Roulette Model of Systematic Sampling

Fumito Muguruma*

Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labour and Welfare, Kasumigaseki, Chiyoda-ku, Tokyo, JAPAN
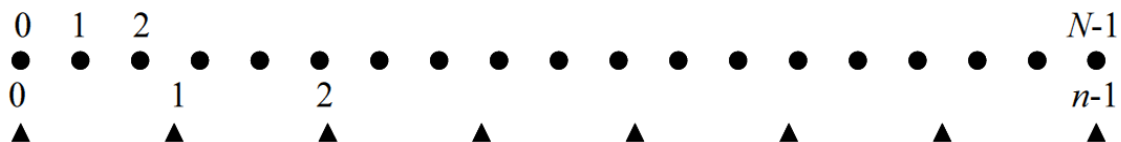*Corresponding author: muguruma-fumito@mhlw.go.jp

Abstract Considering a natural model of systematic sampling, we can calculate the expectation of statistic exactly as it is without regarding it as any other sampling methods. I will demonstrate the details of calculation and get some results. It can be shown that the sample mean of systematic sampling is an unbiased estimator of population mean. Variance of sample mean can be described explicitly and depends on how to sort a population list. Sum of the variance of sample mean and the mean of sample variance is kept constant between random and systematic sampling. As the sample size becomes larger, the variance of sample mean converges to 0.

Cite This Article: Fumito Muguruma, "Roulette Model of Systematic Sampling." *American Journal of Applied Mathematics and Statistics*, vol. 2, no. 5 (2014): 344-351. doi: 10.12691/ajams-2-5-8.

1. Introduction

So many studies have been done about systematic sampling (Kish [1], Cochran [3], Särndal et al. [4], Sharon [5], Thompson [6], Groves et al. [7]). In this paper, on a basis of former studies, I propose a new model of systematic sampling, which procedure is a little bit different from the former ones, and try giving a theory. The model, which I call Roulette Model, can be executed easily and enables us to calculate the expectation of statistic *exactly as it is*.

1.1. What is and what should be Systematic Sampling?

1.1.1. Procedures so far

Let us consider a problem to select *n* out of *N* elements by systematic sampling.

When N is an integral multiple of n, the sampling interval N/n is also an integer. If a random start N_0 is drawn from 0 to $\dfrac{N}{n}-1$, the sample is determined by

$$N_0 + i \cdot \frac{N}{n} - \text{th} \qquad (1.1.1)$$

elements in a population list for $i = 0, 1, 2, \cdots, n-1$. But N is given at first and what we can decide is only n. What if N is a prime number? In general, N/n is not necessarily an integer and problems with the sampling intervals will arise.

By Kish [1], the following four solutions were given (pp. 115-117).

1. Permit the sample size to be either n or $n+1$.
2. Eliminate with epsem (equal probability of selection method).
3. Consider the list to be circular.
4. Using fractional intervals.

He wrote "the sampler should choose the most convenient". Cochran [3] gave an integral sampling interval at first and permitted the sample size to change (pp. 205-206). He also introduced a method suggested by Lahiri [8] (see Murthy[2] p. 139). Särndal et al. [4] organized the former studies and gave the definitions and main results. Anyway, the sampling procedures are optional.

1.1.2. Arrangement and Reconstruction

On a basis of former studies, I will arrange and reconstruct the procedure of systematic sampling in my own way as follows. Let us draw a numerical line across N●-points.
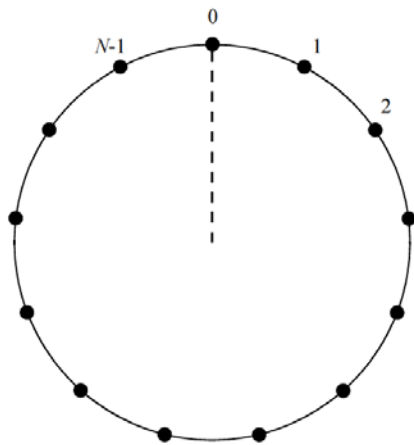
Let the random start N_0 be a random number over $\{0,1,2,\cdots,N-1\}$ and consider the population list to be circular.
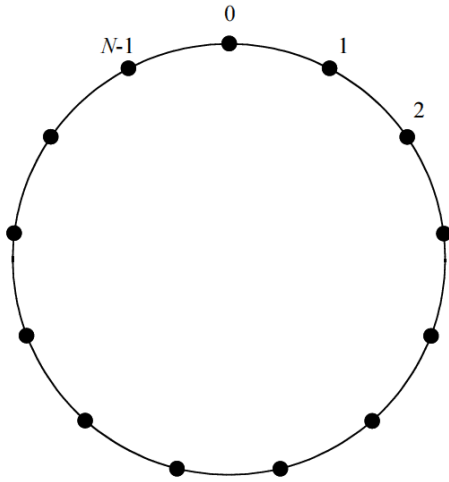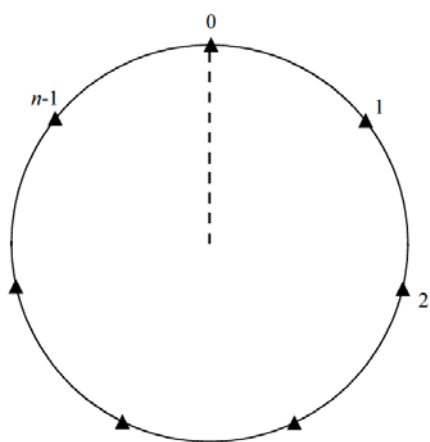

And decide the sample by

$$N_0 + i \cdot \frac{N}{n} - \text{th} \qquad (1.2.1)$$

elements in the list for $i = 0,1,2,\cdots,n-1$. The problem lies in the case when $N_0 + i \cdot N/n$ is not an integer. Which should I select

$$\left\lfloor N_0 + i \cdot \frac{N}{n} \right\rfloor \quad \text{or} \quad \left\lceil N_0 + i \cdot \frac{N}{n} \right\rceil ? \qquad (1.2.2.)$$

Which to select?

In a sampling problem, I don't think it essential whether N/n is an integer or not. I also think it unnatural that only the first sampling point (when $i = 0$) is always an integer. Since N/n is not necessarily an integer, the procedure of systematic sampling should be considered as a continuous problem. The following concept seems to be the most natural for me.

1.1.3. Roulette Model

i. Draw two circles with the same length of circumference N. On one, put N●-points

$$\{0,1,2,\cdots,N-1\}$$

with the same interval 1 and set the point 0 at the top. On another, put n▲-points

$$\{0,1,2,\cdots,n-1\}$$

with the same interval N/n and setthe point 0 at the top. It doesn't matter whether N/n is an integer or not.

ii. Rotate the latter circle by random angle along the circumference *like roulette*.

iii. And lap it over the former one.
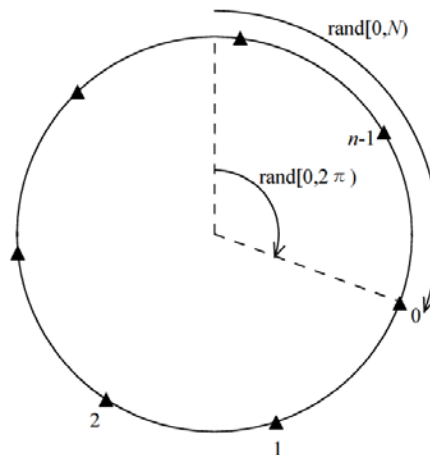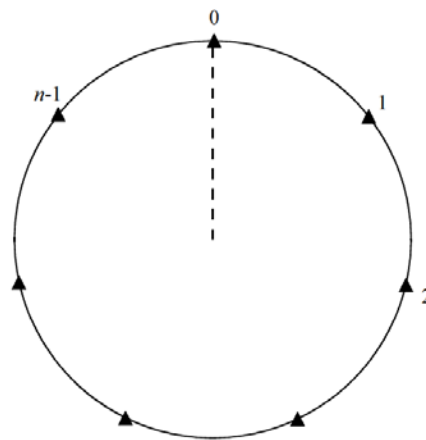
iv. As a sample, choose ●-points correspondent to where ▲-points dropped.

This simple procedure above is a concept of Roulette Model. It can be formulated as follows.

2. Formulation

2.1. Assumption

In this paper, I assume all functions are defined on $\mathbb{R}/N$. That is, if a parameter of a function is outside of the interval
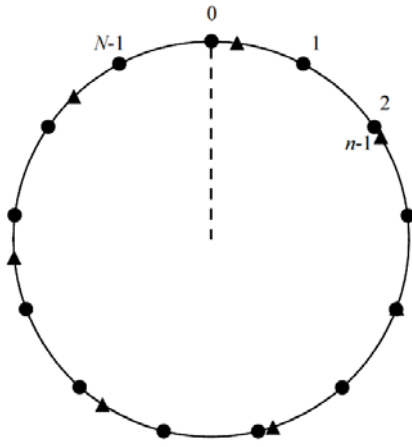
$$[0,N) = \{t|0 \leqq t < N\}, \qquad (2.1.1)$$

I add or subtract N by necessary times so as to settle it within $[0,N)$.

2.2. Identification

i. A circle with the length of circumference N can be regarded as an interval $[0,N)$. So I identify the population list $\{0,1,2,\cdots,N-1\}$ with an interval $[0,N)$.

ii. As the interval $[0,N)$ is a disjoint union of N unit intervals

$$[0,N) = [0,1) \sqcup [1,2) \sqcup \cdots \sqcup [N-1,N), \qquad (2.2.1)$$

I identify an element $i(i = 0,1,2,\cdots,N-1)$ in a population list with a unit interval

$$I_i = [i, i+1). \qquad (2.2.2)$$

iii. If a ▲-point belongs to I_i, select i as an element of sample. Here I don't require any relation between N and n. So, when $n > N$, i can be selected repeatedly the same times as the number of ▲-points which belong to I_i.

2.3. Partition of Random Number

In the above procedure, I used a random angle, which is a uniform random number over an interval

$$[0,2\pi) = \{t|0 \leqq t < 2\pi\}. \qquad (2.3.1)$$

Along the circumference, this is equivalent to a uniform random number over an interval

$$[0,N) = \{t|0 \leqq t < N\}. \qquad (2.3.2)$$

I consider producing the latter from now on.

As an interval $[0,N)$ is a disjoint union of N unit intervals

$$[0,N) = \coprod_{i=0}^{N-1} I_i, \qquad (2.3.3)$$

I consider producing two independent random numbers N_0 and ε as

$$N_0 \sim \text{rand}\{0,1,2,\cdots,N-1\}, \varepsilon \sim \text{rand}[0,1). \qquad (2.3.4)$$

Here N_0 is a discrete uniform random number over $\{0,1,2,\cdots,N-1\}$ and ε is a continuous uniform random number over $[0,1)$. Then $N_0 + \varepsilon$ will surely be a uniform random number over $[0,N)$.

$$N_0 + \varepsilon \sim \text{rand}[0,N). \qquad (2.3.5)$$

By N_0, I select a unit interval I_{N_0} among N intervals $I_0, I_1, I_2, \ldots, I_{N-1}$. And by ε, I focus on a point in the selected interval. That is, I decide a random number over $[0,N)$ in two steps.

2.4. Procedure

In Roulette Model, I decide the first sampling point not by N_0 but by $N_0 + \varepsilon$, identify an element $i(i = 0,1,2,\cdots,N-1)$ in a population list with a unit interval $I_i = [i, i+1)$ and choose i when

$$N_0 + \varepsilon + j \cdot \frac{N}{n} \in I_i \qquad (2.4.1)$$

for $j = 0,1,2,\cdots,n-1$. Then

$$\left\lfloor N_0 + \varepsilon + j \cdot \frac{N}{n} \right\rfloor - \text{th} \qquad (2.4.2)$$

elements for $j = 0,1,2,\cdots,n-1$ re selected as a sample because

$$t \in I_i \Leftrightarrow \lfloor t \rfloor = i. \qquad (2.4.3)$$

The existence of ε-term is a different point from the former procedures. By contribution of ε-term, even if the same first sample N_0 and the same sampling interval N/n are given, stilla different sample can be selected when neither N/n nor n/N is an integer.

3. Calculation

3.1. Expectation of Statistic

The greatest advantage of Roulette Model is, just by adding ε-term, to enable us to calculate the expectation of statistic exactly as it is.

Assume that our interest is value x. The population is described as

$$\{x_0, x_1, x_2, \cdots, x_{N-1}\} \qquad (3.1.1)$$

and the sample is described as

$$\{X_0, X_1, X_2, \cdots, X_{n-1}\}. \qquad (3.1.2)$$

Given N_0 and ε, the sample is decided. Considering any statistic written by

$$f(N_0, \varepsilon; X_0, X_1, X_2, \cdots, X_{n-1}), \qquad (3.1.3)$$

as is known from how we produced two random numbers (N_0, ε), its expectation is calculated by

$$E(f) = \int_0^1 \left(\frac{1}{N} \sum_{i=0}^{N-1} f(i, \varepsilon; X_0, X_1, X_2, \cdots, X_{n-1}) \right) d\varepsilon, \quad (3.1.4)$$
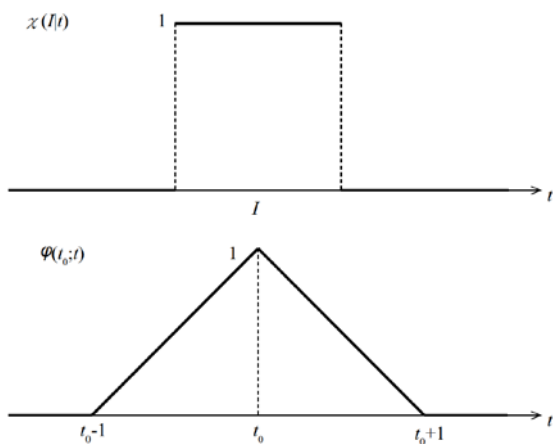
where the order of summation and integral is commutative.

3.2. Mathematical Preparation

I define a characteristic function $\chi(I|t)$ and a pulse function $\varphi(t_0; t)$ by

$$\chi(I|t) = \begin{cases} 1 & t \in I \text{ for } I \subseteq \mathbb{R}/N \\ 0 & \text{otherwise,} \end{cases} \qquad (3.2.1)$$

$$\varphi(t_0; t) = \int_t^{t+1} \chi([t_0, t_0 + 1)|s) ds$$
$$= \begin{cases} 1 - |t - t_0| & t_0 - 1 \leq t \leq t_0 + 1 \\ 0 & \text{otherwise.} \end{cases} \qquad (3.2.2)$$

For an interval $I_i = [i, i+1)(i = 0, 1, 2, \cdots, N-1)$, I write

$$\chi(N_0, \varepsilon; i) = \sum_{j=0}^{n-1} \chi\left(I_i | N_0 + \varepsilon + j \cdot \frac{N}{n}\right). \qquad (3.2.3)$$

Then it leads

$$\chi(N_0, \varepsilon; i)$$
$$= \text{the number of } \blacktriangle - \text{points which belong to } I_i. \qquad (3.2.4)$$

Considering the meaning of χ and φ, we obviously have the following lemmas.

Lemma 1. For any integer j,

$$\chi(N_0, \varepsilon; i \pm j) = \chi(N_0 \mp j, \varepsilon; i). \qquad (3.2.5)$$

Lemma 2.

$$\sum_{i=0}^{N-1} \chi(N_0, \varepsilon; i) \equiv n. \qquad (3.2.6)$$

Lemma 3.

$$\int_0^N \chi(I|t) dt = \int_I 1 dt = |I|,$$
$$\int_0^N \chi(I|t) \chi(I'|t) dt = \int_I \chi(I'|t) dt = |I \cap I'| \qquad (3.2.7)$$

Lemma 4. For any real number s,

$$\varphi(t_0; t \pm s) = \varphi(t_0 \mp s; t). \qquad (3.2.8)$$

Lemma 5.

$$\left\lfloor \frac{n}{N} \right\rfloor \leq \chi(N_0, \varepsilon; i) \leq \left\lceil \frac{n}{N} \right\rceil. \qquad (3.2.9)$$

3.3. Unbiasedness of Sample Mean

Let the population mean and the population variance

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i, \sigma^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \mu)^2 \qquad (3.3.1)$$

respectively.

Given N_0 and ε, the sample is decided. So we can regard the sample as a function of (N_0, ε) and, by the meaning of χ, we have

$$\bar{X}(N_0, \varepsilon) = \frac{1}{n} \sum_{i=0}^{n-1} X_i(N_0, \varepsilon)$$
$$= \frac{1}{n} \sum_{i=0}^{N-1} \chi(N_0, \varepsilon; i) x_i \qquad (3.3.2)$$

as a sample mean. By lemma 1 and 2, we can calculate its expectation as follows.

$$E(\bar{X}) = \int_0^1 \left(\frac{1}{N} \sum_{i=0}^{N-1} \bar{X}(i, \varepsilon) \right) d\varepsilon$$
$$= \int_0^1 \left(\frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1}{n} \sum_{j=0}^{N-1} \chi(i, \varepsilon; j) x_j \right) \right) d\varepsilon$$
$$= \frac{1}{nN} \int_0^1 \left(\sum_{j=0}^{N-1} x_j \left(\sum_{i=0}^{N-1} \chi(i, \varepsilon; j) \right) \right) d\varepsilon$$
$$= \frac{1}{nN} \int_0^1 \left(\sum_{j=0}^{N-1} x_j \left(\sum_{i=0}^{N-1} \chi(0, \varepsilon; j - i) \right) \right) d\varepsilon$$
$$= \frac{1}{nN} \int_0^1 \left(\sum_{j=0}^{N-1} x_j \left(\sum_{i=0}^{N-1} \chi(0, \varepsilon; i) \right) \right) d\varepsilon$$
$$= \frac{1}{nN} \int_0^1 \left(\sum_{j=0}^{N-1} n x_j \right) d\varepsilon$$

$$(3.3.3)$$

$$\left| = \frac{1}{N}\sum_{j=0}^{N-1}\left(x_j\int_0^1 d\varepsilon\right)\right.$$

$$\left| = \frac{1}{N}\sum_{j=0}^{N-1}x_j \right. \qquad (3.3.3)$$

$$\left| = \mu. \right.$$

We hereby get the proof of unbiasedness of sample mean by systematic sampling. As is known from the above process of calculation, ε-term does not contribute at all. So unbiasedness still holds when $\varepsilon \equiv 0$.

Theorem 1. Sample mean by systematic sampling is an unbiased estimator of population mean.

$$E(\bar{X}) = \mu. \qquad (3.3.4)$$

3.4. Variance of Sample Mean

In an identity

$$V(\bar{X}) = E(\bar{X}^2) - \{E(\bar{X})\}^2, \qquad (3.4.1)$$

we have the relation

$$\bar{X}(N_0,\varepsilon) = \frac{1}{n}\sum_{i=0}^{N-1}\chi(N_0,\varepsilon;i)x_i, \qquad (3.4.2)$$

so we have

$$\{\bar{X}(N_0,\varepsilon)\}^2$$
$$= \frac{1}{n^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\chi(N_0,\varepsilon;i)\chi(N_0,\varepsilon;j)x_ix_j. \qquad (3.4.3)$$

The expectation is calculated by

$$E(\bar{X}^2)$$
$$= \frac{1}{n^2 N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\sum_{k=0}^{N-1}\int_0^1\chi(k,\varepsilon;i)\chi(k,\varepsilon;j)d\varepsilon\right)x_ix_j. \qquad (3.4.4)$$

By the definition of χ, we get

$$\sum_{k=0}^{N-1}\int_0^1\chi(k,\varepsilon;i)\chi(k,\varepsilon;j)d\varepsilon$$

$$\left| = \sum_{k=0}^{N-1}\int_0^1\left[\begin{array}{c}\left(\sum_{l=0}^{n-1}\chi\left(I_i|k+\varepsilon+l\cdot\frac{N}{n}\right)\right)\\\left(\sum_{m=0}^{n-1}\chi\left(I_j|k+\varepsilon+m\cdot\frac{N}{n}\right)\right)\end{array}\right]d\varepsilon \right.$$

$$\left| = \sum_{l=0}^{n-1}\sum_{m=0}^{n-1}\left(\sum_{k=0}^{N-1}\int_0^1\left[\begin{array}{c}\chi\left(I_i|k+\varepsilon+l\cdot\frac{N}{n}\right)\\\chi\left(I_j|k+\varepsilon+m\cdot\frac{N}{n}\right)\end{array}\right]d\varepsilon\right)\right.$$

$$\left| = \sum_{l=0}^{n-1}\sum_{m=0}^{n-1}\int_0^N\chi\left(I_i|\varepsilon+l\cdot\frac{N}{n}\right)\chi\left(I_j|\varepsilon+m\cdot\frac{N}{n}\right)d\varepsilon \right. \qquad (3.4.5)$$

Here, if we put

$$j = i+j', m = l+m', \qquad (3.4.6)$$

by lemma 3, we have

$$\int_0^N\chi\left(I_i|\varepsilon+l\cdot\frac{N}{n}\right)\chi\left(I_j|\varepsilon+m\cdot\frac{N}{n}\right)d\varepsilon$$

$$\left| = \left|\left[i-l\cdot\frac{N}{n},i-l\cdot\frac{N}{n}+1\right)\cap\left[j-m\cdot\frac{N}{n},j-m\cdot\frac{N}{n}+1\right)\right|\right.$$

$$\left| = \left|[0,1)\cap\left[j'-m'\cdot\frac{N}{n},j'-m'\cdot\frac{N}{n}+1\right)\right|\right. \qquad (3.4.7)$$

$$\left| = \int_0^N\chi(I_0|\varepsilon)\chi\left(I_{j'}|\varepsilon+m'\cdot\frac{N}{n}\right)d\varepsilon\right.$$

$$\left| = \int_0^1\chi\left(I_{j'}|\varepsilon+m'\cdot\frac{N}{n}\right)d\varepsilon\,d\varepsilon.\right.$$

$$\sum_{k=0}^{N-1}\int_0^1\chi(k,\varepsilon;i)\chi(k,\varepsilon;j)d\varepsilon$$

$$\left| = \sum_{l=0}^{n-1}\sum_{m'=0}^{n-1}\int_0^1\chi\left(I_{j'}|\varepsilon+m'\cdot\frac{N}{n}\right)d\varepsilon\right.$$

$$\left| = n\sum_{m'=0}^{n-1}\int_0^1\chi\left(I_{j'}|\varepsilon+m'\cdot\frac{N}{n}\right)d\varepsilon\right.$$

$$\left| = n\sum_{m'=0}^{n-1}\int_0^1\chi\left(\left[j'-m'\cdot\frac{N}{n},j'-m'\cdot\frac{N}{n}+1\right)|\varepsilon\right)d\varepsilon \right. \quad (3.4.8)$$

$$\left| = n\sum_{m'=0}^{n-1}\varphi\left(j'-m'\cdot\frac{N}{n};0\right)\right.$$

$$\left| = n\sum_{m'=0}^{n-1}\varphi\left(-m'\cdot\frac{N}{n};-j'\right)\right.$$

$$\left| = n\sum_{m'=0}^{n-1}\varphi\left(m'\cdot\frac{N}{n};j'\right).\right.$$

At last, we obtain

$$E(\bar{X}^2) = \frac{1}{nN}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sum_{k=0}^{n-1}\varphi\left(k\cdot\frac{N}{n};j-i\right)x_ix_j \quad (3.4.9)$$

Here I define a function Φ by

$$\Phi(t) = \frac{1}{n}\sum_{i=0}^{n-1}\varphi\left(i\cdot\frac{N}{n};t\right) \qquad (3.4.10)$$

and get the explicit description of the variance of sample mean as

$$V(\bar{X})\left| = E(\bar{X}^2) - \{E(\bar{X})\}^2\right.$$

$$\left| = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\Phi(j-i)x_ix_j - \left(\frac{1}{N}\sum_{i=0}^{N-1}x_i\right)^2\right.$$

$$\left| = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\Phi(j-i)x_ix_j - \left(\frac{1}{N}\sum_{i=0}^{N-1}x_i\right)^2 \right. \quad (3.4.11)$$

$$\left| = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\Phi(j-i) - \frac{1}{N}\right)x_ix_j.\right.$$

Theorem 2. The variance of sample mean by systematic sampling is described as

$$V(\bar{X}) = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\Phi(j-i)-\frac{1}{N}\right)x_i x_j \quad (3.4.12)$$

That a function Φ has a parameter $j-i$ means that $V(\bar{X})$ depends on how to sort a population list.

3.5. Expectation of Sample Variance

The expectation of sample variance

$$\left. S^2(N_0,\varepsilon) \right| = \frac{1}{n}\sum_{i=0}^{n-1}\left(X_i(N_0,\varepsilon)-\bar{X}(N_0,\varepsilon)\right)^2$$

$$\left| = \frac{1}{n}\sum_{i=0}^{n-1}\{X_i(N_0,\varepsilon)\}^2 - \{\bar{X}(N_0,\varepsilon)\}^2 \quad (3.5.1)\right.$$

$$\left| = \frac{1}{n}\sum_{i=0}^{n-1}\chi(N_0,\varepsilon;i)x_i^2 - \{\bar{X}(N_0,\varepsilon)\}^2\right.$$

is calculated, by lemma 1, 2 and 3, as

$$\left. E(S^2) \right| = \frac{1}{n}\sum_{i=0}^{N-1}\left(\int_0^1\left(\frac{1}{N}\sum_{j=0}^{N-1}\chi(j,\varepsilon;i)\right)d\varepsilon\right)x_i^2 - E(\bar{X}^2)$$

$$\left| = \frac{1}{nN}\sum_{i=0}^{N-1}\left(\int_0^1\left(\sum_{j=0}^{N-1}\chi(-i,\varepsilon;-j)\right)d\varepsilon\right)x_i^2 - \left(\begin{array}{c}V(\bar{X})\\+\{E(\bar{X})\}^2\end{array}\right)\right.$$

$$\left| = \frac{1}{nN}\sum_{i=0}^{N-1}\left(\int_0^1\left(\sum_{j=0}^{N-1}\chi(-i,\varepsilon;j)\right)d\varepsilon\right)x_i^2 - \left(V(\bar{X})+\mu^2\right)\right.$$

$$\left| = \frac{1}{nN}\sum_{i=0}^{N-1}\left(\int_0^1 n\,d\varepsilon\right)x_i^2 - \left(V(\bar{X})+\mu^2\right) \quad (3.5.2)\right.$$

$$\left| = \frac{1}{N}\sum_{i=0}^{N-1}x_i^2 - \mu^2 - V(\bar{X})\right.$$

$$\left| = \frac{1}{N}\sum_{i=0}^{N-1}x_i^2 - \mu^2 - V(\bar{X})\right.$$

$$\left| = \sigma^2 - V(\bar{X}).\right.$$

3.6. Conserved Quantity

In general, sample mean and sample variance can be regarded as functions of sample size n. So I write them $\bar{X}(n)$ and $S^2(n)$ respectively from now on. From the result of 3-5, in systematic sampling,

$$V(\bar{X}(n)) + E(S^2(n)) = \sigma^2 \quad (3.6.1)$$

holds. On the other hand, in random sampling with replacement,

$$V(\bar{X}(n)) = \frac{\sigma^2}{n}, E(S^2(n)) = \frac{n-1}{n}\sigma^2, \quad (3.6.2)$$

so we have

$$V(\bar{X}(n)) + E(S^2(n)) = \sigma^2. \quad (3.6.3)$$

In random sampling without replacement, because

$$V(\bar{X}(n)) = \frac{N-n}{n(N-1)}\sigma^2, E(S^2(n)) = \frac{N(n-1)}{n(N-1)}\sigma^2, \quad (3.6.4)$$

we again have

$$V(\bar{X}(n)) + E(S^2(n)) = \sigma^2. \quad (3.6.5)$$

While $\bar{X}(n)$ and $S^2(n)$ are random variables, μ and σ^2 are constants. In general, $V(\bar{X}(n)) + E(S^2(n))$ is considered to depend on sample size n and how to select a sample. So the above results tell us the conserved quantity between random and systematic sampling.

Theorem 3. Between random and systematic sampling, the following quantities are conserved.

$$E(\bar{X}(n)) \equiv \mu, V(\bar{X}(n)) + E(S^2(n)) \equiv \sigma^2 \quad (3.6.6)$$

3.7. Limiting Behavior and Some Other Results

In

$$\left. \Phi(j-i) \right| = \frac{1}{n}\sum_{i=0}^{n-1}\varphi\left(i\cdot\frac{N}{n};j-i\right)$$

$$\left| = \frac{1}{n}\sum_{i=0}^{n-1}\int_{j-i}^{j-i+1}\chi\left(\left[i\cdot\frac{N}{n},i\cdot\frac{N}{n}+1\right)\middle|t\right)dt\right.$$

$$\left| = \frac{1}{n}\sum_{i=0}^{n-1}\int_{j-i}^{j-i+1}\chi\left([0,1)\middle|t-i\cdot\frac{N}{n}\right)dt\right.$$

$$\left| = \frac{1}{n}\sum_{i=0}^{n-1}\int_0^1\chi\left([0,1)\middle|t+(j-i)-i\cdot\frac{N}{n}\right)dt \quad (3.7.1)\right.$$

$$\left| = \frac{1}{n}\int_0^1\left(\sum_{i=0}^{n-1}\chi\left(I_0\middle|t+(j-i)-i\cdot\frac{N}{n}\right)\right)dt\right.$$

$$\left| = \frac{1}{n}\int_0^1\left(\sum_{i=0}^{n-1}\chi\left(I_0\middle|(j-i)+t+i\cdot\frac{N}{n}\right)\right)dt\right.$$

$$\left| = \frac{1}{n}\int_0^1\chi(j-i,t;0)dt\right.$$

from lemma 5, we immediately have inequalities

$$\frac{n}{N}-1 < \left\lfloor\frac{n}{N}\right\rfloor \leq \chi(j-i,t;0) \leq \left\lceil\frac{n}{N}\right\rceil < \frac{n}{N}+1, \quad (3.7.2)$$

$$\frac{1}{N}-\frac{1}{n} < \Phi(j-i) = \frac{1}{n}\int_0^1\chi(j-i,t;0)dt < \frac{1}{N}+\frac{1}{n}, \quad (3.7.3)$$

so we obtain

$$\left|\Phi(j-i)-\frac{1}{N}\right| < \frac{1}{n}. \quad (3.7.4)$$

Applying this estimation to

$$V\left(\bar{X}(n)\right) = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\Phi(j-i) - \frac{1}{N}\right)x_i x_j, \quad (3.7.5)$$

we get an analogy of law of large numbers.

Theorem 4. In systematic sampling, while

$$V\left(\bar{X}(n)\right) + E\left(S^2(n)\right) \equiv \sigma^2 \qquad (3.7.6)$$

is kept,

$$\lim_{n\to\infty} V\left(\bar{X}(n)\right) = 0. \qquad (3.7.7)$$

Moreover, while we obviously have

$$V\left(\bar{X}(N)\right) = 0 \qquad (3.7.8)$$

for random sampling without replacement, we also have

$$V\left(\bar{X}(N)\right) = 0 \qquad (3.7.9)$$

for systematic sampling because, when $n = N$,

$$1 = \left\lfloor \frac{N}{N} \right\rfloor \leqq \chi(j-i,t;0) \leqq \left\lceil \frac{N}{N} \right\rceil = 1 \qquad (3.7.10)$$

and

$$\Phi(j-i) = \frac{1}{N}\int_0^1 \chi(j-i,t;0)dt = \frac{1}{N}. \qquad (3.7.11)$$

That is, $V\left(\bar{X}(n)\right)$ of systematic sampling has both properties of random samplings with and without replacement.

When x is nonnegative and $\mu > 0$, from (3.7.4), we can estimate

$$
\begin{aligned}
\left| V\left(\bar{X}(n)\right) \right| &\leqq \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left|\Phi(j-i) - \frac{1}{N}\right| x_i x_j \\
&< \frac{1}{nN}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x_i x_j \\
&= \frac{1}{nN}\left(\sum_{i=0}^{N-1} x_i\right)^2 \\
&= \frac{1}{nN}(N\mu)^2 \\
&= \frac{N}{n}\mu^2 \\
&= \frac{N}{n}\left\{E\left(\bar{X}(n)\right)\right\}^2.
\end{aligned}
\qquad (3.7.12)
$$

So the coefficient of variation of $\bar{X}(n)$ is less than $\sqrt{N/n}$.

Corollary 1. x When is nonnegative and $\mu > 0$,

$$CV\left(\bar{X}(n)\right) < \sqrt{\frac{N}{n}} \qquad (3.7.13)$$

Applying lemma 5 to

$$\bar{X}(n) = \frac{1}{n}\sum_{i=0}^{N-1}\chi(N_0,\varepsilon;i)x_i, \qquad (3.7.14)$$

we get the following estimation.

$$\left\lfloor \frac{n}{N} \right\rfloor \frac{1}{n}\sum_{i=0}^{N-1}x_i \leqq \bar{X}(n) \leqq \left\lceil \frac{n}{N} \right\rceil \frac{1}{n}\sum_{i=0}^{N-1}x_i \qquad (3.7.15)$$

and

$$\left(\left\lfloor \frac{n}{N} \right\rfloor - \frac{n}{N}\right)\frac{N}{n}\mu \leqq \bar{X}(n) - \mu \leqq \left(\left\lceil \frac{n}{N} \right\rceil - \frac{n}{N}\right)\frac{N}{n}\mu. \qquad (3.7.16)$$

Here, as

$$-1 < \left\lfloor \frac{n}{N} \right\rfloor - \frac{n}{N} \leqq 0, 0 \leqq \left\lceil \frac{n}{N} \right\rceil - \frac{n}{N} < 1, \qquad (3.7.17)$$

we have

$$\left|\bar{X}(n) - \mu\right| < \frac{N}{n}\mu. \qquad (3.7.18)$$

This result suggests the robustness of systematic sampling.

Corollary 2. When x is nonnegative and $\mu > 0$,

$$\frac{\left|\bar{X}(n) - \mu\right|}{\mu} < \frac{N}{n}. \qquad (3.7.19)$$

4. Discussion

Define a vector x and an $N \times N$ matrix $A(n,N)$ as

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}, A(n,N) = \left(\Phi(j-i) - \frac{1}{N}\right)_{i,j=0}^{N-1}. \qquad (4.1)$$

Then $V\left(\bar{X}\right)$ can be regarded as a quadratic form

$$V\left(\bar{X}(n)\right) = \frac{1}{N}x^{\mathsf{T}}A(n,N)x, \qquad (4.2)$$

where $A(n,N)$ s obviously a positive semi definite matrix.

Proposition. For any $n, N \in \{1, 2, \cdots\}$ and any real vector $x \in \mathbb{R}^N$,

$$x^{\mathsf{T}}A(n,N)x \geqq 0 \qquad (4.3)$$

and

$$x^{\mathsf{T}}A(n,N)x = 0 \Leftrightarrow N \mid n \text{ or } x_0 = x_1 = \cdots = x_{N-1}. (4.4)$$

How to sort a population list is nothing but how to permutate a population list mathematically. So, if I write

$$\tau(x) = \begin{pmatrix} x_{\tau(0)} \\ x_{\tau(1)} \\ x_{\tau(2)} \\ \vdots \\ x_{\tau(N-1)} \end{pmatrix} \qquad (4.5)$$

for $\tau \in \mathfrak{S}_N$ where $\mathfrak{S}_N$ is a permutation group of degree N, $V\left(\overline{X}(n)\right)$ can be regarded as a function of $\tau \in \mathfrak{S}_N$ and written as

$$V\left(\tau;\overline{X}(n)\right) = \frac{1}{N}\tau(x)^{\mathsf{T}} A(n,N)\tau(x). \qquad (4.6)$$

Here the next question will arise.

What kind of τ givesless $V\left(\tau;\overline{X}(n)\right)$?

Obviously the value is conserved according to a cyclic permutation. I leave this problem to the readers. Thank you so much for reading my poor English to the end.

Acknowledgement

References

[1]   Leslie Kish, *Survey Sampling*, John Wiley & Sons, Inc., 1965.

[2]   Murthy, M. N., *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta, India, 1967.

[3]   William G. Cochran, *Sampling Techniques*, John Wiley & Sons, Inc., 1977.

[4]   Carl-Erik Särndal, Bengt Swensson and Jan Wretman, *Model Assisted Survey Sampling*, Springer, 1992.

[5]   Sharon L. Lohr., *Sampling: Design and Analysis.* Duxbury Press, 1999.

[6]   Steven K. Thompson, *Sampling*, John Wiley & Sons, Inc., 2002.

[7]   Robert M. Groves, Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau., *Survey Methodology*. John Wiley & Sons, Inc., 2004.

[8]   Lahiri, D. B., "A method for sample selection providing unbiased ratio estimates". Bull. Int. Stat. Inst., 33, 2, 133-140, 1951.