BL College Journal
Full PDF
www.blcollege.edu.bd
ISSN 2664-228X (Print)
ISSN 2710-3684 (Online)
Volume -IV, Issue-II, December 2022
Pp. 73-89

Inclusion Probability in Simple Random Sampling by Hypergeometric Distribution

Anwar H. Joarder and A.M. Mujahidul Islam

Abstract

We review important probability issues in sampling from simple random sampling without replacement. The inclusion probability can be calculated by enumerating samples which is formidable for most cases of large samples or large population. A good number of possible situations have been considered. We prove that hypergeometric mass function provides an elegant solution to the problem.

Anwar H. Joarder
Department of
Computer Science & Engineering
Faculty of Science & Engineering
Northern University of
Business and Technology
Khulna-9100, Bangladesh
e-mail : ajstat@gmail.com
anwar.joarder@nubtkhulna.ac.bd

A.M. Mujahidul Islam Upazila Statistical Office
Bangladesh Bureau of Statistics
Dighalia, Khulna-9220, Bangladesh
e-mail : mujahid4659@gmail.com

Keywords

simple random sampling, inclusion probability, hypergeometric distribution, conditional probability

1. Introduction to Simple Random Sampling (SRS)

Over the last few decades, there have been important progresses in the methods of sampling. The book by Tille (2006) has described forty-six sampling methods. A simple random sample (srs) of size n is the one which is selected in such a way that every sample of the given size n has an equal probability of being selected. It is worth noting that it is a property of simple random sampling that every element in the population has an equal chance of being included in the sample.

Suppose that the population size is N and the sample size is n. If repetition is allowed, then there will be a total of N^n samples. If we select one of them, the sample is a SRS with replacement. If the repletion is not allowed we will have a total of $\binom{N}{n}$ samples (note that we ignore order). If we select one of them, the sample is a SRS without replacement. If sampling is required in scientific investigations, we popularly adopt SRS without replacement.

Bebbington (1975), Mcleaod and Bellhouse (1983) and Tille (2006) and Ting (2021) have discussed many methods for drawing SRS from a finite population. In what follows, we will be using the Pochammer notation:

$$n^{\{0\}} = 1 \text{ and } n^{\{a\}} = n(n-1)...(n-a+1), \quad a \geq 1. \tag{1.1}$$

The following combinatorial identity is extensively used in the development of inclusion probabilities:

$$\binom{N-a}{n-a} \div \binom{N-b}{n-b} = \frac{(n-b)^{\{a-b\}}}{(N-b)^{\{a-b\}}},$$

where $a > b \geq 1$, or, simply, $\binom{N-c}{n-c} \div \binom{N}{n} = \dfrac{n^{\{c\}}}{N^{\{c\}}}, \; c \geq 1.$ (1.2)

The last identity is obvious by the following steps:

$$\frac{N^{\{c\}}}{n^{\{c\}}}\binom{N-c}{n-c} = \frac{N^{\{c\}}}{n^{\{c\}}} \times \frac{(N-c)!}{(n-c)!(N-n)!}, \text{ or,}$$

$$\frac{N^{\{c\}}}{n^{\{c\}}}\binom{N-c}{n-c} = \frac{N^{\{c\}}(N-c)!}{\left[n^{\{c\}}(n-c)!\right](N-n)!}, \text{ or,}$$

$$\frac{N^{\{c\}}}{n^{\{c\}}}\binom{N-c}{n-c} = \frac{N!}{n!(N-n)!}, \text{ or,}$$

$$\frac{N^{\{c\}}}{n^{\{c\}}}\binom{N-c}{n-c} = \binom{N}{n}.$$

We review some relevant probabilities of SRS in Section 2. Inclusion probabilities of two and three elements are discussed in Sections 3 and 4 respectively. Conditional inclusion probabilities are discussed in Section 5. The general results are presented in Section 6. In Section 7, we conclude by providing a direction for further application of inclusion probability of SRS without replacement to other popular methods of sampling.

2. Probabilities Related to SRS

If X_1 is the first value drawn from a population of size N, X_2 is the second value drawn, $\ldots X_n$ is the n-th value drawn, and the joint probability distribution of these n random variables is given by

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{N^{\{n\}}}, \text{ where } N^{\{n\}} = N(N-1)\ldots(N-n+1) \tag{2.1}$$

(Miller and Miller, 1999) for each ordered n-tuple, then $\{X_1, X_2, \ldots, X_n\}$ is said to constitute a random sample from the given finite population. The probability for each subset of n of the N elements of the finite population is

$$\frac{n!}{N^{\{n\}}} = 1 \div \binom{N}{n}. \tag{2.2}$$

This is often given as an alternative definition or as a criterion for the selection of a random sample of size n from a finite population of size N: Each of the $\binom{N}{n}$ possible samples must have the same probability. This is proved in Corollary 6.1 in many ways.

We now discuss some probabilities related to sampling without replacement by considering a small population. Consider a population of three doctors and two nurses denoted by A, B, C and D, E respectively. Notice that the individuals are distinctly identified. The sample space of a sample of 3 persons selected without replacement is given by

$$\{A,B,C\},\ \{A,B,D\},\ \{A,B,E\},\ \{A,C,D\},\ \{A,C,E\},$$
$$\{A,D,E\},\ \{B,C,D\},\ \{B,C,E\},\ \{B,D,E\},\ \{C,D,E\}.$$

(i) Probability That a Person is Included in a Particular Draw

Let $A_i (i = 1,2,3)$ be the event that Doctor A is included in the i-th selection. Then the probability that A is included in the 1st selection is $= P(A_1) = 1/5$. Since the sampling is without replacement, the probability that A is included in the 2nd selection is given by

$$P(A_1' A_2) = P(A_1')P(A_2 \mid A_1') \text{ which equals}$$

$P(A_1'A_2) = [1-P(A_1)]P(A_2 \mid A_1')$ which equals

$$\left(1-\frac{1+0}{1+4}\right)\left(\frac{1+0}{1+3}\right)=\frac{1}{5}.$$

Also the probability that A is included in the 3rd selection is given by

$P(A_1'A_2'A_3) = P(A_1')P(A_2' \mid A_1')P(A_3 \mid A_1'A_2')$ which equals

$P(A_1'A_2'A_3) = [1-P(A_1)][1-P(A_2 \mid A_1')]P(A_3 \mid A_1'A_2')$ which equals

$$\left(1-\frac{1+0}{1+4}\right)\left(1-\frac{1+0}{1+3}\right)\left(\frac{1+0}{1+2}\right)=\frac{4}{5}\times\frac{3}{4}\times\frac{1}{3}=\frac{1}{5}.$$

Obviously, the probability that unit j of the population of N units is included in the i-th selection is given by

$$\left(1-\frac{1}{N}\right)\left(1-\frac{1}{N-1}\right)\cdots\left(1-\frac{1}{N-(i-2)}\right)\frac{1}{N-(i-1)}=\frac{1}{N}.$$

The marginal distribution of X_r in (2.1) is thus given by

$$f(x_r)=\frac{1}{N},\ x_r=c_1,c_2,\ldots,c_N \tag{2.3}$$

for $r=1,2,\ldots,n.$

(ii) *Probability That a Person Is Included in a Sample*

The probability that Doctor A is included in the sample is given by

$$P(A_1)+P(A_1'A_2)+P(A_1'A_2'A_3)=\frac{1}{5}+\frac{1}{5}+\frac{1}{5}=\frac{3}{5}.$$

Thus each of the 5 persons have the same chance $(3/5)$ of being selected in a without replacement sample of size 3.

Let $M(j)=\binom{N-j}{n-j}$, $j=1,2,\cdots,n.$ The number of samples of size n that contains unit j of the population of N units is $M(1)=\binom{N-1}{n-1}$. Since the total number of samples of size n is given by $M(0)=\binom{N}{n}$, the probability that unit j of the population of N units is included in the sample is given by $\dfrac{M(1)}{M(0)} = \binom{N-1}{n-1} \div \binom{N}{n} = \dfrac{n}{N}$.

(iii) Inclusion Probability of a Unit for SRS Without Replacement

Let π_j be the probability that unit $j(j = 1, 2, \ldots, N)$ of the population is included in the sample. Then $\pi_j = P(j \in \underset{\frown}{s})$, or, $\pi_j = \sum_{\underset{\frown}{s} \ni j} P(\underset{\frown}{s})$, where $\underset{\frown}{s}$ is the set containing all possible $\binom{N}{n}$ samples.

Example 2.1 Let $\{A, B, C, D, E\}$ be a population of 5 units. Then there will be 10 possible samples of size 3.

a. Write out all possible samples.

Solution :
$\{A, B, C\}$, $\{A, B, D\}$, $\{A, B, E\}$, $\{A, C, D\}$, $\{A, C, E\}$,
$\{A, D, E\}$, $\{B, C, D\}$, $\{B, C, E\}$, $\{B, D, E\}$, $\{C, D, E\}$.

b. The probability that the third unit C is included in the sample is given by

$\pi_3 = \pi_C = P(C \in \underset{\frown}{s})$, or, $\pi_3 = \sum_{\underset{\frown}{s} \ni C} P(\underset{\frown}{s})$, or,

$\pi_3 = \pi_C = P(\{A, B, C\}, \{A, C, D\}, \{A, C, E\}, \{B, C, D\}, \{B, C, E\}, \{C, D, E\})$, or,

$\pi_3 = \pi_C = P\{A, B, C\} + P\{A, C, D\} + P\{A, C, E\} + P\{B, C, D\} + P\{B, C, E\} + P\{C, D, E\})$.

Note that $\{A, B, C\} = \{ABC, ACB, BAC, BCA, CAB, CBA\}$. The event $\{A, B, C\}$ means that A, B and C are selected in the sample in any order whereas the event $\{ABC\} = \{A_1 B_2 C_3\}$ and the order is important.

$P\{A, B, C\} = P(ABC, ACB, BAC, BCA, CAB, CBA)$.

Obviously, $P(ABC) = P(A_1 B_2 C_3) = P(A_1) P(B_2 \mid A_1) P(C_3 \mid A_1 B_2)$, which equals

$P(ABC) = \dfrac{1}{5} \times \dfrac{1}{4} \times \dfrac{1}{3} = \dfrac{1}{60}$, i.e.,

$f(x_1, x_2, x_3) = \dfrac{1}{5(4)(3)}$.

Similarly, it can be proved that

$P(ACB) = P(BAC) = P(BCA) = P(CAB) = P(CBA) = \dfrac{1}{60}$ so that

$$P\{A,B,C\} = P(ABC) + P(ACB) + P(BAC) + P(BCA) + P(CAB) + P(CBA) = 6 \times \frac{1}{60} = \frac{1}{10}.$$

Each of the $\binom{N}{n} = 10$ possible samples has the same probability as $1 \div \binom{5}{3}$ Hence $\pi_3 = 6 \div \binom{5}{3}$.

Since in each of the above samples, we have the element C, we are choosing (3-1) other units from the population with $\{A,B,D,E\}$. Since this new 'population' has 4 elements, this can be done in $\binom{5-1}{3-1} = \binom{N-1}{n-1}$ ways. A notation for the number of samples that contain the unit C of the population would be $\sum_{s \ni C} 1 = \binom{5-1}{3-1} = 6$. Finally, we have

$$\pi_3 = \binom{5-1}{3-1} \div \binom{5}{3} = \frac{3}{5}.$$

c. What is the probability that the second unit (i.e. unit B) of the population is selected in the sample?

Solution:

$$\pi_B = P(\{A,B,C\},\ \{A,B,D\},\ \{A,B,E\},\ \{B,C,D\},\ \{B,C,E\},\ \{B,D,E\}),$$

$$\pi_B = P\{A,B,C\} + P\{A,B,D\} + P\{A,B,E\} + P\{B,C,D\} + P\{B,C,E\} + P\{B,D,E\},$$

$$\pi_B = \frac{1}{10} + \frac{1}{10} + \ldots + \frac{1}{10},$$

$$\pi_B = \frac{6}{10} = \frac{3}{5}.$$

A Combinatorial Solution is presented below:

Possible samples are given by the following

$\{A,B,C\},\ \{A,B,D\},\ \{A,B,E\},\ \{A,C,D\},\ \{A,C,E\},$

$\{A,D,E\},\ \{B,C,D\},\ \{B,C,E\},\ \{B,D,E\},\ \{C,D,E\}$

which are $\binom{5}{3} = 10$ in number.

The number of samples that contains unit B of the population is $\binom{5-1}{3-1} = 6$ and they are given below: $\{A,B,C\},\ \{A,B,D\},\ \{A,B,E\},\ \{B,C,D\},\ \{B,C,E\},\ \{B,D,E\}$. This can also be viewed as number of ways of selecting 2 units from $\{A,C,D,E\}$ so that B can be associated with each of the 6 samples. Then the probability that the second unit (i.e. unit B) of the population is selected in the sample is given by

$$\pi_2 = P(a_2 = 1) = \frac{6}{10}.$$

Lemma 2.1 The probability that unit j $(j = 1, 2, \ldots, N)$ of the population is included in a simple random sample of size n without replacement is given by

$$\pi_j = \frac{n}{N}, \quad j = 1, 2, \ldots, N \tag{2.4}$$

Proof. A notation for the number of samples that contain the unit j of the population is $\sum_{\underset{\sim}{s} \ni j} 1 = \binom{N-1}{n-1}.$

The total number of samples of size n that can be drawn from a population of size N is given by $\binom{N}{n}$. Then, $\pi_j = \binom{N-1}{n-1} \div \binom{N}{n}$, which equals $\pi_j = \frac{n^{\{1\}}}{N^{\{1\}}}$ by (1.2). This is the same as (2.4).

3. Inclusion Probability of Two Units

Lemma 3.1 The probability that units i and j $(i = 1, 2, \ldots, N; j = 1, 2, \ldots, N; i \neq j)$ of the population are included in the sample (in any order; order is not important) is given by $\pi_{ij} = \frac{n^{\{2\}}}{N^{\{2\}}}$, (3.1)

where $n^{\{a\}}$ is defined by (1.1).

Combinatorial Proof

Since in case of SRS without replacement, the probability that a sample of $n(< N)$ units is selected from a population of N units is given by $P(\underset{\sim}{s}) = 1 \div \binom{N}{n}$, so that the probability that units i and j are included in the sample is

$$\pi_{ij} = P(i, j \in \underset{\sim}{s}) = \sum_{\underset{\sim}{s} \ni i,j} P(\underset{\sim}{s}),$$

$$\pi_{ij} = \left(\sum_{\underset{\sim}{s} \ni i,j} 1\right) \div \binom{N}{n}.$$

The summand is the number of samples that have both units i and j. This will amount to choosing $(n-2)$ elements from a "population" that does not have units i and j. That is

$$\sum_{\underset{\sim}{s} \ni i,j} 1 = \binom{N-2}{n-2}.$$

Hence, $\pi_{ij} = \binom{N-2}{n-2} \div \binom{N}{n}$.

Then by (1.2), we have $\pi_{ij} = \dfrac{n^{\{2\}}}{N^{\{2\}}}$,

where $a^{\{n\}}$ is defined in (1.1). This is the same as (3.1).

Proof by Conditional Probability

$$\pi_{ij} = P(a_i = 1, a_j = 1) = P(a_i = 1)P(a_j = 1 \mid a_i = 1),$$

$$\pi_{ij} = \pi_i \pi_{j|i} = \left[\binom{N-1}{n-1} \div \binom{N}{n}\right] \times \left[\binom{N-2}{n-2} \div \binom{N-1}{n-1}\right].$$

By (1.2), $\pi_{ij} = \dfrac{n^{\{1\}}}{N^{\{1\}}} \times \dfrac{(n-1)^{\{2-1\}}}{(N-1)^{\{2-1\}}}$,

where $n^{\{a\}}$ is defined in (1.1). This is the same as (3.1).

Proof by Hypergeometric Probability

	Units $\{i, j\}$	Other units	Total
Population	2	$N-2$	N
Sample	2	$n-2$	n

$$\pi_{ij} = \left[\binom{2}{2}\binom{N-2}{n-2}\right] \div \binom{N}{n} = \frac{n^{\{2\}}}{N^{\{2\}}},$$

where $n^{\{a\}}$ is defined in (1.1).

Example 3.1 Consider simple random sampling (without replacement) of 3 units from a population of 5 units $\{A, B, C, D, E\}$. What is the probability that the second and the third units (i.e., units B and C) of the population are selected (order is not important) in the sample?

Solution by Enumeration

Possible samples are given by

$\{A, B, C\}$, $\{A, B, D\}$, $\{A, B, E\}$, $\{A, C, D\}$, $\{A, C, E\}$,
$\{A, D, E\}$, $\{B, C, D\}$, $\{B, C, E\}$, $\{B, D, E\}$, $\{C, D, E\}$.

The samples $\{A,B,C\}$, $\{B,C,D\}$, $\{B,C,E\}$ contain units B and C in any order. The required probability is

$$\pi_{23} = P\left(\{A,B,C\},\ \{B,C,D\},\ \{B,C,E\}\ \right), \text{ or,}$$

$$\pi_{23} = P\left(\{A,B,C\},\ \{B,C,D\},\ \{B,C,E\}\ \right) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10}.$$

Solution by Combinatorial Method

The possible samples are $\binom{5}{3} = 10$, and they are given below:

$$\{A,B,C\},\ \{A,B,D\},\ \{A,B,E\},\ \{A,C,D\},\ \{A,C,E\},$$
$$\{A,D,E\},\ \{B,C,D\},\ \{B,C,E\},\ \{B,D,E\},\ \{C,D,E\}.$$

The samples that have B and C are given by $\{A,B,C\}$, $\{B,C,D\}$, $\{B,C,E\}$. The number of samples can be viewed as choosing $3-2$ units from $\{A,D,E\}$. Then the required probability is $\pi_{23} = \binom{5-2}{3-2} \div \binom{5}{3} = \frac{3}{10}$.

Solution by Conditional Probability

The number of samples that contains unit B of the population is $\binom{5-1}{3-1} = 6$ and they are given below: $\{A,B,C\}$, $\{A,B,D\}$, $\{A,B,E\}$, $\{B,C,D\}$, $\{B,C,E\}$, $\{B,D,E\}$.

The number $\binom{5-1}{3-1} = 6$ can also be viewed as choosing $3-1$ units from $5-1$ units, i.e., from $\{A,C,D,E\}$. These are $\{A,C\}$, $\{A,D\}$, $\{A,E\}$, $\{C,D\}$, $\{C,E\}$, $\{D,E\}$. Then

$$\pi_2 = P(a_2 = 1) = \frac{6}{10}.$$

Thus once unit B of the population is selected in the sample, the unit C of the population can then be selected into sample in $\binom{5-2}{3-2} = \binom{3}{1} = 3$ ways and they are given below: $\{A,C\}$, $\{C,D\}$, $\{C,E\}$. This means we choose one unit C from $\{A,C,E\}$. $\pi_{3|2} = P(a_3 = 1 \mid a_2 = 1) = \frac{3}{6}$. Finally, we have $\pi_{23} = \pi_2 \pi_{3|2} = \frac{6}{10} \times \frac{3}{6} = \frac{3}{10}$.

The above method can be summarized as

$$\pi_{23} = \pi_2 \pi_{3|2} = \left[\binom{5-1}{3-1} \div \binom{5}{3}\right] \times \left[\binom{5-2}{3-2} \div \binom{5-1}{3-1}\right] = \frac{6}{10} \times \frac{3}{6} = \frac{3}{10}.$$

Solution by Repeating Sampling Fractions

$$\pi_{23} = \frac{n}{N} \times \frac{n-1}{N-1} = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}.$$

Solution by Pochammer Method

$$\pi_{23} = \frac{n^{\{2\}}}{N^{\{2\}}} = \frac{3(3-1)}{5(5-1)} = \frac{3}{10}.$$

Example 3.2 Let $\{i, j\}, (i = 1, 2, \cdots, 5; j = 1, 2, \ldots, 5)$ be the two units of the population we want to select in the sample. Then the total of the inclusion probabilities is $\pi = \sum_{i=1}^{i=5} \sum_{j=1}^{j=5} \pi_{ij}$. We want to prove that $\pi = 3^2$.

	1	2	3	4	5
1	0.6	0.3	0.3	0.3	0.3
2	0.3	0.6	0.3	0.3	0.3
3	0.3	03	0.6	0.3	0.3
4	0.3	0.3	0.3	0.6	0.3
5	0.3	0.3	0.3	0.3	0.6

What is the total of these probabilities?

Solution: Note that the inclusion probability matrix is symmetric.

By definition, we have

$$\pi = \sum_{i=1}^{i=5} \sum_{j(\neq i)1}^{j=5} \pi_{ij} + \sum_{i=j=1}^{i=j=5} \pi_{ii},$$

$$\pi = \sum_{i=1}^{i=5} \sum_{j(\neq i)1}^{j=5} \pi_{ij} + \sum_{i=1}^{i=5} \pi_i,$$

$$\pi = \sum_{i=1}^{i=5} \sum_{j(\neq i)1}^{j=5} \frac{3(3-1)}{5(5-1)} + \sum_{i=1}^{5} \frac{3}{5},$$

$$\pi = 3(3-1) + 3 = 3^2.$$

4. Inclusion Probability of Three Elements

The probability that units i, j and k $(i = 1, 2, \ldots, N; j = 1, 2, \ldots, N; j = 1, 2, \ldots, N; i \neq j$

of the population are included in the sample is given by

$$\pi_{ijk} = \frac{n^{\{3\}}}{N^{\{3\}}},$$

where $n^{\{a\}}$ is defined in (1.1).

Proof. $\pi_{ijk} = P(i, j, k \in \underset{\sim}{s}) = \sum_{\underset{\sim}{s} \ni i,j,k} P(\underset{\sim}{s})$.

Combinatorial Solution

Since in case of SRS without replacement, $P(\underset{\sim}{s}) = 1 \div \binom{N}{n}$, we have

$$\pi_{ijk} = \left(\sum_{\underset{\sim}{s} \ni i,j,k} 1\right) \div \binom{N}{n}.$$

The summand is the number of samples that have units i, j and k. This will amount to choosing $(n-3)$ elements from a "population" that does not have units i, j and k.

That is $\sum_{\underset{\sim}{s} \ni i,j,k} 1 = \binom{N-3}{n-3}$.

Hence, $\pi_{ijk} = \binom{N-3}{n-3} \div \binom{N}{n}$. Then by (1.2), we have

$$\pi_{ijk} = \frac{n^{\{3\}}}{N^{\{3\}}},$$

where $n^{\{a\}}$ is defined in (1.1).

Solution by Conditional Probability

$$\pi_{ijk} = P(a_i = 1, a_j = 1, a_k = 1) = P(a_i = 1)P(a_j = 1 \mid a_i = 1)P(a_k = 1 \mid a_i = 1, a_j = 1),$$

$$\pi_{ijk} = \left[\binom{N-1}{n-1} \div \binom{N}{n}\right] \times \left[\binom{N-2}{n-2} \div \binom{N-1}{n-1}\right] \times \left[\binom{N-3}{n-3} \div \binom{N-2}{n-2}\right].$$

By (1.2), we have

$$\pi_{ijk} = \frac{n^{\{1\}}}{N^{\{1\}}} \times \frac{(n-1)^{\{1\}}}{(N-1)^{\{1\}}} \times \frac{(n-2)^{\{1\}}}{(N-2)^{\{1\}}}, \text{ or,}$$

$$\pi_{ijk} = \frac{n^{\{3\}}}{N^{\{3\}}},$$

where $n^{\{a\}}$ is defined in (1.1).

The definition is true even if n is not an integer. Since in the sampling context, n is an integer, the Pochammer polynomial can be interpreted to be a truncated factorial.

Solution by Hypergeometric Probability

	Units $\{i, j, k\}$	Other units	Total
Population	3	$N-3$	N
Sample	3	$n-3$	n

$$\pi_{ijk} = \left[\binom{3}{3}\binom{N-3}{n-3}\right] \div \binom{N}{n}.$$

By (1.2), we have $\pi_{ijk} = \dfrac{n^{\{3\}}}{N^{\{3\}}}$.

5. Inclusion Probability of Two Elements Given Three Elements Selected

Lemma 5.1 The probability that units l and m $(l = 1, 2, \ldots, N; m = 1, 2, \ldots, N; m \neq l)$ of the population are included in the sample given that units i, j and k $(i = 1, 2, \ldots, N;$ $j = 1, 2, \ldots, N;$ $k = 1, 2, \ldots, N;$ $i \neq j \neq k)$ of the population are already included in the sample is given by

$$\pi_{lm|ijk} = \frac{(n-3)^{\{2\}}}{(N-3)^{\{2\}}},$$

where $n^{\{a\}}$ is defined in (1.1).

Proof by Conditional Probability

$$\pi_{lm|ijk} = P(a_l = 1, a_m = 1 \mid a_i = 1, a_j = 1, a_k = 1),$$

$$\pi_{lm|ijk} = P(a_l = 1 \mid a_i = 1, a_j = 1, a_k = 1) \times P(a_m = 1 \mid a_i = 1, a_j = 1, a_k = 1, a_l = 1),$$

$$\pi_{lm|ijk} = \left[\binom{N-4}{n-4} \div \binom{N-3}{n-3}\right] \times \left[\binom{N-5}{n-5} \div \binom{N-4}{n-4}\right].$$

By (1.2), we have

$$\pi_{lm|ijk} = \frac{(n-3)^{\{4-3\}}}{(N-3)^{\{4-3\}}} \times \frac{(n-4)^{\{5-4\}}}{(N-4)^{\{5-4\}}}, \text{ or,}$$

$$\pi_{lm|ijk} = \frac{(n-3)^{\{2\}}}{(N-3)^{\{2\}}}, \text{ where } a^{\{n\}} \text{ is defined in (1.1).}$$

Proof by Hypergeometric Probability

	Units $\{l, m\}$	Other units	Total
Conditional Population	2	$(N-3)-2$	$N-3$
Sample	2	$(n-3)-2$	$n-3$

$$\pi_{lm|ijk} = \left[\binom{2}{2}\binom{N-3-2}{n-3-2}\right] \div \binom{N-3}{n-3}.$$

By (1.2), we have $\pi_{lm|ijk} = \dfrac{(n-3)^{\{2\}}}{(N-3)^{\{2\}}}$,

where $n^{\{a\}}$ is defined in (1.1).

6. The General Result

Theorem 6.1 In general, at the first draw the probability that one of the $b(<n)$ specified units is selected is b/N. At the second draw the probability that one of the remaining $(b-1)$ specified units is drawn is $(b-1)/(N-1)$, and so on. Hence the probability that all b specified units are selected in b draws is

$$\pi_{i_1 i_2 \dots i_b} = \frac{n^{\{b\}}}{N^{\{b\}}}, \tag{6.1}$$

where $n^{\{a\}}$ is defined in (1.1).

Proof by Conditional Probability

$$\pi_{i_1 i_2 \dots i_b} = P(a_{i_1} = 1, a_{i_2} = 1, \dots, a_{i_b} = 1)$$

$$\pi_{i_1 i_2 \dots i_b} = P(a_{i_1} = 1) P(a_{i_2} = 1 | a_{i_1} = 1) P(a_{i_3} = 1 | a_{i_1} = 1, a_{i_2} = 1) \dots P(a_{i_b} = 1 | a_{i_1} = 1, a_{i_2} = 1, \dots, a_{i_{b-1}} = 1),$$

$$\pi_{i_1 i_2 \dots i_b} = \frac{b}{N} \cdot \frac{b-1}{N-1} \cdot \frac{b-2}{N-2} \cdots \frac{b-(b+1)}{N-(b+1)},$$

$$\pi_{i_1 i_2 \dots i_b} = \frac{n^{\{b\}}}{N^{\{b\}}},$$ where $n^{\{a\}}$ is defined in (1.1).

Proof by Hypergeometric Probability

	Units $(i_1, i_2, \ldots, i_b)$	Other units	Total
Population	b	$N-b$	N
Sample	b	$n-b$	n

$$\pi_{i_1,i_2,\ldots,i_b} = \left[\binom{b}{b}\binom{N-b}{n-b}\right] \div \binom{N}{n}.$$

By (1.2), we have, $\pi_{i_1,i_2,\ldots,i_b} = \dfrac{n^{\{b\}}}{N^{\{b\}}}$,

where $a^{\{n\}}$ is defined in (1.1).

An alternative argument is provided now. Since the number of samples that includes units $j_1, j_2, \ldots, j_b$ of the population are included in the sample is given by $M(b) = \binom{N-b}{n-b}$, the probability that units $j_1, j_2, \ldots, j_b$ of the population will be included in the sample is given by

$$\frac{M(b)}{M(0)} = \binom{N-b}{n-b} \div \binom{N}{n} = \frac{n^{\{b\}}}{N^{\{b\}}},$$ where $n^{\{a\}}$ is defined in (1.1).

Corollary 6.1 In general, at the first draw the probability that one of the n specified units is selected is n/N . At the second draw the probability that one of the remaining $(n-1)$ specified units is drawn is $(n-1)/(N-1)$, and so on. Hence the probability that all n specified units are selected in n draws is

$$\pi_{i_1 i_2 \ldots i_n} = 1 \div \binom{N}{n}. \tag{6.2}$$

Proof by Conditional Probability

$$\pi_{i_1 i_2 \ldots i_n} = P(a_1 = i_1, a_2 = i_2, \ldots, a_n = i_n)$$

$$\pi_{i_1 i_2 \ldots i_n} = P(a_1 = i_1) P(a_2 = i_2 \mid a_1 = i_1) \ldots P(a_n = i_n \mid a_1 = i_1, a_2 = i_2, \ldots, a_{n-1} = i_{n-1}),$$

$$\pi_{i_1 i_2 \ldots i_n} = \frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} \cdots \frac{n-(n+1)}{N-(n+1)},$$

$$\pi_{i_1 i_2 \ldots i_n} = \frac{n^{\{n\}}}{N^{\{n\}}} = \frac{n!(N-n)!}{N!} = 1 \div \binom{N}{n}, \text{ (Cochran, 1977, 18).}$$

We remark that since the number of samples that include specific n units of the population is $M(n) = \binom{N-n}{n-n}$, by the arguments of Theorem 4.2, the probability that specified n units of the population of N units is included in a sample of size n is given by

$$\frac{M(n)}{M(0)} = \binom{N-n}{n-n} \div M(0). = \frac{1}{M(0).} = 1 \div \binom{N}{n}.$$

Proof by Hypergeometric Probability

	Units $(i_1, i_2, \ldots, i_n)$	Other units	Total
Population	n	$N-n$	N
Sample	n	0	n

$$\pi_{i_1, i_2, \ldots, i_n} = \left[\binom{n}{n}\binom{N-n}{0}\right] \div \binom{N}{n} = 1 \div \binom{N}{n}.$$

Theorem 6.2 For a sample of size n from a population of size N, the inclusion probabilities that two units in the population will be in the sample add to n^2.

Proof. Let $\{i, j\}, (i = 1, 2, \cdots, N; j = 1, 2, \ldots, N)$ be the two units of the population we want to select in the sample. Then the total of the inclusion probabilities is $\pi = \sum_{i=1}^{N} \sum_{j=1}^{j=N} \pi_{ij}$. We want to prove that $\pi = n^2$. By definition, we have

$$\pi = \sum_{i=1}^{N} \sum_{j(\neq i)1}^{j=N} \pi_{ij} + \sum_{i=j=1}^{N} \pi_{ii},$$

$$\pi = \sum_{i=1}^{N} \sum_{j(\neq i)1}^{j=N} \pi_{ij} + \sum_{i=1}^{N} \pi_i,$$

$$\pi = \sum_{i=1}^{N} \sum_{j(\neq i)1}^{j=N} \frac{n(n-1)}{N(N-1)} + \sum_{i=1}^{N} \frac{n}{N},$$

$\pi = n(n-1) + n = n^2.$

Theorem 6.3 The probability that b units $(j_{a+1}, j_{a+2}, \ldots, j_{a+b})$ where $(j_{a+1} \neq j_{a+2} \neq \ldots \neq j_{a+b})$ of the population will be included in the sample given that other a units $(j_1, j_2, \ldots, j_a)$ where $(j_1 \neq j_2 \neq \ldots \neq j_a)$ of the population are already included in the sample is given by

$\pi_{j_{a+1}, j_{a+2}, \ldots, j_{a+b} | j_1, j_2, \ldots, j_a} = \dfrac{(n-a)^{\{b\}}}{(N-a)^{\{b\}}}$, where $n^{\{a\}}$ is defined in (1.1).

Proof by Conditional Probability

$$\pi_{j_{a+1}, j_{a+2}, \ldots, j_{a+b} | j_1, j_2, \ldots, j_a} = P(a_{j_{a+1}} = 1, a_{j_{a+2}} = 1, \ldots a_{j_{a+b}} = 1 \,|\, a_{j_1} = 1, a_{j_2} = 1, \ldots a_{j_a} = 1),$$

$$\pi_{j_{a+1}, j_{a+2}, \ldots, j_{a+b} | j_1, j_2, \ldots, j_a} = \binom{N-a-b}{n-a-b} \div \binom{N-a}{n-a},$$

By (1.2), we have $\pi_{j_{a+1}, j_{a+2}, \ldots, j_{a+b} | j_1, j_2, \ldots, j_a} = \dfrac{(n-a)^{\{b\}}}{(N-a)^{\{b\}}}$,

where $n^{\{a\}}$ is defined in (1.1).

	Units $\{l, m\}$	Other units	Total
Conditional Population	2	$(N-3)-2$	$N-3$
Sample	2	$(n-3)-2$	$n-3$

Proof by Hypergeometric Method

$$\pi_{j_{a+1}, j_{a+2}, \ldots, j_{a+b} | j_1, j_2, \ldots, j_a} = \left[\binom{b}{b}\binom{N-a-b}{n-a-b}\right] \div \binom{N-a}{n-a}.$$

By (1.2), we have $\pi_{j_{a+1}, j_{a+2}, \ldots, j_{a+b} | j_1, j_2, \ldots, j_a} = \dfrac{(n-a)^{\{b\}}}{(N-a)^{\{b\}}}$,

where $a^{\{n\}}$ is defined in (1.1).

7. Conclusion

We have discussed inclusion probabilities in a number of situations in simple random sampling without replacement. The formidable task of writing algorithms for inclusion probabilities is elegantly simplified by the standard hypergeometric mass function. We believe this will be of immense benefit to students and instructors. The idea can be applied to many other sampling methods, say, to stratified sampling, cluster sampling, ranked set sampling etc.

References

Bebbington, A.C. (1975). A simple method of drawing a sample without replacement. *Applied Statistics*, 24, 136.

Cochran, W.G. (1977). *Sampling Techniques*. John Wiley and Sons.

McLeaod, A.I. and Bellhouse, David R. (1983). A convenient algorithm for drawing a simple random sample. *Applied Statistics*, 32(2), 182-184.

Miller, Irwin and Miller, Merylees (1999). *John E Freund's Mathematical Statistics*, Prentice Hall, New Jersey, USA.

Tille, Yves (2006). *Sampling Algorithms*. Springer.

Ting, Daniel (2021). Simple, Optimal algorithms for simple random sampling without replacement. https://arxiv.org/pdf/2104.05091.pdf