

## An extension of Horvitz–Thompson estimator used in adaptive cluster sampling to continuous universe

Tomasz Bąk

To cite this article: Tomasz Bąk (2017) An extension of Horvitz–Thompson estimator used in adaptive cluster sampling to continuous universe, Communications in Statistics - Theory and Methods, 46:19, 9777-9786, DOI: [10.1080/03610926.2016.1218028](https://doi.org/10.1080/03610926.2016.1218028)

To link to this article: <https://doi.org/10.1080/03610926.2016.1218028>



Published online: 20 Jun 2017.



Submit your article to this journal [↗](#)



Article views: 104



View related articles [↗](#)



View Crossmark data [↗](#)



# An extension of Horvitz–Thompson estimator used in adaptive cluster sampling to continuous universe

Tomasz Bąk

Department of Statistics, Econometrics and Mathematics, University of Economics in Katowice, Katowice, Poland

## ABSTRACT

In this paper, an extension of Horvitz–Thompson estimator used in adaptive cluster sampling to continuous universe is developed. Main new results are presented in theorems. The primary notions of discrete population are transferred to continuous population. First and second order inclusion probabilities for networks are delivered. Horvitz–Thompson estimator for adaptive cluster sampling in continuous universe is constructed. The unbiasedness of the estimator is proven. Variance and unbiased variance estimator are delivered. Finally, the theory is illustrated with an example.

## ARTICLE HISTORY

Received 19 February 2016  
Accepted 25 July 2016

## KEYWORDS

Adaptive sampling;  
continuous universe;  
Horvitz–Thompson  
estimator; spatial sampling.

## MATHEMATICS SUBJECT CLASSIFICATION

62D05

## 1. Introduction to adaptive sampling

The adaptive sampling results from the need of researchers which have not been met by non adaptive sampling schemes. The best way to understand the idea of the adaptive sampling is an analysis of examples. Chaudhuri (2000) took up the problem of exploring locations to find mineral oilfields. In countries which are geographically extensive, like India, it is hard to find sources of oil. On the other hand, sources of fuel are crucial from commercial point of view. By using non adaptive methods, one would probably select a sample with many elements which do not have any signs of oil. The exploration would be of bad efficiency.

Instead of sampling many unexpected elements, one can focus on the selection of the elements which are close to the elements with traces of oil. The same approach was also used in the case of low-density populations of freshwater mussels research Vilella et al. (2003) or in the case of hydroacoustic surveys in oceans Elizabeth Connors and Schwager (2002).

The examples above were not chosen coincidentally. Samples and estimators for these surveys were developed according to the theory constructed by Thompson (1990) and later developed by Thompson and Seber (1996), Chaudhuri (2000), Chaudhuri and Saha (2004), and Chaudhuri et al. (2005). This theory was devoted for finite populations. Therefore, populations such as area of ocean or grounds have to be divided into finite number of elements to adjust them to finite population theory.

However, population such as area of ocean or grounds could be treated as continuous (infinite) population. Theory of sampling from continuous universe was constructed in the '90s; see, for example, De Gruijter and Ter Braak (1990), Cordy (1993), and Brus and de Gruijter (1997). Over the past few years, continuous universe approach has been extensively

considered Marcheselli et al. (2012). Generally, the need for using continuous universe approach for continuous population is noticeable. An example is validation of soil maps Brus et al. (2011). Technically map contains infinite number of points. Sampling from continuous universe gives a possibility to validate values presented on the map more precisely—exactly at the point. Moreover, soil maps could contain information about characteristics which are rare, i.e., information about oil. Therefore, there is a need for sampling scheme which combines continuous and adaptive approaches.

The aim of this paper is to extend the adaptive cluster sampling theory to the continuous (infinite) population. It is delivered through the extension of the Horvitz–Thompson estimator used in adaptive sampling to the continuous universe. This development is conducted in two stages. In Sec. 2, notions of cluster and newtork, which are crucial for adaptive cluster sampling, are transferred to continuous universe. This section is a theoretical introduction to continuous modification of Horvitz–Thompson estimator for adaptive cluster sampling. In Sec. 3, Horvitz–Thompson estimator used in adaptive cluster sampling Thompson (1990) is extended to the continuous population using results delivered by Cordy (1993). As a result, a continuous modification for both concepts of the sampling theory and the estimation method is proposed.

## 2. Continuous adaptive cluster sampling

Sampling design for the continuous population is defined similarly to the common approach of the finite population framework. The difference is that continuous probability distribution is used instead of sampling probabilities.

Let us consider sample of  $n$  points, which are selected from the continuous population  $\mathcal{U} \subset \mathbb{R}^k$ . It is assumed that  $\mathcal{U}$  is bounded open set. The characteristic under study is denoted by  $Y$ . Strictly speaking,  $Y$  is a random process defined as

$$\{Y(x) : x \in \mathcal{U}\} \quad (1)$$

The sample space  $\mathcal{P}$  is the set of ordered samples, i.e.,  $\mathcal{P} = \mathcal{U}^n = \{(x_1, \dots, x_n), \quad x_i \in \mathcal{U}, \quad i = 1, \dots, n\}$ . On sample space  $\mathcal{P}$ , we define probability density function  $f$ . The aim of the estimation - the total value of the  $Y$  characteristic is equal to

$$I_Y = \int_{\mathcal{U}} Y(x) dx \quad (2)$$

The  $f$  function could be considered as a joint probability density function of  $n$  random variables  $X_1, \dots, X_n$ , defined on the  $\mathcal{U}$ . As an analogy to inclusion probability for finite population Cordy (1993) introduced inclusion density function

$$\pi(x) = \sum_{i=1}^n f_i(x) \quad (3)$$

where  $f_i$  denotes the marginal density of  $X_i$ ,  $i = 1, \dots, n$ . Next, let  $f_{ij}$ ,  $i \neq j$ ,  $i = 1, \dots, n$  denotes the joint marginal density of  $X_i$  and  $X_j$ . The pairwise inclusion density function, which is equivalent to second order inclusion probability in finite population, is defined as Cordy (1993)

$$\pi(x, x') = \sum_{i=1}^n \sum_{i \neq j} f_{ij}(x, x') \quad (4)$$

A continuous adaptive cluster sampling is the transformation of the idea of adaptive cluster sampling Thompson (1990) and Thompson and Seber (1996). The transformation is made using theory developed by Cordy (1993). In the first step, we need to extend notions of cluster and network, which are crucial for sampling theory and estimation method, respectively, Thompson (1990). A neighborhood  $A(i)$  of element  $i$  of finite population  $(1, \dots, N)$  is a set composed of the  $i$ th element and other elements in the population which bear a reciprocal unalterable relationship with the element  $i$  Chaudhuri (2014). Suppose that the population under study contains two types of elements. First group contains elements at which value of  $Y$  is zero. Second group contains elements for which realizations of the spatial process  $Y$  are not zero-valued. For a given unit with a positive value of  $Y$ , a cluster is formed by associating with it all the units in its neighborhood and successively extending the association accommodating the units in the neighborhoods of any unit with positive values of characteristic under study and stopping only on encountering units with zero-values of  $Y$  Chaudhuri (2014). These elements with zero  $Y$  values are called the edge-units. The set of units with non zero values of  $Y$  in the cluster of the  $i$ th element is called a network of  $i$ th element. Each element with the zero value of  $Y$  (also the edge-units) is called a singleton network. All “normal” networks and “singleton” networks are disjoint and together cover the whole population.

Let us again discuss the continuous population. In the continuous population, there is no need to distinguish the notions of the cluster and the network. Therefore, one definition is introduced for these two notions. Generally, if the continuous population is considered, the values of  $Y$  characteristic should be described by continuous function. We assume that  $Y$  is continuous function on  $\mathcal{U}$ . On this assumption, let us define network (cluster) for an element  $x_0$  of population  $\mathcal{U}$ . When  $Y(x_0) > 0$ , then network (cluster)  $A(x_0)$  is the biggest connected set which contains  $x_0$  and other elements at which value of  $Y$  is greater than zero. When  $Y(x_0) = 0$ , then network (cluster)  $A(x_0)$  contains all elements at which value of  $Y$  is zero. These networks are disjoint and together cover the whole population - similarly to finite population case.

In adaptive cluster sampling process of sampling is divided into two stages. First stage is the selection of initial sample. Generally, at the initial stage simple random sample with or without replacement are used Chaudhuri (2014). At the second stage all elements from clusters which consist the elements sampled in the first stage are added to sample. Final sample is called adaptive sample Chaudhuri (2014).

Let us consider the adaptive cluster sampling in the continuous population. We can assume that at the initial stage simple random sample with or without replacement are used. Thus, all  $n$  elements of initial sample are selected independently. Therefore, the probability density function which determinates sampling probabilities of different initial samples fulfills  $f = \prod_{i=1}^n f_i$ . At the second stage some modifications have to be introduced, because in the continuous population selection of all elements of the cluster is impossible. Instead of this, sampling inside a cluster is conducted. Different sampling schemes can be used at this stage. An example of the sampling scheme is proposed in Sec. 4.

### 3. Horvitz-Thompson estimator in continuous adaptive cluster sampling

Thompson (1990) and Thompson and Seber (1996) applied the classic approach to the estimation based on sampling with unequal probabilities Horvitz and Thompson (1952) to adaptive cluster sampling. Let us consider  $K$  networks  $A_1, \dots, A_K$  with the cardinalities  $m_1, \dots, m_K$ , resp. Let  $\mathbb{1}_k$ ,  $i = 1, \dots, K$  be the indicator variable defined as

$$\mathbb{1}_k = \begin{cases} 1, & \text{if any element of } k\text{th network is in } s \text{ sample,} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The total value  $y_k^*$  of the  $Y$  characteristics in the network  $k$  is equal to  $y_k^* = \sum_{i \in m_k} y_i$ . The Horvitz–Thompson estimator for adaptive cluster sampling is defined as Thompson and Seber (1996)

$$\hat{\mu}_{HT} = \sum_{k=1}^K \frac{y_k^* \mathbb{1}_k}{\pi_k} \quad (6)$$

where  $\pi_k$  is the inclusion probability of the network  $k$ ,  $k = 1, \dots, K$ . Inclusion probability  $\pi_k$  is a probability that the initial sample intersects network  $k$ ,  $k = 1, \dots, K$ . The estimator is unbiased with variance equal to Thompson and Seber (1996)

$$\text{Var}(\hat{\mu}_{HT}) = \sum_{k=1}^K \sum_{h=1}^K \frac{y_k^* y_h^* (\pi_{kh} - \pi_k \pi_h)}{\pi_k \pi_h}, \quad (7)$$

where  $\pi_{kh}$  is the second order probability of inclusion for networks with indexes  $k$  and  $h$ . Inclusion probability  $\pi_{kh}$  is a probability that the initial sample intersects both networks  $k$  and  $h$ ,  $k, h = 1, \dots, K$ . An unbiased estimator of the variance of the estimator  $\hat{\mu}_{HT}$  is

$$\widehat{\text{Var}}(\hat{\mu}_{HT}) = \sum_{k=1}^K \sum_{h=1}^K \frac{y_k^* y_h^* \mathbb{1}_k \mathbb{1}_h (\pi_{kh} - \pi_k \pi_h)}{\pi_k \pi_h \pi_{kh}} \quad (8)$$

Estimators (6) and (8) are based on the inclusion probabilities defined on the network level. In practice, spatial distribution of  $Y$  characteristics is such that number of areas on which values of  $Y$  is greater than zero is finite. Therefore, number of networks is also finite. In turn, characteristic under study is defined on continuous population. Therefore, process of estimation is connection of estimation on finite and infinite (continuous) sets of elements.

If the population is continuous, probabilities of inclusion are replaced by inclusion density functions. Therefore, probabilities of inclusion of networks have to be replaced too. For each network first order inclusion function is defined as

$$\pi_k = \int_{x \in A_k} \pi(x) dx, \quad k = 1, \dots, K \quad (9)$$

The second-order inclusion functions are defined as

$$\pi_{kh} = \int_{x \in A_k} \int_{x' \in A_h} \pi(x, x') dx dx', \quad k, h = 1, \dots, K, \quad k \neq h \quad (10)$$

Using (9), a Horvitz–Thompson estimator for adaptive cluster sampling in continuous population can be defined as

$$\hat{I}_Y = \sum_{k=1}^K \frac{y_k^* \mathbb{1}_k}{\pi_k} \quad (11)$$

where  $\mathbb{1}_k$  is defined as in (5) and  $y_k^* = \int_{x \in A_k} Y(x) dx$ ,  $k = 1, \dots, K$ .

**Theorem 1.** *If the first-order inclusion function fulfills  $\pi_k > 0$ ,  $k = 1, \dots, K$ , then  $\hat{I}_Y$  is unbiased estimator of total value of  $Y$  characteristics with finite values.*

**Proof.**

$$\begin{aligned}
 E_{\mathcal{P}}(\hat{I}_Y) &= E_{\mathcal{P}}\left(\sum_{k=1}^K \frac{y_k^* \mathbb{1}_k}{\pi_k}\right) = \int_{\mathcal{P}} \sum_{k=1}^K \frac{y_k^* \mathbb{1}_k}{\pi_k} f(s) dx_1 \dots dx_n \\
 &= \int_{\mathcal{U}} \dots \int_{\mathcal{U}} \sum_{k=1}^K \frac{y_k^* \mathbb{1}_k}{\pi_k} f(x_1, \dots, x_n) dx_1 \dots dx_n \\
 &= \int_{\mathcal{U}} \dots \int_{\mathcal{U}} \sum_{i=1}^n \frac{y_{k_i}^*}{\pi_{k_i}} f(x_1, \dots, x_n) dx_1 \dots dx_n \\
 &= \sum_{i=1}^n \int_{\mathcal{U}} \dots \int_{\mathcal{U}} \frac{y_{k_i}^*}{\pi_{k_i}} f(x_1, \dots, x_n) dx_1 \dots dx_n \\
 &= \sum_{i=1}^n \int_{\mathcal{U}} \frac{y_{k_i}^*}{\pi_{k_i}} f_i(x) dx
 \end{aligned}$$

Values of  $Y$  characteristics in the population are limited. Furthermore, for the clarity of calculation, let us introduce  $y^* = \sum_{k=1}^K \mathbb{1}_{A_k} y_k^*$  and  $\pi^* = \sum_{k=1}^K \mathbb{1}_{A_k} \pi_k$ . Therefore, by applying Fubini's theorem Rudin (1974) and using definition (9), we obtain

$$\begin{aligned}
 \sum_{i=1}^n \int_{\mathcal{U}} \frac{y_{k_i}^*}{\pi_{k_i}} f_i(x) dx &= \sum_{i=1}^n \int_{\mathcal{U}} \frac{y^*}{\pi^*} f_i(x) dx = \int_{\mathcal{U}} \frac{y^*}{\pi^*} \pi(x) dx \\
 &= \sum_{k=1}^K \int_{A_k} \frac{y_k^*}{\pi_k} \pi(x) dx = \sum_{k=1}^K \frac{y_k^*}{\pi_k} \int_{A_k} \pi(x) dx \\
 &= \sum_{k=1}^K \int_{A_k} Y(x) dx = \int_{\mathcal{U}} Y(x) dx = I_Y
 \end{aligned}$$

□

**Theorem 2.** Let us assume that values of  $Y$  are limited. If the first order inclusion function fulfills  $\pi_k > 0$ ,  $k = 1, \dots, K$ , then the variance of  $\hat{I}_Y$  is equal to

$$\text{Var}(\hat{I}_Y) = \sum_{k=1}^K \frac{y_k^{*2}}{\pi_k} + \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K y_k^* y_h^* \frac{\pi_{kh} - \pi_k \pi_h}{\pi_k \pi_h} \quad (12)$$

Moreover, if  $\pi_{kh} > 0$  for each  $k, h = 1, \dots, K$ ,  $k \neq h$ , then the unbiased estimator of  $\text{Var}(\hat{I}_Y)$  is given by

$$\widehat{\text{Var}}(\hat{I}_Y) = \sum_{k=1}^K \frac{(\mathbb{1}_k y_k^*)^2}{\pi_k^2} + \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \mathbb{1}_k \mathbb{1}_h y_k^* y_h^* \left( \frac{\pi_{kh} - \pi_k \pi_h}{\pi_{kh} \pi_k \pi_h} \right) \quad (13)$$

**Proof.** The variance of  $\hat{I}_Y$  is given by  $\text{Var}(\hat{I}_Y) = E(\hat{I}_Y^2) - (E(\hat{I}_Y))^2$ . The first element of the difference on the right-hand side of the equation can be written as

$$\begin{aligned} \mathbb{E}(\hat{I}_Y^2) &= \int_{\mathcal{P}} \left( \sum_{k=1}^K \frac{y_k^* \mathbb{1}_k}{\pi_k} \right)^2 f(s) dx_1 \dots dx_n = \int_{\mathcal{P}} \sum_{k=1}^K \frac{y_k^{*2} \mathbb{1}_k^2}{\pi_k^2} f(s) dx_1 \dots dx_n \\ &\quad + \int_{\mathcal{P}} \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^* \mathbb{1}_k \mathbb{1}_h}{\pi_k \pi_h} f(s) dx_1 \dots dx_n \end{aligned}$$

The first element of the sum on the right-hand side is

$$\begin{aligned} \int_{\mathcal{P}} \sum_{k=1}^K \frac{y_k^{*2} \mathbb{1}_k^2}{\pi_k^2} f(s) dx_1 \dots dx_n &= \int_{\mathcal{U}} \dots \int_{\mathcal{U}} \sum_{k=1}^n \frac{y_{k_i}^{*2}}{\pi_{k_i}^2} f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \sum_{k=1}^n \int_{\mathcal{U}} \frac{y^{*2}}{\pi^{*2}} f_i(x) dx = \int_{\mathcal{U}} \frac{y^{*2}}{\pi^{*2}} \pi(x) dx = \sum_{k=1}^K \int_{A_k} \frac{y_k^{*2}}{\pi_k^2} \pi(x) dx \\ &= \sum_{k=1}^K \frac{y_k^{*2}}{\pi_k^2} \int_{A_k} \pi(x) dx = \sum_{k=1}^K \frac{y_k^{*2}}{\pi_k} \end{aligned}$$

To calculate the second element of the sum we will use Fubini's theorem and [Theorem 1](#)

$$\begin{aligned} &\int_{\mathcal{P}} \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^* \mathbb{1}_k \mathbb{1}_h}{\pi_k \pi_h} f(s) dx_1 \dots dx_n \\ &= \int_{\mathcal{U}} \dots \int_{\mathcal{U}} \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^* \mathbb{1}_k \mathbb{1}_h}{\pi_k \pi_h} f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \int_{\mathcal{U}} \dots \int_{\mathcal{U}} \frac{y_k^* y_h^* \mathbb{1}_k \mathbb{1}_h}{\pi_k \pi_h} f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq k}}^n \int_{\mathcal{U}} \int_{\mathcal{U}} \frac{y_{k_i}^* y_{h_j}^*}{\pi_{k_i} \pi_{h_j}} f_{k_i h_j}(x, x') dx dx' = \int_{\mathcal{U}} \int_{\mathcal{U}} \frac{y^*(x) y^*(x')}{\pi^*(x) \pi^*(x')} \pi(x, x') dx dx' \\ &= \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \int_{A_k} \int_{A_h} \frac{y_k^* y_h^*}{\pi_k \pi_h} \pi(x, x') dx' dx \\ &= \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^*}{\pi_k \pi_h} \int_{A_k} \int_{A_h} \pi(x, x') dx' dx = \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^*}{\pi_k \pi_h} \pi_{kh}. \end{aligned}$$

From [Theorem 1](#) we obtain

$$\begin{aligned} (\mathbb{E}_{\mathcal{P}}(\hat{I}_Y))^2 &= \left( \int_{\mathcal{U}} Y(x) dx \right)^2 = \int_{\mathcal{U}} Y(x) dx \int_{\mathcal{U}} y(x') dx' \\ &= \sum_{k=1}^K y_k^* \sum_{\substack{h=1 \\ h \neq k}}^K y_h^* = \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K y_k^* y_h^*. \end{aligned}$$

Finally,

$$\begin{aligned}\text{Var}(\hat{I}_Y) &= \sum_{k=1}^K \frac{y_k^{*2}}{\pi_k} + \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^*}{\pi_k \pi_h} \pi_{kh} - \sum_{k=1}^K \sum_{h=1}^K y_k^* y_h^* \\ &= \sum_{k=1}^K \frac{y_k^{*2}}{\pi_k} + \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K y_k^* y_h^* \frac{\pi_{kh} - \pi_k \pi_h}{\pi_k \pi_h}.\end{aligned}$$

To prove the unbiasedness of (13) we calculate

$$\begin{aligned}\mathbb{E}(\widehat{\text{Var}}(\hat{I}_Y)) &= \mathbb{E}\left(\sum_{k=1}^K \frac{(\mathbb{1}_k y_k^*)^2}{\pi_k^2} + \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \mathbb{1}_k \mathbb{1}_h y_k^* y_h^* \left(\frac{\pi_{kh} - \pi_k \pi_h}{\pi_k \pi_h}\right)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^n \frac{y_{k_i}^{*2}}{\pi_{k_i}^2}\right) + \mathbb{E}\left(\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n y_{k_i}^* y_{h_j}^* \left(\frac{\pi_{k_i h_j} - \pi_{k_i} \pi_{h_j}}{\pi_{k_i} \pi_{h_j}}\right)\right)\end{aligned}$$

To calculate the first expected value let  $z_k = \frac{y_k^{*2}}{\pi_k}$ ,  $k = 1, \dots, K$ . Then, by analogous calculations as above, we obtain

$$\mathbb{E}\left(\sum_{i=1}^n \frac{(\mathbb{1}_{k_i} y_{k_i}^*)^2}{\pi_{k_i}^2}\right) = \mathbb{E}\left(\frac{z_{k_i}}{\pi_{k_i}}\right) = \sum_{k=1}^K z_k = \sum_{k=1}^K \frac{y_k^{*2}}{\pi_k}$$

The second expected value can be expressed as

$$\mathbb{E}\left(\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n y_{k_i}^* y_{h_j}^* \left(\frac{\pi_{k_i h_j} - \pi_{k_i} \pi_{h_j}}{\pi_{k_i} \pi_{h_j}}\right)\right) = \mathbb{E}\left(\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{y_{k_i}^* y_{h_j}^*}{\pi_{k_i} \pi_{h_j}}\right) - \mathbb{E}\left(\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{y_{k_i}^* y_{h_j}^*}{\pi_{k_i} \pi_{h_j}}\right)$$

The first expected value was calculated before and is equal to

$$\sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K \frac{y_k^* y_h^*}{\pi_k \pi_h} \pi_{kh}$$

The calculations of the expected value of the second term are similar to the calculations performed for the first term. As a result, we obtain

$$\mathbb{E}\left(\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{y_{k_i}^* y_{h_j}^*}{\pi_{k_i} \pi_{h_j}}\right) = \sum_{k=1}^K \sum_{\substack{h=1 \\ h \neq k}}^K y_k^* y_h^*$$

Combining results of three expected values, we find the unbiasedness of  $\widehat{\text{Var}}(\hat{I}_Y)$  □



#### 4. An example

Let us consider two-dimensional continuous population  $X_1 \times X_2 = [0, 5] \times [0, 5]$  and define the variable  $Y$  as

$$\begin{aligned} Y(x_1, x_2) = & \max(0.2 - (x_1 - 0.7)^2 - (x_2 - 0.4)^2, 0) \\ & + \max(0.2 - (x_1 - 1.55)^2 - (x_2 - 4.45)^2, 0) \\ & + \max(0.2 - (x_1 - 2.45)^2 - (x_2 - 2.2)^2, 0) \\ & + \max(0.2 - (x_1 - 3.9)^2 - (x_2 - 4.1)^2, 0) \end{aligned}$$

All the elements in the population are gathered in five networks, four of them collect elements with non zero values of characteristic under study. The size of each of the “non zero network” is equal to  $0.2\pi$ .

A sampling without replacement was conducted on this population. A sample of 100 elements was obtained. This sample was treated as initial sample for adaptive cluster sampling. In that case, inclusion density function (3) is equal to  $\pi(x) = 4$ ,  $x \in [0, 5] \times [0, 5]$  and pairwise inclusion density function (4) is equal to  $\pi(x, x') = 15.84$ ,  $x, x' \in [0, 5] \times [0, 5]$ . Furthermore, first order inclusion function for network (9) with non zero values of the characteristic under study is equal to  $\pi_k = 0.8\pi$ ,  $k = 1, \dots, 4$  and second-order inclusion function for networks (10) with non zero values of the characteristic under study is equal to  $\pi_{kh} = 0.6336\pi^2$ ,  $k, h = 1, \dots, 4$ ,  $k \neq h$ .

At the next step, for each  $(x_{1j}, x_{2j})$  for which  $Y(x_{1j}, x_{2j}) > 0$  additional 20 elements are sampled using the following probability density function:

$$f_j(x_1, x_2) = \begin{cases} (\alpha_j \text{dist}_j(x_1, x_{1j}, x_2, x_{2j}))^{-1}, & \text{when } x_1 \neq x_{1j} \text{ or } x_2 \neq x_{2j}, \\ 0, & \text{when } x_1 = x_{1j} \text{ and } x_2 = x_{2j} \end{cases} \quad (14)$$

where  $\alpha_j = \int_{x_1, x_2} \text{dist}_j(x_1, x_{1j}, x_2, x_{2j})^{-1} dx_1 dx_2$ , where  $\text{dist}_j$  is defined as

$$\text{dist}_j(x_1, x_{1j}, x_2, x_{2j}) = \begin{cases} 1 / ((x_1 - x_{1j})^2 + (x_2 - x_{2j})^2) & \text{when } x_1 \neq x_{1j} \text{ or } x_2 \neq x_{2j} \\ 0, & \text{when } x_1 = x_{1j} \text{ and } x_2 = x_{2j} \end{cases} \quad (15)$$

Generally, the  $\text{dist}_j$  functions, which are used to build probability density functions, have to satisfy the definition of distance function (in other words, have to satisfy the conditions of non negativity, identity of indiscernibles, symmetry, and subadditivity). Moreover, functions  $\text{dist}_j^{-1}$  have to be integrable B  k (2014).

According to this sampling plan, 180 elements were sampled: 100 elements as the initial sample and additional 80 elements in the second stage of the sampling. At the next step, an interpolation of the  $Y$  characteristic in each “non zero network” was conducted. For this purpose, bilinear interpolation algorithm built in R package *akima* Akima (1996) was used. On the basis of the interpolation, the total values of the  $Y$  characteristic in each cluster were calculated.

Total value of the  $Y$  characteristic in population was equal to 0.115664. Modified Horvitz–Thompson estimator (11) based on interpolated values of  $Y$  delivered value equal to 0.104711. Variance of the estimator was calculated using (13). It was equal to 0.002706. Distribution of the  $Y$  characteristic and interpolated distribution of the  $Y$  are presented in Figures 1 and 2.

Sampling approach is designed for this example. The aim of the example is rather to illustrate theoretical notions than to give general sampling approach. However, sampling using probability density function based on distance function (14) was considered in practice B  k (2014).

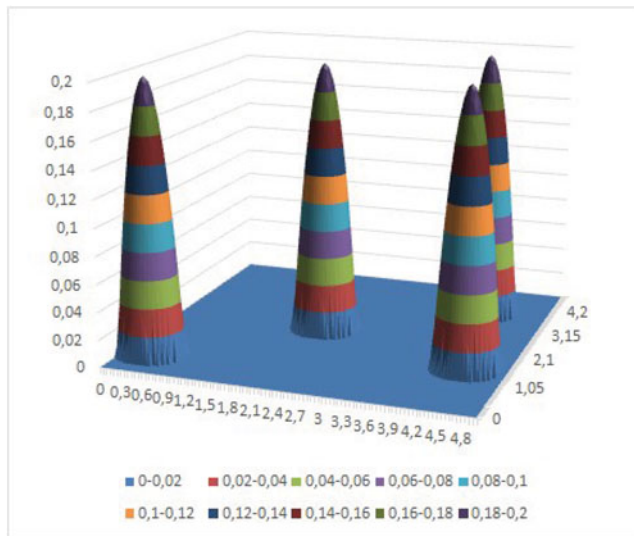


Figure 1. Y characteristic distribution.

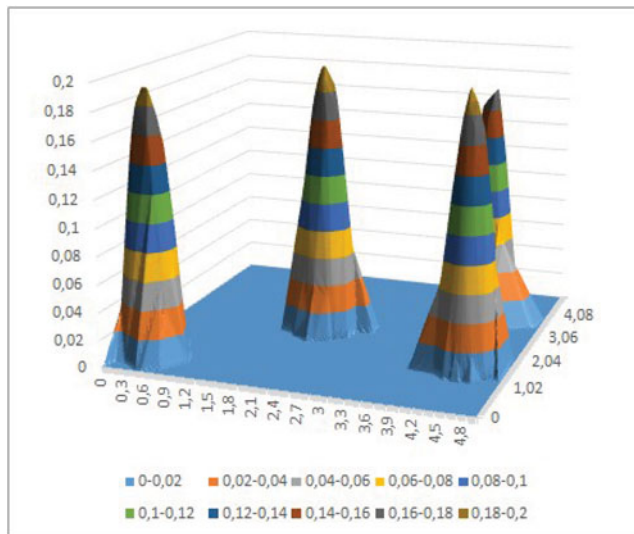


Figure 2. Y characteristic distribution from interpolation.

## References

- Akima, H. (1996). Algorithm 761: Scattered-data surface fitting that has the accuracy of a cubic polynomial. *ACM Trans. Math. Software* 22:362–371.
- Bak, T. (2014). Triangular method of spatial sampling. *Stat. Trans.* 15(2):9–22.
- Brus, D.J., de Gruijter, J.J. (1997). Random sampling or geostatistical modeling? Choosing between design-based and model-based sampling strategies for soil. *Geoderma* 80:1–44.
- Brus, D.J., Kempen, B., Heuvelink, G.B.M. (2011). Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* 62:394–407.
- Chaudhuri, A. (2000). Network and adaptive sampling with unequal probabilities. *Cal. Stat. Assoc. Bull.* 50:238–253.
- Chaudhuri, A. (2014). *Network and Adaptive Sampling*. First edition. Boca Raton: CRC Press.
- Chaudhuri, A., Bose, M., Dihidar, K. (2005). Sample-size restrictive adaptive sampling: An application in estimating localized elements. *J. Stat. Plann. Inference* 134:254–267.

- Chaudhuri, A., Saha, A. (2004). On adaptive cluster sampling and generalized regression method of estimating localized elements. *J. Probab. Stat. Sci.* 2(1):35–45.
- Cordy, C.B. (1993). An extension of the horvitz-thompson theorem to point sampling from a continuous universe. *Stat. Probab. Lett.* 18:353–362.
- De Gruijter, J.J., Ter Braak, C.J.F. (1990). Model-free estimation from spatial samples: A reappraisal of classical sampling theory. *Math. Geol.* 22:407–415.
- Elizabeth Conners, M., Schwager, S.J. (2002). The use of adaptive cluster sampling for hydroacoustic surveys. 59(6):1314–1325.
- Horvitz, D.G., Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.* 47:663–685.
- Marcheselli, M., Barabesi, L., Franceschi, S. (2012). Properties of design-based estimation under stratified spatial sampling with application to canopy coverage estimation. *Ann. Appl. Stat.* 6(1):210–228.
- Rudin, W. (1974). *Real and Complex Analysis*. New York: McGraw-Hill.
- Thompson, S.K., Seber, G.A.F. (1996). *Adaptive Sampling*. New York, USA: Wiley.
- Thompson, S.K. (1990). Adaptive cluster sampling. *J. Am. Stat. Assoc.* 85(412):1050–1059.
- Villella, R., Smith, D., Lemarie, D. (2003). Application of adaptive cluster sampling to low-density populations of freshwater mussels. *Environ. Ecol. Stat.* 10:7–15.