i}{T_N}$, *i* = 1,2,…, *N*.** • **Repeat the procedure *n* times to get a sample of size *n*.**
2	X_2	$T_2 = X_1 + X_2$		
$\vdots$	$\vdots$	$\vdots$		
$i-1$	X_{i-1}	$T_{i-1} = \sum_{j=1}^{i-1} X_j$		
i	X_i	$T_i = \sum_{j=1}^{i} X_j$		
$\vdots$	$\vdots$	$\vdots$		
N	X_N	$T_N = \sum_{j=1}^{N} X_j$		

1. Cumulative Total Method:

In this case, the probability of selection of *i^{th}* unit is

$$P_i = \frac{T_i - T_{i-1}}{T_N} = \frac{X_i}{T_N}$$

$$\Rightarrow P_i \propto X_i.$$

Note that T_N is the population total which remains constant.

1. Cumulative Total Method: Drawback

This procedure involves writing down the successive cumulative totals. This is time consuming and tedious if the number of units in the population is large.

This problem is overcome in the Lahiri's method.

Lahiri's Method:

Let $M = \underset{i=1,2,...,N}{Max} \; X_i,$ **i.e., maximum of the sizes of *N* units in the population or some convenient number greater than *M*.**

The sampling procedure has following steps:

1. **Select a pair of random number (*i, j*) such that** $1 \le i \le N, \; 1 \le j \le M.$

2. **If** $j \le X_i,$ **then *i*** th **unit is selected otherwise rejected and another pair of random number is chosen.**

3. **To get a sample of size *n*, this procedure is repeated till *n* units are selected.**

Lahiri's Method:

Now we see how this method ensures that the probabilities of selection of units are varying and are proportional to the size.

Probability of selection of i^{th} unit at a trial depends on the following two possible outcomes:

- **either it is selected at the first draw**
- **or it is selected in the subsequent draws preceded by ineffective draws.**

Lahiri's Method:

Such probability is given by

$$P(1 \le i \le N)P(1 \le j \le M \mid i) = \frac{1}{N}.\frac{X_i}{M} = P_i^*, \text{ say.}$$

Probability that no unit is selected at a trial $= \dfrac{1}{N}\sum_{i=1}^{N}\left(1 - \dfrac{X_i}{M}\right)$

$$= \frac{1}{N}\left(N - \frac{N\bar{X}}{M}\right)$$

$$= 1 - \frac{\bar{X}}{M} = Q, \text{ say.}$$

Lahiri's Method:

The probability that unit *i* is selected at a given draw (all other previous draws result in the non selection of unit *i*)

$$= P_i^* + QP_i^* + Q^2 P_i^* + \ldots$$

$$= \frac{P_i^*}{1-Q}$$

$$= \frac{X_i / NM}{\bar{X} / M} = \frac{X_i}{N\bar{X}} = \frac{X_i}{X_{total}} \propto X_i.$$

Thus the probability of selection of unit *i* is proportional to the size X_i. So this method generates a pps sample.

Lahiri's Method: Advantages

1. It does not require writing down all cumulative totals for each unit.

2. Sizes of all the units need not be known before hand.

   We need only some number greater than the maximum size and the sizes of those units which are selected by the choice of the first set of random numbers 1 to N for drawing sample under this scheme.

Lahiri's Method: Disadvantage

It results in the wastage of time and efforts if units get rejected.

A draw is ineffective if one of the ineffective random numbers is selected.

Sample Selection: Example

Consider the following data set of number of workers in 10 factories and its output. We illustrate the selection of units using the cumulative total method.

Sample Selection: Example

Factory no.	Number of workers (X) (in thousands)	Industrial production (in metric tons) (Y)	Cumulative total of sizes
1	2	30	$T_1 = 2$
2	5	60	$T_2 = 2+5 = 7$
3	10	12	$T_3 = 2+5+10 = 17$
4	4	6	$T_4 = 17+4 = 21$
5	7	8	$T_5 = 21+7 = 28$
6	12	13	$T_6 = 28+12 = 30$
7	3	4	$T_7 = 30+3 = 33$
8	14	17	$T_8 = 33+14 = 47$
9	11	13	$T_9 = 47+11 = 58$
10	6	8	$T_{10} = 58+6 = 64$

Selection of Sample using Cumulative Total Method:

1. First draw:

Draw a random number between 1 and 64.

- **Suppose it is 23.**
- $T_4 < 23 < T_5$
- **Unit** Y **is selected and** $Y_5 = 8$ **enters in the sample.**

Selection of Sample using Cumulative Total Method:

2. Second draw:

- **Draw a random number between 1 and 64.**
- **Suppose it is 38.**
- $T_7 < 38 < T_8$
- **Unit 8 is selected and Y_8 = 17 enters in the sample and so on.**
- **This procedure is repeated till the sample of required size is obtained.**

Sample Selection: Example

Factory no.	Number of workers (X) (in thousands)	Industrial production (in metric tons) (Y)	Cumulative total of sizes
1	2	30	$T_1 = 2$
2	5	60	$T_2 = 2 + 5 = 7$
3	10	12	$T_3 = 2 + 5 + 10 = 17$
4	4	6	$T_4 = 17 + 4 = 21$
5	7	8	$T_5 = 21 + 7 = 28$
6	12	13	$T_6 = 28 + 12 = 30$
7	3	4	$T_7 = 30 + 3 = 33$
8	14	17	$T_8 = 33 + 14 = 47$
9	11	13	$T_9 = 47 + 11 = 58$
10	6	8	$T_{10} = 58 + 6 = 64$

Selection of Sample using Lahiri's method:

In this case

$$M = \underset{i=1,2,...,10}{Max} X_i = 14$$

So we need to select a pair of random number (*i, j*) such that

$$1 \le i \le 10,\ 1 \le j \le 14$$

Following table shows the sample obtained by Lahiri's scheme:

Selection of Sample using Lahiri's Method:

Random no $1 \le i \le 10$	Random no $1 \le j \le 14$	Observation	Selection of unit
3	7	$j = 7 < X_3 = 10$	trial accepted (y_3)
8	13	$j = 13 < X_8 = 14$	trial accepted (y_8)
4	7	$j = 7 > X_4 = 4$	trial rejected
2	9	$j = 9 > X_2 = 5$	trial rejected
9	2	$j = 2 < X_9 = 11$	trial accepted (y_9)