Sigma J Eng Nat Sci, Vol. 42, No. 6, pp. 1780−1785, December, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00134

Research Article

On inclusion probabilities for weighted random sampling without replacement

Murat GÜNGÖR[1,*]

[1]*Department of Industrial Engineering, İstanbul Medeniyet University, İstanbul, 34700, Türkiye*

ARTICLE INFO

Article history
Received: 10 August 2023
Revised: 18 September 2023
Accepted: 11 January 2024

Keywords:
Hajj Draws; Inclusion Probabilities; Weighted Random Sampling; Without Replacement

ABSTRACT

Hajj is an annual Islamic pilgrimage to Mecca, Saudi Arabia. It is performed on certain dates of the lunar year. The Saudi government sets quotas for various countries to keep the pilgrims' number at a manageable level. While some countries maintain waiting lists and evaluate applications on a first-come-first-served basis, others conduct draws to determine who will be admitted to the journey. Türkiye is one of the latter, where candidates' odds are, in a sense, proportional to the square of the number of years they have been waiting for, or to be more accurate, to the square of the number of times they made an application. This policy, which is called "katsayılı kura sistemi" in Turkish, is adopted by countries like Bosnia and Herzegovina and Belgium as well. The sampling process described above is referred to as "weighted random sampling without replacement with defined weights" (WRS) in the literature. The purpose of this paper is to investigate the inclusion probabilities in WRS for which no efficient method exists. First, we take up an analytical approach and derive theoretical lower and upper bounds on the inclusion probabilities. Second, for situations where these bounds are not as tight as desired, we propose an estimation procedure by simulation. The simulation design is based on an ingenious idea from computer science. We apply our results to estimate applicants' chances in Türkiye's last hajj draw before the COVID-19 pandemic. It turns out that one who participates in the draws for the first time has a chance in between 0.12% and 0.13%; similar bounds for one who participates for the eleventh time (for one with the largest number of applications) are 13.22% and 14.16%. These bounds actually rely on a conjecture relating WRS to a more general problem for which we provide a supportive example.

Cite this article as: Güngör M. On inclusion probabilities for weighted random sampling without replacement. Sigma J Eng Nat Sci 2024;42(6):1780−1785.

INTRODUCTION

Suppose N items are partitioned into a number of groups. Let us denote the size and the weight of group i by $n_i \in \mathbb{N}$ and $w_i > 0$, respectively. We study the sampling process of randomly choosing n items in this fashion: in each round one of the groups is selected at random to include in the sample all items in that group, and the chance of selecting a group equals its weight divided by the yet

*Corresponding author.
*E-mail address: murat.gungor@medeniyet.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
Copyright 2021, Yıldız Technical University. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

unselected groups' total weight. The particular situation where each group is a singleton, namely a set with only one item, is called "weighted random sampling without replacement with defined weights" by Efraimidis [1]. We shorten this as WRS, and the general case, which we shall refer to as "weighted random sampling of groups without replacement with defined weights", will be abbreviated to WRSG. Let p_i stand for the probability that group i is included in the sample. Even for WRS no efficient method is known to exist to calculate p_i exactly [1]. The purpose of this paper is to investigate these inclusion probabilities. Our incentive stems from a real-world application about hajj draws conducted every year by some countries.

Research on unequal probability sampling, unlike the subject matter of this paper, mainly focuses on devising sampling designs. Indeed, the primary objective in sampling is to estimate some unknown population parameter, and the choice of sampling design is important since it determines the properties of the estimator used [2]. In this context, the p_i are commonly denoted π_i and called first-order inclusion probabilities. These are usually prescribed beforehand for all items such that the relevant estimator has desirable properties [3].

Hansen and Hurwitz [4] were the first to suggest the use of unequal probabilities in sampling. Yates and Grundy [5] discuss the so-called Yates-Grundy draw-by-draw procedure. They explicitly give the probability of obtaining a specific sample when the sample size is two. Fellegi [6], in his doctoral thesis, carefully distinguishes between the "direct problem" and the "reverse problem" associated with sampling without replacement from a finite population. Rao [7] develops an asymptotic theory thereon. Brewer [8] proposes a simple procedure for sampling without replacement where the probability of inclusion is strictly proportional to size. Hanif and Brewer [9] present a comprehensive review on sampling with unequal probabilities without replacement. They list 50 selection procedures along with their properties and principal reference. These procedures are described in detail in their book [10]. Li [11] introduces an efficient computer implementation of the Yates-Grundy draw-by-draw procedure. This is also discussed by Gelman and Meng [12]. Tille [13] gives a precise definition of draw-by-draw algorithms. Yu [14] obtains some comparison results for the inclusion probabilities in some unequal probability sampling plans without replacement.

More recently, Tille [15] examines the modern development of the theory and foundations of survey sampling. Stamatelatos and Efraimidis [16] view growing preferential attachment models as an application of unequal probability random sampling. Dumelle et al. [17] compare design-based and model-based approaches for finite population spatial sampling and inference. Chauvet [18] and Aubry [19] attract attention to certain issues about a sampling design that is used in popular commercial statistical packages. Tille [20] provides remarks on some misconceptions about unequal probability sampling without replacement.

Weighted random sampling has attracted the attention of computer scientists as well. However, their concern is to develop, under various circumstances, efficient algorithms that actually generate the sample in question. For WRS, Efraimidis and Spirakis [21] present such an algorithm, which yields a sample in one pass over the entire population.

Here is an outline of the paper: in Section 2, we make a conjecture that relates WRSG to WRS. In Section 3, for the inclusion probabilities in WRS, we derive theoretical lower and upper bounds, and discuss how simulation can be used for estimation. In Section 4, we apply our results to Türkiye's 2020 hajj draw. Finally, in Section 5, we conclude with some future research opportunities.

A Conjecture Relating *WRSG* to *WRS*

Suppose that two groups j, k unite to form a new group l with and $n_l := n_j + n_k$ and $w_l := (n_j w_j + n_k w_k)/(n_j + n_k)$. Thus, w_l is the weighted average of the component weights. This operation decreases the number of groups by 1, and changes the inclusion probabilities of all groups in question. We conjecture that, for non-uniting groups, the change in inclusion probabilities is asymptotically zero.

Conjecture. Let p'_i be the inclusion probability of group i after two groups other than i unite. For fixed n/N, the probability p'_i tends to p_i in the limit $N \to \infty$.

We cannot prove this conjecture for the time being, but we will provide a supportive example. Assume all groups have size 1 and all weights are 1. Let p' be the inclusion probability of a group after two other groups unite to form a new group of size 2. We shall compute $q' := 1 - p'$. The exclusion probability given that the new group is not included in the sample is

$$\frac{N-3}{N-1}\cdot\frac{N-4}{N-2}\cdots\frac{N-n-1}{N-n+1}\cdot\frac{N-n-2}{N-n-1}. \qquad (1)$$

Note that the last factor's denominator is $N - n - 1$ rather than $N - n$ since in the final round the new group is automatically discarded as its size exceeds remaining capacity. The exclusion probability given that the new group is included in the sample is, by conditioning on the round it is selected,

$$(n-1)\cdot\frac{(N-3)(N-4)\cdots(N-n)}{(N-1)(N-2)\cdots(N-n+1)}. \qquad (2)$$

Summing up (1) and (2) we get, after simplification,

$$q' = \frac{(N-n)(N-3)}{(N-1)(N-2)}. \qquad (3)$$

Thus, for fixed n/N, we have $\lim_{N\to\infty} q' = 1 - n/N = q$, implying $\lim_{N\to\infty} p' = p$ as desired.

If the conjecture is correct, then by decomposing groups other than i into singletons, we obtain almost a WRS with p_i approximately equal to the original inclusion probability of group i (decomposition yields an exact WRS if $n_i = 1$).

In other words, inclusion probability calculation in WRSG essentially reduces to that in WRS. Therefore, in the sequel, we examine the inclusion probabilities for WRS rather than for WRSG. As a consequence, we do not distinguish between items and groups from this point on.

Lower and Upper Bounds on Inclusion Probabilities

As specified above, we consider WRS, namely the procedure of randomly selecting n out of N items without replacement, where items have associated weights so that in each round the probability of selecting an item equals its weight divided by the yet unselected items' total weight.

Let $q_i := 1 - p_i$. Thus, q_i is the exclusion probability for item i. Let Y_t represent the index selected in round t ($1 \le t \le n$). These are dependent random variables with support $N := \{1,2,\ldots,N\}$. Let $N(i) := N\backslash\{i\}$ for short. Then

$$\begin{aligned} q_i &= P(Y_1 \neq i, \ldots, Y_n \neq i) \\ &= P\big(Y_1 \times \cdots \times Y_n \in \mathcal{N}(i) \times \cdots \times \mathcal{N}(i)\big). \end{aligned} \quad (4)$$

Let w_j be the weight of item j ($1 \le j \le N$). We denote the sum of all weights by $W := \sum_{j=1}^{N} w_j$. For $n = 1$, we have $q_i = 1 - P(Y_1 = i) = 1 - w_i/W$. For $n = 2$, conditioning Y_2 on Y_1,

$$\begin{aligned} q_i &= \sum_{j \in \mathcal{N}(i)} P(Y_1 = j) P(Y_2 \in \mathcal{N}(i) \mid Y_1 = j) \\ &= \sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} P(Y_1 = j)\, P(Y_2 = k \mid Y_1 = j) \\ &= \sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} \frac{w_j}{W} \frac{w_k}{W - w_j} \end{aligned} \quad (5)$$

where $N(i,j) := N\backslash\{i,j\}$. Similarly, for $n = 3$, conditioning Y_2 on Y_1, and Y_3 on Y_2 and Y_1,

$$\begin{aligned} q_i &= \sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} P(Y_1 = j)\, P(Y_2 = k \mid Y_1 = j)\, P(Y_3 \in \mathcal{N}(i) \mid Y_2 = k, Y_1 = j) \\ &= \sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} \sum_{l \in \mathcal{N}(i,j,k)} P(Y_1 = j)\, P(Y_2 = k \mid Y_1 = j)\, P(Y_3 = l \mid Y_2 = k, Y_1 = j) \\ &= \sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} \sum_{l \in \mathcal{N}(i,j,k)} \frac{w_j}{W} \frac{w_k}{W - w_j} \frac{w_l}{W - w_k - w_l} \end{aligned} \quad (6)$$

where $N(i,j,k) := N\backslash\{i,j,k\}$. This formula is straightforward to generalize for arbitrary n [22]. Although it provides a means of computing q_i exactly, the number of operations is of order N^n, making it useless for large parameter values. However, the formula can be used to obtain lower and upper bounds on q_i.

Theorem 1. Let a_t and b_t denote the sums of respectively the smallest and the largest t weights other than w_i, and let a_0 and b_0 be defined as zero. Then

$$\prod_{t=0}^{n-1} \left(1 - \frac{w_i}{W - a_t}\right) \ge q_i \ge \prod_{t=0}^{n-1} \left(1 - \frac{w_i}{W - b_t}\right). \quad (7)$$

Proof. Suppose without loss of generality that weights are in increasing order, namely $w_1 \le \ldots \le w_n$. Then $w_1 \le w_j \le w_N$ for any j so that $P(Y_2 \in N(i) \mid Y_1 = j) = 1 - P(Y_2 = i \mid Y_1 = j) = 1 - w_i/(W - w_j)$ is greater than or equal to $1 - w_i/(W - w_N)$ and less than or equal to $1 - w_i/(W - w_1)$.. Consequently, for $n = 2$,

$$\sum_{j \in \mathcal{N}(i)} P(Y_1 = j)\left(1 - \frac{w_i}{W - w_N}\right) \le q_i \le \sum_{j \in \mathcal{N}(i)} P(Y_1 = j)\left(1 - \frac{w_i}{W - w_1}\right); \quad (8)$$

that is,

$$\left(1 - \frac{w_i}{W}\right)\left(1 - \frac{w_i}{W - w_N}\right) \le q_i \le \left(1 - \frac{w_i}{W}\right)\left(1 - \frac{w_i}{W - w_1}\right). \quad (9)$$

Similarly $P(Y_3 \in N(i) \mid Y_2 = k, Y_1 = j) = 1 - P(Y_3 = i \mid Y_2 = k, Y_1 = j) = 1 - w_i/(W - w_k - w_j)$, is greater than or equal to $1 - w_i/(W - w_N - w_{N-1})$ and less than or equal to $1 - w_i/(W - w_1 - w_2)$. Hence, for $n = 3$,

$$\begin{aligned} &\sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} P(Y_1 = j)\, P(Y_2 = k \mid Y_1 = j)\left(1 - \frac{w_i}{W - w_N - w_{N-1}}\right) \\ &\le q_i \le \sum_{j \in \mathcal{N}(i)} \sum_{k \in \mathcal{N}(i,j)} P(Y_1 = j)\, P(Y_2 = k \mid Y_1 = j)\left(1 - \frac{w_i}{W - w_1 - w_2}\right), \end{aligned} \quad (10)$$

implying

$$\begin{aligned} &\left(1 - \frac{w_i}{W}\right)\left(1 - \frac{w_i}{W - w_N}\right)\left(1 - \frac{w_i}{W - w_N - w_{N-1}}\right) \\ &\le q_i \le \left(1 - \frac{w_i}{W}\right)\left(1 - \frac{w_i}{W - w_1}\right)\left(1 - \frac{w_i}{W - w_1 - w_2}\right). \end{aligned} \quad (11)$$

The statement can be proved with the same reasoning by mathematical induction on n. ■

As a consequence of Theorem 1,

$$1 - \prod_{t=0}^{n-1} \left(1 - \frac{w_i}{W - a_t}\right) \le p_i \le 1 - \prod_{t=0}^{n-1} \left(1 - \frac{w_i}{W - b_t}\right). \quad (12)$$

Since $1 - w_i/(W - a_t)$ and $1 - w_i/(W - b_t)$ are both decreasing sequences of t, the inequalities

$$\left(1 - \frac{w_i}{W}\right)^n \ge q_i \ge \left(1 - \frac{w_i}{W - b_{n-1}}\right)^n \quad (13)$$

and

$$1 - \left(1 - \frac{w_i}{W}\right)^n \le p_i \le 1 - \left(1 - \frac{w_i}{W - b_{n-1}}\right)^n \quad (13)$$

also follow from Theorem 1.

If many items possess the same weight, then inclusion probabilities can also be estimated via simulation by using an efficient algorithm of Efraimidis and Spirakis [21]. Let m_i be the total number of items in the population with weight w_i, and suppose that in the sample generated by the Algorithm below there are x_i such items. Then x_i/m_i. is a point estimate for p_i. Generating sufficiently many samples, one can also construct a confidence interval.

Algorithm (Efraimidis and Spirakis [21]).

1. For each item i, generate a uniform random variable u_i in (0,1) and define $k_i := u_i^{1/w_i}$.
2. Select the n items with the largest k_i.

That the Algorithm above yields samples according to WRS rests upon a proposition of Efraimidis and Spirakis [21], given here as Theorem 2. Indeed, for $\alpha = 2$, Theorem 2 says

$$P(K_1 \leq \cdots \leq K_N) = \prod_{i=1}^{N} \frac{w_i}{w_1 + \cdots + w_i}. \quad (14)$$

Left-hand side is the probability that the descending order of the K_i is $K_N, K_{N-1}, \ldots, K_2, K_1$; right-hand side is the probability that WRS, when $n = N$, yields the permutation $(N, N - 1, \ldots, 2, 1)$. In fact, the foregoing relation holds for any permutation of $1, \ldots, N$. Hence, the probability that WRS yields the sample $\{i_1, \ldots, i_n\}$ is equal to the probability that the set $\{k_{i_1}, \ldots, k_{i_n}\}$ is composed of the largest n numbers among the k_i.

Theorem 2 (Efraimidis and Spirakis [21]). Let U_i be independent uniform random variables in (0,1). For $w_i > 0$, let $K_i := U_i^{1/w_i}$. Then, for any $\alpha \in [0,1]$,

$$P(K_1 \leq \cdots \leq K_N \leq \alpha) = \alpha^{w_1 + \cdots + w_N} \prod_{i=1}^{N} \frac{w_i}{w_1 + \cdots + w_i}. \quad (15)$$

Proof. The density function of K_i is $f_i(\alpha) = w_i \alpha^{w_i - 1}$. The proof proceeds by mathematical induction on N, using the equality

$$P(K_1 \leq \cdots \leq K_N \leq \alpha) = \int_0^{\infty} P(K_1 \leq \cdots \leq K_{N-1} \leq t) f_N(t) dt. \quad (16)$$

Application to Hajj Draws

The Saudi government sets hajj quotas for countries to keep the pilgrims' number at a manageable level. Some countries maintain waiting lists and evaluate applications on a first-come-first-served basis, while others conduct draws to determine who will be admitted to the journey. Türkiye is one of the latter, where candidates' odds are proportional to the square of the number of times they made an application [23]. More precisely, in each round of a draw, someone applying for the second time is four times as likely to be selected as one who applies for the first time, someone applying for the third time is nine times as likely, and so on. This policy, which is known as "katsayılı kura sistemi" in Turkish, is adopted by countries like Bosnia and Herzegovina and Belgium as well. We shall estimate a candidate's chance given the number of applications he or she has made. This is a crucial piece of information for applicants as well as policymakers.

The sampling process described above is an example of WRS. However, in reality, the situation is more complex since people usually apply as a group as they want to perform hajj together with one or more of their relatives. Weights associated to such groups are the average of the individual weights, in compliance with the definition in Section 2. Therefore, Türkiye's hajj draws are actually examples of WRSG. Nevertheless, assuming the conjecture in Section 2, we shall proceed as if we have a WRS. Subsequent results must be interpreted accordingly.

Let N and n denote the total number of applicants and the quota; c the maximum number of applications; w_u, r_u, and m_u the weight, the ratio, and the number of applicants with u applications ($1 \leq u \leq c$). Consequently, $w_u = u^2$ and $m_u = Nr_u$. The question is to calculate a particular applicant's chance (inclusion probability) p_u given that he or she has made u applications.

We note that weighted random sampling without replacement, in this context, is also related to a generalization of the hypergeometric distribution. Let X_u be the random variable representing the number of people in the sample with u applications. The joint distribution of the X_u is referred to as the multivariate Wallenius' noncentral hypergeometric distribution [24]. Then each p_u above can be seen as a marginal probability associated with an additional singleton to be extracted from the related set of candidates. Nevertheless, calculation of p_u this way is tractable only for very small values of n.

For 2020's hajj draw in Türkiye, N = 2,298,800 and n = 83,430. This means that over two million Turks declare their intention to perform hajj, whereas the quota is about eighty thousand. We have $c = 11$; in other words, there is no applicant with 12 or more applications. This is a consequence of a previous decree by the Turkish government that led to the admission of a certain group of candidates who had been waiting for too long. The ratios r_u, up to two decimal points, as well as the numbers $m_u = Nr_u$ rounded to nearest integers are given in Table 1. We adjusted m_1 so that the group sizes add up to N.

Lower and upper bounds on p_u provided by (12) are denoted $\underline{p_u}$ and $\overline{p_u}$, and given in the second and third columns of Table 2. Values are rounded to four decimal points

Table 1. Ratio and number of people with u applications in Türkiye's 2020 hajj draw

u	r_u (%)	m_u
1	7.59	174,480
2	7.73	177,697
3	12.97	298,154
4	13.87	318,844
5	22.75	522,977
6	9.60	220,685
7	8.30	190,800
8	8.21	188,731
9	4.25	97,699
10	2.02	46,436
11	2.71	62,297
Total	100	2,298,800

Table 2. Lower and upper bounds, point estimates, and 99% confidence intervals for p_u expressed as percentages

u	$\underline{p_u}$ (%)	$\overline{p_u}$ (%)	$\widehat{p_u}$ (%)	$\text{CI}_{99\%}$ (%)
1	0.12	0.13	0.120	(0.118, 0.123)
2	0.47	0.50	0.481	(0.476, 0.485)
3	1.05	1.13	1.081	(1.076, 1.086)
4	1.86	2.00	1.920	(1.915, 1.926)
5	2.89	3.11	2.987	(2.981, 2.993)
6	4.13	4.44	4.264	(4.252, 4.276)
7	5.58	6.00	5.760	(5.747, 5.773)
8	7.23	7.76	7.456	(7.443, 7.469)
9	9.06	9.72	9.352	(9.329, 9.374)
10	11.06	11.86	11.422	(11.388, 11.456)
11	13.22	14.16	13.623	(13.587, 13.660)

and expressed as percentages. Also we implemented the Algorithm in Section 3 in R, and ran 100 replications. On a PC with Intel(R) Core(TM) i7-4790 CPU (3.60 GHz) and 8 GB RAM, each replication takes less than half a second. Point estimates $\widehat{p_u}$ and 99% confidence intervals are given in the fourth and fifth columns of Table 2. Values are rounded to five decimal points and expressed as percentages.

Table 2 shows that one who takes part in the draws for the first time has a chance in between 0.12% and 0.13%; a point estimate for p_1 is 0.120%, and with 99% probability $p_1 \in (0.118\%, 0.123\%)$. Clearly, in this case, the left endpoint of the confidence interval can be replaced with the theoretical lower bound 0.12%. Bounds and point estimates for p_u gradually increase with respect to u. Lastly, one who participates for the eleventh time has a chance in between 13.22% and 14.16%; a point estimate for p_{11} is 13.623%, and with 99% probability $p_{11} \in (13.587\%, 13.660\%)$.

CONCLUSION

Every year Türkiye conducts a draw to determine the citizens that will be admitted to hajj. Candidates' odds are proportional to the square of the number of times they made an application. This policy is called "katsayılı kura sistemi" in Turkish. Abstraction of this sampling procedure is known as "weighted random sampling without replacement with defined weights" (WRS) in the literature. In this paper, we investigated the inclusion probabilities in WRS for which no efficient method exists. More precisely, we derived lower and upper bounds on inclusion probabilities in terms of item weights. As an application thereof, we estimated applicants' chances in Türkiye's 2020 hajj draw. The computational study shows that one who takes part in the draws for the first time has a probability in between 0.12% and 0.13%; similar bounds for one who takes part for the eleventh time (for one with the largest number of applications) are 13.22% and 14.16%. These results are supported with point estimates based on one-pass simulation experiments that yield very tight confidence intervals. Our findings depend on a conjecture for which we give a confirmational example.

It would be of interest for future research to investigate the problem of inclusion probability calculation from a computational complexity point of view, and to find, if possible, point estimates that do not rely on simulation. It is also worthwhile to prove the conjecture stated in Section 2, relating "weighted random sampling of groups without replacement with defined weights" (WRSG) to WRS.

ACKNOWLEDGMENTS

I am grateful to H. Yunus Apaydın and Osman Balcı for helping me obtain the information in Table 1, namely the ratio of applicants with different number of applications in Türkiye's 2020 hajj draw.

CODE AVAILABILITY

The R code used for the computational study is available from the author upon request.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Efraimidis PS. Weighted random sampling over data streams. In Zaroliagis C, Pantziou G, Kontogiannis S (Eds.). Algorithms, Probability, Networks, and Games, vol. 9295, 2015. Midtown Manhattan, New York City: Springer; p. 183–195. [CrossRef]

[2] Grafstrom, A. (2010). On Unequal Probability Sampling Designs (Doctorial dissertation). Umea: Umea University; 2010. [CrossRef]

[3] Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 1952;47:663–685. [CrossRef]
[4] Hansen MN, Hurwitz WN. On the theory of sampling from finite Populations. Annal Math Stat 1943;14:333–362. [CrossRef]
[5] Yates F, Grundy PM. Selection without replacement from within strata with probability proportional to size. J Royal Stat Soc Ser B Method 1953;15:253–261. [CrossRef]
[6] Fellegi IP. Sampling without replacement with probabilities proportional to size. J Am Stat Assoc 1963;58:183–201. [CrossRef]
[7] Rao JN. Sampling procedures involving unequal probability selection (Doctorial dissertation). Iowa: Iowa State University; 1961.
[8] Brewer KR. A Simple Procedure for Sampling pipswor. Australian J Stat 1975;17:166–172. [CrossRef]
[9] Hanif M, Brewer KR. Sampling with unequal probabilities without replacement: A review. Int Stat Rev 1980;48:317. [CrossRef]
[10] Brewer KR, Hanif M. Sampling With Unequal Probabilities. Midtown Manhattan, New York City: Springer; 1983. [CrossRef]
[11] Li KH. A computer implementation of the yates-grundy draw by draw procedure. J Stat Comput Simul 1994;50:147–151. [CrossRef]
[12] Gelman A, Meng XL. Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family. Hoboken, New Jersey: Wiley; 2004. [CrossRef]
[13] Tille Y. Sampling Algorithms. Midtown Manhattan, New York City: Springer; 2006.
[14] Yu Y. On the inclusion probabilities in some unequal probability sampling plans without replacement. Bernoulli 2012;18:279–289. [CrossRef]
[15] Tille Y. Sampling and Estimation from Finite Populations. Hoboken, New Jersey: Wiley; 2020. [CrossRef]
[16] Stamatelatos G, Efraimidis PS. About Weighted Random Sampling in Preferential Attachment Models, 2021. Preprint http://arxiv.org/abs/2102.08173
[17] Dumelle M, Higham JM, Ver Hoef JM, Olsen AR, Madsen L. A comparison of design-based and model-based approaches for finite population spatial sampling and inference. Methods Ecol Evol 2022;13:2018–2029. [CrossRef]
[18] Chauvet G. A cautionary note on the Hanurav-Vijayan sampling algorithm. J Survey Stat Method 2022;10:1276–1291. [CrossRef]
[19] Aubry P. On the correct implementation of the Hanurav-Vijayan selection procedure for unequal probability sampling without replacement. Commun Stat Simul Comput 2023;52:1849–1877. [CrossRef]
[20] Tille Y. Remarks on some misconceptions about unequal probability sampling without replacement. Comput Sci Rev 2023;47:100533. [CrossRef]
[21] Efraimidis PS, Spirakis PG. Weighted random sampling with a reservoir. Inform Process Lett 2006;97:181–185. [CrossRef]
[22] Fellegi IP. Sampling with varying probabilities without replacement: Rotating and non-rotating samples. J Am Stat Assoc 1963;58:183–201. [CrossRef]
[23] Anadolu Ajansı. Hac kurasına ilişkin merak edilenler, 2021. Available at: https://www.aa.com.tr/tr/turkiye/hac-kurasina-iliskin-merak-edilenler/1694685 Last Accessed Date: 30.07.2024.
[24] Fog A. Calculation methods for Wallenius' noncentral hypergeometric distribution. Commun Stat Simul Comput 2008;37:258–273. [CrossRef]