
Assignment2

1 Question 58 - 2019 UGC

A sample of size $n = 2$ is drawn from a population of size $N = 4$ using probability proportional to size without replacement sampling scheme, where the probabilities proportional to size are

The probability of inclusion of unit 1 in the sample is

i	1	2	3	4
Pi	0.4	0.2	0.2	0.2

1. 0.4

2. 0.6

3. 0.7

4. 0.75

Answer : Option - 3 - (0.7)

2 Definition of probability proportional to size without replacement:

In a simple random sampling scheme, every unit in the population, has an equal probability of getting selected for a sample subset. But when if in a given population sizes of different classes are different, then probability that a sample belonging to a certain class (or unit) differs. Sampling is needed when working with large set of population is difficult to handle, and we reduce it and also ensure that all the classes or units have proportional representation. For example lets say if there are 4 classes in a population and class 1 has 40 % of representation then after sampling the probability of class 1's representation is 40%. There are various schemes for doing the same. Here we discuss PPS without replacement.

In a probability scheme without replacement, when the initial probabilities of selection are unequal, then the probability of drawing a specified unit of the population at a given draw changes with the draw. Also immediate repetition of a sample from same class is not allowed i.e, if a sample from class 1 is selected in the first draw, then it is not allowed to draw a sample from same class in the subsequent draw.

3 Solution:

If S_1 is from unit 1 (or class 1), then the probability that S_1 is drawn in the first draw is $P_i = P_1$, which is already given as 0.4.

Now, $P_{i(2)}$ which is probability that i^{th} unit is drawn from the population in the second draw. It can be derived as follows:

$P_{i(2)}$ can occur in following possible ways:

- S_1 is selected in 1st draw and S_i is selected at 2 draw
- S_2 is selected in 1st draw and S_i is selected at 2 draw
- .
- .
- S_{i-1} is selected in 1st draw and S_i is selected at 2 draw
- S_{i-1} is selected in 1st draw and S_i is selected at 2 draw
- S_N is selected in 1st draw and S_i is selected at 2 draw

It has to be noted that S_i is being skipped in for first draw because repetition is not allowed in PPS without replacement scheme.

S_1 is selected in 1st draw and S_i is selected at 2 draw is derived as

Let S_1 is selected in 1st draw be Event A, S_i is selected at 2 as Event B

$$P(A, B) = P(A) \times P(B/A), \quad \text{where, } P(A) = P_1,$$

$$P(B) = \frac{P_i}{1 - (P_1)}$$

Population of P_i over total remaining population, which is $1 - P_1$, because P_1 can not be selected again

It should be noted that dependency of event B on A is removed inherently when, S_i is skipped in the first draw when calculating the entire probability.

$$\begin{aligned} P_{i(2)} &= P_1 \times \frac{P_i}{1 - P_1} + P_2 \times \frac{P_i}{1 - P_2} + \dots \\ &\quad + P_{i-1} \times \frac{P_i}{1 - P_{i-1}} + P_{i+1} \times \frac{P_i}{1 - P_{i+1}} + P_N \times \frac{P_i}{1 - P_N} \\ P_{i(2)} &= \sum_{j=1}^N P_j \times \frac{P_i}{1 - P_j} - P_i \times \frac{P_i}{1 - P_i} \end{aligned}$$

Now, in the given problem, it is asked the inclusion probability of 1st unit. Which means, the probability that 1st unit is drawn in the 1st draw or that is drawn in the second draw. Since in the derivations we removed dependency inherently, it is equal sum of those two probabilities (we subtracted the probability that 1st unit is drawn in the 1st draw **and** is drawn in the second draw which is $P_i \times \frac{P_i}{1 - P_i}$).

Now P_1 is given as 0.4. Need to find $P_{1(2)}$.

$$\begin{aligned} P_{i(2)} &= \sum_{j=1}^N P_j \times \frac{P_i}{1 - P_j} - P_i \times \frac{P_i}{1 - P_i} \\ &= 0.4 \times \left(\frac{0.2}{1 - 0.2} + \frac{0.2}{1 - 0.2} + \frac{0.2}{1 - 0.2} + \frac{0.4}{1 - 0.4} - \frac{0.4}{1 - 0.4} \right) \\ &= 0.4 \times \left(\frac{0.6}{0.8} \right) \\ &= 0.3 \end{aligned}$$

Hence The probability of inclusion of unit 1 in the sample is $0.4 + 0.3 = 0.7$