



Balanced sampling of boxes from batches for assessing quality of fruits and vegetables in EU countries

Sara Franceschi¹ · Gianni Betti¹ · Lorenzo Fattorini¹ · Francesca Gagliardi¹ · Gianni Montrone²

Accepted: 21 September 2021 / Published online: 2 October 2021
© The Author(s) 2021

Abstract

The best evaluation for the proportion of defective units in a batch of fruits and vegetables can be achieved by an exhaustive checking of all the boxes in the batch, that is prohibitive to perform in most cases. Usually, only a sample of boxes is checked. In EU countries, EU regulations establish to estimate the proportion of defective units in a batch by the proportion of defective units in the sample, without giving any rule for selecting boxes. Therefore, results are highly dependent on the subjective choice of boxes. In the present study, an objective design-based approach is considered to select boxes from batches, adopting balanced spatial schemes with equal inclusion probabilities. The schemes are able to select samples of boxes evenly spread throughout the batch also ensuring good statistical properties for the proportion of defective units in the sample as estimator of the proportion of defective units in the batch. The performance of these strategies is evaluated by means of a simulation study performed on real and artificial batches of apples, peppers and strawberries. A case study is considered to estimate the proportion of defective units in a batch of courgettes stored in a distribution center of a supermarket chain in Central Italy.

Keywords Equal probability sampling · EU regulation · Horvitz-Thompson estimator · Monte Carlo simulation · Sample proportion of defective units · Spatial balance

1 Introduction

The distribution center of a supermarket chain has the aim of receiving and storing goods, performing a daily distribution of them to sale points. Owing to deteriorating propension of fruit and vegetables, there is the necessity of ensuring the quality of these products by suitable control protocols to be performed on the batches stored in the warehouse before their distribution to sale points. Product controls are essential to guarantee high qualitative standards, to reduce complaints and to construct defect indexes ruling relationships with providers.

✉ Sara Franceschi
sara.franceschi@unisi.it

¹ Department of Economics and Statistics, University of Siena, Siena, Italy

² Quality Assurance, Conad del Tirreno, Pistoia, Italy

In most cases, controls are performed at two levels. The first-level control consists on a visual checking of batches without any destructive testing of fruits and vegetables and of their packages. The second-level control consists on a quantitative assessment of deterioration for those batches that do not pass the first-level control. More precisely, the second-level control consists in determining the fraction of fruit and vegetable units showing characteristics that do not comply with current legislations, henceforth referred to as defective units. Regarding the EU countries, unit defects are those mentioned in the EU Reg 543/2011 for all fruits and vegetables ruled by CMO codification or those not satisfying UNECE standards for fruits and vegetables not ruled by EU Reg 543/2011. In practice, defects of fruits and vegetables are determined by: i) presence of molds and rots; ii) absence of the main characteristics of the product; iii) size outside the established range. The occurrence of at least one of these defects is sufficient to declare the unit as defective.

From a procedural point of view, it is obvious that the second-phase control cannot be performed for all the items in the stored batch, as controls entails item destruction in most cases. Therefore, controls can be performed only for a subset/sample of units. In particular, regarding fruits and vegetables supplied in boxes, EU Reg 543/2011 establishes strict guidelines for performing the control: a fixed number of boxes to be checked is determined as function of the total number of boxes in the batch (see Table 1); all the units in the boxes is checked and classified as defective or not; the fraction of defective units is reckoned and the batch is refused if the fraction is greater than 10%.

Despite the strictness of these rules, surprisingly EU Reg 543/2011 does not give any detail about the way in which the fixed number of boxes have to be selected. It simply states that the selected boxes must be evenly distributed (both in vertical and horizontal sense) throughout the pallets they are located. Therefore, the fractions of defective units achieved by these rules strictly depend on the subjective selection of boxes to check, precluding any objective evaluation about their accuracy and precision and about the subsequent acceptance or rejection of batches.

For these reasons, we were recently involved by managers of the Distribution Center of Conad del Tirreno, a large supermarket chain settled in Central Italy, to construct a statistically sound procedure for estimating fractions of defective units with the possibility of estimating sampling errors and providing confidence intervals but, at the same time, respecting the guidelines established by EU Reg 543/2011 regarding the number of boxes to select, their even distribution throughout batches, and the estimation criterion.

The novel proposal of this work, never pursued before in the framework of fruits and vegetables quality control, is to adopt probabilistic sampling schemes of boxes able to ensure balanced samples with respect to box positions in the batch, in such a way that the fraction of defective units in the sample can constitute a design-based unbiased estimator of the fraction of defective units in the batch with a design-based variance that can be conservatively estimated in order to provide a cautionary estimation of the precision. A simulation study is

Table 1 Number of boxes to be checked as function of the number of boxes in the batch in accordance with EU Reg 543/2011

| Total number of boxes (N) | Number of boxes to be checked (n) |
|-------------------------------|---------------------------------------|
| $N \leq 100$ | 5 |
| $100 < N \leq 300$ | 7 |
| $300 < N \leq 500$ | 9 |
| $500 < N \leq 1000$ | 10 |
| $N > 1000$ | 11 |

performed starting from data from some real batches of fruits and vegetables in order to assess the power of the rejection rule established by EU Reg 543/2011.

Details about sampling boxes from a batch, and the description of the sampling schemes mentioned in the paper are given in the Appendix, together with some practical examples. Data adopted in the simulation study are reported in a supplementary material file.

2 Statement of the problem

The target population U is constituted by a batch of N boxes of fruits or vegetables (see Fig. 4 in the Appendix). Then, denote by y_j the number of fruit or vegetable units in the box j and by d_j the number of defective units in the same box. Let the fraction of defective units in the whole batch

$$P = \frac{\sum_{j \in U} d_j}{\sum_{j \in U} y_j} = \frac{D}{Y}$$

be the target parameter to be estimated from a sample S of n boxes selected from U .

Following the EU Reg 543/2011 guidelines, once the sample S is selected, all the fruit or vegetable units in each selected box are exhaustively examined to count the total of units y_j and the defective units d_j in the box $j \in S$. Then, the unknown proportion P has to be estimated by means of the proportion of deteriorated items observed in the whole sample, i.e. by means of the estimator

$$\hat{P}_{EU} = \frac{\sum_{j \in S} d_j}{\sum_{j \in S} y_j} \quad (1)$$

irrespective of the sampling scheme adopted for selecting boxes.

On the other hand, if boxes were selected by means of a without-replacement sampling scheme, this scheme univocally induces the first-order inclusion probabilities π_j , i.e., for each $j \in U$, the probability that the box j enters the sample, as well as the second-order inclusion probabilities $\pi_{j,h}$, i.e., for each $h \neq j \in U$, the probability that boxes j and h both enter the sample. Therefore, the unknown totals D and Y could be estimated exploiting the first-order inclusion probabilities of the selected boxes by means of the Horvitz-Thompson (HT) estimators \hat{D}_{HT} and \hat{Y}_{HT} , respectively, and the proportion P could be estimated in a very natural way by means of their ratio, i.e.

$$\hat{P}_{HT} = \frac{\hat{D}_{HT}}{\hat{Y}_{HT}} = \frac{\sum_{j \in S} \frac{d_j}{\pi_j}}{\sum_{j \in S} \frac{y_j}{\pi_j}} \quad (2)$$

Being the ratio of two estimators, \hat{P}_{HT} has an intractable mean and variance. However using Taylor series linearization (Särndal et al., 1992, Result 5.6.2, p.178), \hat{P}_{HT} constitutes an approximately unbiased estimator for P , with approximated variance

$$AV(\hat{P}_{HT}) = \frac{1}{Y^2} \sum_{j \in U} \sum_{h > j \in U} (\pi_j \pi_h - \pi_{j,h}) \left(\frac{d_j - P y_j}{\pi_j} - \frac{d_h - P y_h}{\pi_h} \right)^2 \quad (3)$$

It is worth noting that for some fruits and vegetables (e.g. melons, cabbages, etc) there is an equal number of units stored in the boxes, i.e. $y_j = y_0$ for each $j \in U$. In these cases, the total number of units Y is known to be Ny_0 while the total number of units in the sample is known to be ny_0 . Accordingly, the proportion of defective units in the sample becomes

$$\hat{P}_{EU} = \frac{\sum_{j \in S} d_j}{ny_0} \tag{4}$$

while the estimator of P achieved exploiting the first-order inclusion probabilities becomes

$$\hat{P}_{HT} = \frac{\sum_{j \in S} \frac{d_j}{\pi_j}}{Y} = \frac{\hat{D}_{HT}}{Ny_0} \tag{5}$$

Because the estimator (5) is simply a linear transformation of the HT estimator \hat{D}_{HT} , it is unbiased with a well-known variance expression that is given by

$$V(\hat{P}_{HT}) = \frac{1}{N^2 y_0^2} \sum_{j \in U} \sum_{h > j \in U} (\pi_j \pi_h - \pi_{j,h}) \left(\frac{d_j}{\pi_j} - \frac{d_h}{\pi_h} \right)^2 \tag{6}$$

(e.g., Hedayat and Sinha , 1991, equation 3.4, p.48). It is apparent from the previous expressions that the estimators (2) and (5) coincide with the EU estimation rules (1) and (4), respectively, when boxes are selected with constant first-order inclusion probabilities $\pi_j = n/N$. Practically speaking, in order to provide a statistically sound estimation strategy that at the same time fulfils EU estimation guidelines, the sampling scheme adopted to select boxes should ensure the same first-order inclusion probabilities for all boxes in the batch.

3 Problems in performing balanced, equal-probability sampling of boxes

The simplest way to draw n boxes out of the N in the batch with constant first-order inclusion probabilities equal to n/N is to use simple random sampling without replacement, usually referred to as SRSWOR (see Appendix for a description of the scheme). In this case, from equation (3), the approximated variance of (1) reduces to

$$AV(\hat{P}_{EU}) = \frac{N(N-n)}{Y^2} \frac{S_e^2}{n} \tag{7}$$

where

$$S_e^2 = \frac{1}{N-1} \sum_{j \in U} (d_j - Py_j)^2$$

is the variance of the errors occurred in the population predicting the d_j s as Py_j . Usually (6) is estimated by

$$\hat{V}_{SRSWOR}^2 = \frac{N(N-n)}{\hat{Y}_{HT}^2} \frac{s_e^2}{n} = \frac{N-n}{N\bar{y}^2} \frac{s_e^2}{n} \quad (8)$$

where

$$\bar{y} = \frac{1}{n} \sum_{j \in S} y_j$$

is the sample mean, and

$$s_e^2 = \frac{1}{n-1} \sum_{j \in S} (d_j - \hat{P}_{EU} y_j)^2$$

is the variance of the errors occurred in the sample predicting the d_j s as $\hat{P}_{EU} y_j$ (e.g. Särndal et al., 1992, Example 5.6.2, p.179). When boxes in the batch contain an equal number of units, from equation (6) the variance of (4) reduces to

$$V(\hat{P}_{EU}) = \frac{N-n}{N y_0^2} \frac{S_d^2}{n} \quad (9)$$

where

$$S_d^2 = \frac{1}{N-1} \sum_{j \in U} (d_j - \bar{D})^2$$

is the variance of the number of defective units in the boxes of the population and $\bar{D} = D/N$ is their mean. The very familiar estimator of (9) is given by

$$\hat{V}_{SRSWOR}^2 = \frac{N-n}{N y_0^2} \frac{s_d^2}{n} \quad (10)$$

where

$$s_d^2 = \frac{1}{n-1} \sum_{j \in S} (d_j - \bar{d})^2$$

is the variance of the number of defective units in the boxes of the sample and

$$\bar{d} = \frac{1}{n} \sum_{j \in S} d_j$$

is their mean.

Despite its simplicity, SRSWOR does not completely accomplish EU guidelines that recommend an even distribution of the selected boxes throughout the batch. In order to achieve a balanced selection, it should be pointed out that the population U of N boxes is partitioned into K pallets U_1, \dots, U_K and, in turn, each pallet U_k is constituted by L_k levels $U_{k,1}, \dots, U_{k,L_k}$ each constituted by 4 or 6 boxes. In some cases, L_k may be equal for all the pallets in the batch, but this is not true in general (see e.g., the case study).

Generally speaking, the sampling scheme to be adopted should be able to ensure a well-spread distribution of boxes throughout pallets and levels. Indeed, it could be

likely that nearby boxes or boxes belonging to the same pallet contain a similar proportion of deteriorated items than boxes far apart, as well as boxes at lower levels in the pallets contain more deteriorated pieces than boxes at higher levels due, for example, to higher moisture expositions. On the other hand, SRSWOR can even provide samples of boxes all belonging to the same pallet or to the same level (see Fig. 5 in the Appendix). For these reasons, SRSWOR cannot constitute a solution and it is simply considered as a benchmark to determine the gain achieved when using balanced schemes.

A balanced selection could be readily ensured if we were able to partition the population U into n blocks/strata of contiguous levels, in such a way that each block belonged to the same pallet and was constituted by the same number of boxes. In this case, balance was simply achieved by using systematic sampling or one-per-stratum sampling (Breidt, 1995), i.e. selecting a box in a block and repeating its position in the remaining blocks, or randomly and independently selecting a box in each block. Unfortunately, in most situations, the partition of population into n blocks constituted by an equal number of contiguous levels belonging to the same pallet is impossible. The issue is considered in the Appendix, where it is exemplified as the systematic selection of boxes with equal first-order inclusion probabilities is always possible but at the cost of allowing the sample size to vary from $n - 1$ to n and the first-order inclusion probabilities to be different from n/N . Similarly in the Appendix it is also shown as the one-per-stratum sampling of n boxes is always possible but at the cost of allowing the first-order inclusion probabilities to vary with blocks. As a consequence, the subsequent estimators (2) and (5) differ from that requested by EU regulations.

4 Balanced spatial sampling schemes as solution

Grafström and Tillé (2013) propose the use of the doubly balanced spatial scheme (DBSS) obtained combining a generalization of the local pivotal method (LPM) by Grafström et al. (2012) and the cube method by Deville and Tillé (2004). DBSS is tailored to well-spread samples with pre-fixed first-order inclusion probabilities. While the generalized LPM ensures spatially balanced samples avoiding the selection of neighbouring units, the cube method ensures nearly balanced samples with respect to M auxiliary variables, i.e. variables whose values are known for each unit in the population. Like in the generalized regression estimator, the cube method ensures that the totals of these variables estimated from the sample are almost the same as in the population (see the Appendix for a more accurate description of LPM, cube method and DBSS).

As stated in the Appendix, in order to obtain samples of fixed size n , the first-order inclusion probabilities of points must be chosen as an auxiliary variable, in such a way that their sum equals n . Moreover, regarding the choice of the other balancing variables, Grafström and Tillé (2013) suggest to use the unit coordinates. In this case, DBSS not only provides samples of units well spread throughout the population, but the sample estimates of the average coordinates are similar to the true averages in the population. In this sense, the sampling is doubly balanced.

From these considerations, it is at once apparent that DBSS can be suitably applied for selecting well-spread samples of boxes from a batch with equal inclusion probabilities n/N , thus providing a suitable solution that fulfills EU guidelines. Moreover, because boxes are equipped by four spatial x -coordinates $x_{j,1}, x_{j,2}, x_{j,3}, x_{j,4}$ identifying their position in the batch, the transformed z -coordinates $z_{j,1} = 100x_{j,1}, z_{j,2} = 10x_{j,2}, z_{j,3} = x_{j,3}, z_{j,4} = x_{j,4}$ can be used as

additional auxiliary variables, in such a way to emphasize the Euclidean distances between boxes belonging to different pallets and, with smaller amounts, between boxes belonging to different levels (see the Appendix for more details on this points). In this way, by the use of $M = 5$ auxiliary variables, the resulting samples of boxes are not only well spread throughout the pallet, but the sample means of the four z -coordinates are similar to the averages coordinates in the pallets (see the example of Fig. 8 in the Appendix).

However, as stated by Deville and Tillé (2004, pp. 901-903), the rounding problem - that invariably occurs for satisfying the balancing equations - strictly depends on the number of balancing variables M and quickly becomes negligible if the sample size is large with respect to M . Therefore, it could be unsuitable to adopt $M = 5$ auxiliary variables with very small sample sizes that, in accordance with Table 1, never exceed $n = 11$ boxes. Accordingly, we also consider the alternative, simpler use of LPM, a scheme fully described in the Appendix, that can be viewed as a particular case of DBSS when no balancing variable is adopted with the exception of the first-order inclusion probabilities. Obviously, LPM only provides the even spread of the selected boxes throughout the batch without no calibration with respect to the box coordinates.

Because DBSS and LPM avoid the selection of neighbouring boxes, the variance of (1) that is approximated by equation (3) cannot be estimated unbiasedly. Indeed, it is a well known result of finite population sampling that the variance of an HT estimator can be estimated unbiasedly if, and only if, the sampling scheme ensures second-order inclusion probabilities that are invariably positive (see e.g., Hedayat and Sinha, 1991, Corollary 3.2, p.49). In this case, a solution is to adopt the Hansen-Hurvitz (HH) estimator for both the schemes, that in this case gives rise to

$$\hat{V}_{HH}^2 = \frac{N^2}{\hat{Y}_{HT}^2} \frac{s_e^2}{n} = \frac{s_e^2}{n\bar{y}^2} \quad (11)$$

The HH estimator constitutes by far the most familiar alternative when unbiased variance estimation is not possible and, as proven by Wolter (1985, Theorem 2.4.6, pp. 44–45), it tends to be conservative for most designs. When boxes in the batch contain an equal number of fruit or vegetable units, the HH estimator for the variance of (4) reduces to

$$\hat{V}_{HH}^2 = \frac{s_d^2}{ny_0^2} \quad (12)$$

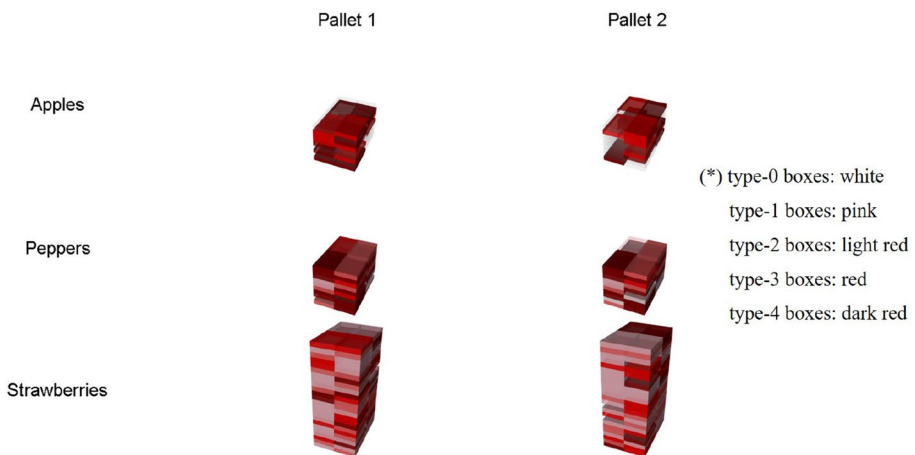
DBSS is also referred to as the local cube in the widely used `BalancedSampling` R package by Grafström and Lisic (2019) that also performs the LPM when the first-order inclusion probabilities are adopted as the sole balancing variable.

5 Simulation study

The performance of the estimator \hat{P}_{EU} under SRSWOR, DBSS and LPM was empirically checked by means of a simulation study carried out on real batches.

Table 2 Descriptive characteristics of the three batches adopted in the simulation

| Product | APPLES | PEPPERS | STRAW- BER- RIES |
|--|--------|-----------|------------------------|
| Number of pallets | 2 | 2 | 2 |
| Number of levels | 9 | 10 | 24 |
| Number of boxes per level | 4 | 4 | 4 |
| Number of units per box | 33 | Not fixed | 10 |
| Total number of boxes (N) | 72 | 80 | 192 |
| Number of boxes to select (n) | 5 | 5 | 7 |
| Total number of units | 2376 | 1243 | 1920 |
| Total number of defective units | 49 | 189 | 552 |
| Percentage of defective units ($P \times 100$) | 2.06 | 15.21 | 28.75 |

**Fig. 1** Graphical representation of box allocations within pallets and levels in the real cases (SA1) for the three batches of apples, peppers and strawberries(*)

5.1 Real and artificial batches

For simulating different situations in which defective fruits and vegetables are placed within boxes in the batch, we started from three real batches that were exhaustively checked in the Conad del Tirreno Distribution Center at Pistoia (Tuscany):

(APPLES) a batch of 2 pallets each of them constituted by 9 levels of 4 boxes of 33 apples;

(PEPPERS) a batch of 2 pallets each of them constituted by 10 levels of 4 boxes of an unknown number of peppers;

(STRAWBERRIES) a batch of 2 pallets each of them constituted by 24 levels of 4 boxes of 10 strawberries.

The characteristics of the three batches are summarized in Table 2 and graphically represented in Fig. 1, where boxes with no defective units are represented in white (type-0 boxes) and boxes with a fraction of defective units lower than the first quartile

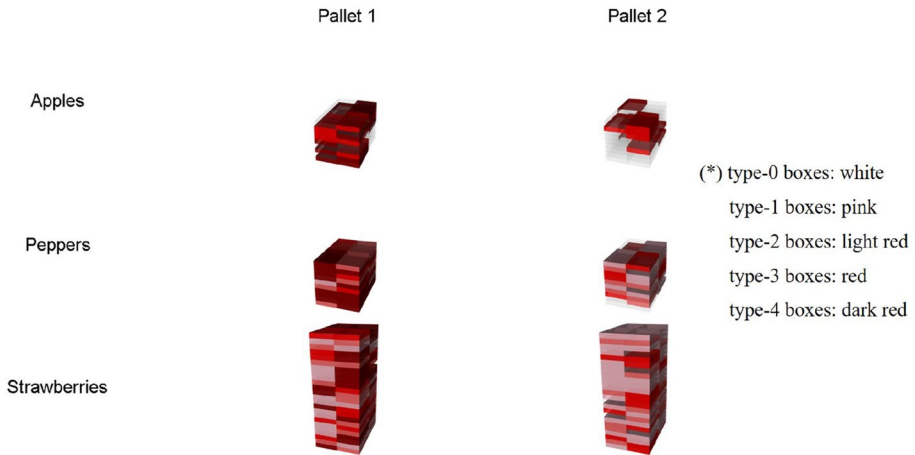


Fig. 2 Graphical representation of box allocations within pallets and levels in the real cases (SA2) for the three batches of apples, peppers and strawberries(*)

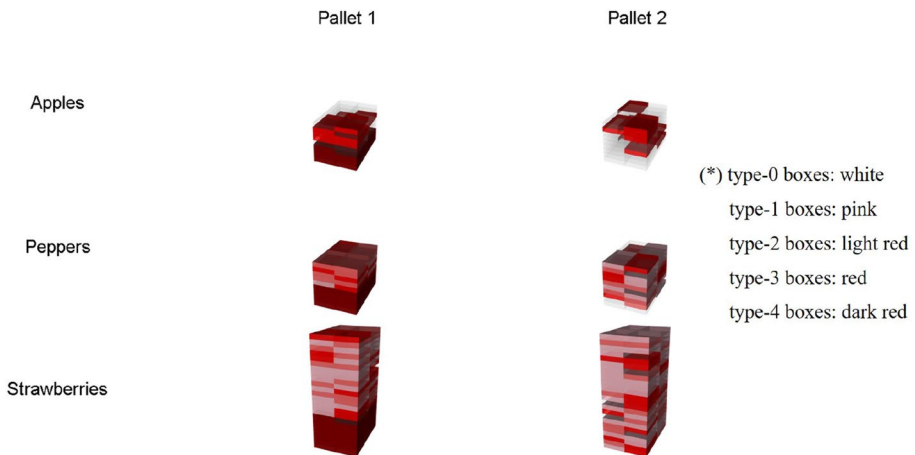


Fig. 3 Graphical representation of box allocations within pallets and levels in the real cases (SA3) for the three batches of apples, peppers and strawberries(*)

(type-1 boxes), between first and second quartile (type-2 boxes), between second and third quartile (type-3 boxes) and greater than third quartile (type-4 boxes) are represented in pink, light red, red and dark red, respectively.

For each population, three different allocations of boxes were artificially generated throughout pallets. The first simulated allocation (SA1) is the real one (see Fig. 1). In the second simulated allocation (SA2), boxes were re-allocated in such a way that type-4 boxes were present only in the first pallet and were randomly distributed throughout levels of that pallet (see Fig. 2). Finally, in the third simulated allocation (SA3), starting from SA2 allocation, type-4 boxes of the first pallet were re-allocated in the lower pallet levels (see Fig. 3). Fractions of deteriorated pieces for each population do not vary with allocations and are reported in Table 2. The complete data set, that contains the number of defective units in each boxes of the three simulated population and the three batches

is reported as supplementary material in the excel file `Data_apples_peppers_strawberries.xlsx`.

5.2 Sampling and estimation

SRSWOR, DBSS and LPM were considered for selecting n boxes from the batches, where the sample size n was determined according to EU guidelines of Table 1. DBSS was performed adopting equal inclusion probabilities and the four z -coordinates of boxes as the $M = 5$ balancing variables. For each product, each box allocation and each sampling scheme, 100000 samples of size n were independently selected. LPM was performed adopting equal inclusion probabilities as the unique balancing variable. For each generated sample, the estimate of the proportion of deteriorated units was computed by means of equation (1) in the case of peppers, where the number of units per box is unknown, and by means of equation (4) in the case of apples and strawberries, where the number of unit per box was known. Moreover, in the case of peppers, sampling variances were estimated by means of equation (8) under SRSWOR and by means of equation (11) under DBSS and LPM, while in the case of apples and strawberries sampling variances were estimated by means of equation (10) under SRSWOR and by means of equation (12) under DBSS and LPM. Once the variance estimates were achieved, the standard errors estimates were computed by the square root of variance estimates.

5.3 Performance indexes

On the basis of the resulting Monte Carlo distributions, the bias values for the estimator of the proportion of defective units (BIAS), the standard errors (SE) and the expectations of the standard error estimators (ESEE) were empirically determined. Because all the simulated strategies ensured unbiased or approximately unbiased estimators of P , true bias values were invariably equal to or close to 0. Therefore, bias values achieved from simulation were reported just to confirm the reliability of the simulation study.

Moreover, because EU guidelines entail the rejection of the whole batch when the proportion of defective items in the sample is greater than 10%, the percentage of times in which the estimated proportion is greater than 10% (OV10) was computed in order to determine the probability of taking the right decision.

5.4 Simulation results

Results from the simulation study are reported in Tables 3. The Monte Carlo values of bias invariably turned out to be smaller than 0.1%, confirming the reliability of the simulation study to approximate estimator performance.

For all the three populations and the three allocations, the performance of DBSS and LPM are practically identical both in terms of standard errors and probabilities of taking the right decision, as well as in term of variance estimation. Moreover, the use of DBSS or LPM does not provide significant improvements with respect to SRSWOR when the real allocation of boxes is considered (SA1). That is probably due to the fact that in these situations the presence of defective units does not depend to box allocation, as it can be noticed from Fig. 1, where boxes with high percentages of defective units are evenly spread over pallets and levels of the three batches. On the other hand, DBSS and

Table 3 Monte Carlo percentage values of bias (BIAS), standard error (SE), expectation of the standard error estimator (ESEE), and percentage of times in which the proportion estimate is greater than 10% (OV10) for the three batches of apples, peppers and strawberries, for each simulated allocations (SA1, SA2, SA3) under simple random sampling without replacement (SRSWOR), doubly balanced spatial sampling (DBSS) and local pivotal method (LPM)

| Product | Allocation Scheme | SA1 | | | SA2 | | | SA3 | | |
|---|-------------------|--------|-------|-------|--------|-------|-------|--------|-------|-------|
| | | SRSWOR | DBSS | LPM | SRSWOR | DBSS | LPM | SRSWOR | DBSS | LPM |
| APPLES <i>P</i> = 2.06 Acceptation | BIAS | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | SE | 1.14 | 1.15 | 1.16 | 1.14 | 1.08 | 1.07 | 1.14 | 0.90 | 0.88 |
| | ESEE | 1.14 | 1.17 | 1.17 | 1.13 | 1.19 | 1.19 | 1.13 | 1.23 | 1.23 |
| | OV10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| PEPPERS <i>P</i> = 15.21 Rejection | BIAS | -0.01 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | -0.02 | -0.01 | -0.01 |
| | SE | 4.29 | 4.17 | 4.25 | 4.31 | 3.78 | 3.82 | 4.30 | 3.46 | 3.36 |
| | ESEE | 4.31 | 4.47 | 4.46 | 4.30 | 4.56 | 4.56 | 4.30 | 4.62 | 4.64 |
| | OV10 | 88.36 | 88.67 | 88.54 | 88.16 | 91.07 | 90.98 | 88.20 | 92.77 | 93.72 |
| STRAWBERRIES <i>P</i> = 28.75 Rejection | BIAS | 0.02 | -0.01 | -0.02 | 0.03 | 0.00 | -0.03 | 0.01 | 0.04 | 0.00 |
| | SE | 5.44 | 5.37 | 5.45 | 5.45 | 5.06 | 5.05 | 5.44 | 4.46 | 4.38 |
| | ESEE | 5.44 | 5.54 | 5.54 | 5.44 | 5.59 | 5.60 | 5.44 | 5.69 | 5.70 |
| | OV10 | 99.99 | 99.99 | 99.98 | 99.98 | 99.99 | 99.99 | 99.98 | 99.99 | 99.99 |

LPM outperform SRSWOR in terms of SE when artificial allocations are considered (SA2 and SA3) with gains in precision that increase as the uneven distribution of defective units becomes more marked, i.e. passing from SA2 to SA3. In fact, while gains in precision under SA2 are smaller than 10% for the two schemes and the three batches, gains increase to about 20 – 25% under SA3. In the case of peppers, the precision gain also increases the probability of taking the right decision, i.e. rejecting the batch, that under SA3 passes from about 88% in the case of SRSWOR to about 93% in the case of DBSS. For the other two batches, the true percentages of defective units are so far from the threshold of 10% that the right decision is taken almost certainly in all the cases.

Regarding the estimation of uncertainty, the standard error estimation based on the HH estimator provides quite satisfactory results for both DBSS and LPM, being moderately conservative with overestimation smaller than one percentage point in most cases. Overestimations slightly greater than 1% only occur for both schemes in the extreme case of SA3.

Stated the equivalence of the performance achieved by DBSS and LPM, the simulation results suggest that the aim of selecting an evenly distributed sample of boxes with equal first-order inclusion probabilities is more straightforwardly achieved by the use of LPM.

6 A case study

Sampling and estimation of defective units in a batch of courgettes was performed as a training activity for the staff of the Conad del Tirreno Distribution Center in May 2017. The batch was composed by $K = 3$ pallets, in turn composed by $L_1 = 9$, $L_2 = 9$ and $L_3 = 5$ levels of 4 boxes each. In total, the batch was composed by $N = 92$ boxes. Therefore, in accordance with Table 1, the boxes to be checked were $n = 5$. The selection of the boxes was automatically performed by DBSS, adopting the constant inclusion probabilities $\pi_j = 5/92 = 0.0543478$ and the four z -coordinates identifying the spatial positions of boxes in the batch as balancing variables. The average pallet coordinate in the population was 193, the average level index was 45.6, and the averages of vertical and horizontal positions of boxes were 1.5 in both cases. Owing to the calibration ensured by the DBSS, the sample means of these coordinates should be similar with their population counterparts.

The use of DBSS gave rise to the selection of two boxes in the first pallet, one at level 4 with vertical position 1 and horizontal position 2 and one at level 6 with vertical position 1 and horizontal position 2, two boxes in the second pallet, one at level 2 with vertical position 2 and horizontal position 1 and one at level 8 with vertical position 1 and horizontal position 1, and one box in the third pallet at level 2 with vertical position 2 and horizontal position 1. Besides the even distribution of the selected boxes throughout the batch, the average pallet coordinate in the sample was 180, the average level coordinate was 44, and the sample estimates of vertical and horizontal positions were both equal to 1.4, i.e. values very similar to the averages in the whole batch.

After the staff check, the units counted in the five boxes resulted 28, 27, 32, 31, 30 and the defective units detected within were 1, 0, 2, 3, 2, respectively. In total the number of sampled courgettes were 148 with 8 of them defective. Therefore, from equation (1), the estimate of the percentage of defective units was 5.41%. Moreover, from equation (11), the standard error estimate was 1.58% with a nominal 95% confidence interval of 2.25% – 8.56%.

7 Concluding remarks

EU regulations establish strict mandatory guidelines for checking the quality of fruits and vegetables. The number of boxes n to be checked is determined as a function of the total number N of boxes in the batch, and the batch is rejected when the proportion of defective units in the n boxes exceeds 10%. No strict guidelines are instead given regarding the way in which boxes be selected from the batch, only recommending an even distribution of selected boxes throughout pallets and levels. Therefore, from a statistical point of view, sampling effort and estimation criterion are fixed and none has the advantage or the will of performing changes. From a side, warehouse managers have no convenience in increasing the number of boxes to be checked that would increase expenses (in terms of timed and staff) devoted to control. From the other side, providers would not accept any rejection based on a criterion differing from that established by EU regulations.

In this framework, statisticians can only be involved in the choice of a suitable n -sized sampling scheme for ensuring a balanced sample of boxes, in accordance with EU recommendations, at the same time ensuring good statistical properties for the proportion of defective units in the sample that is adopted as estimator of defective units in the batch.

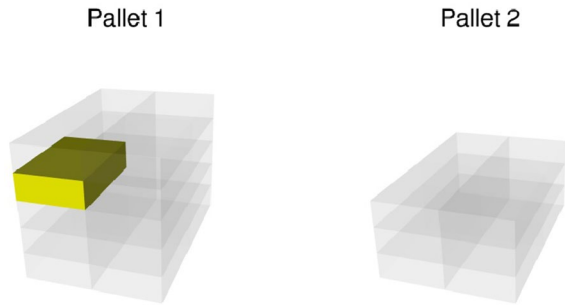
In this framework, the simplest as well effective solution seems to be the use of the equal-probability DBSS and LPM. The equal-probability sampling of boxes ensures that the proportion of defective units in the sample coincides with the HT estimator or the ratio of two HT estimators, thus ensuring unbiasedness or approximate unbiasedness and the possibility of achieving objective, conservative estimates of precision. In addition, both DBSS and LPM ensure an even distribution of selected boxes in the batch. No changes are necessary in the usual way of performing controls, except for the fact that positions identifying the n sampled boxes must be communicated to warehouse staff before controls. However, based on simulation results, the use of LPM seems to be more suitable, being more simple and ensuring a precision similar to that achieved by DBSS.

In this way, the sample proportion of defective units is equipped by an estimate regarding the precision of the strategy. If the estimated proportion is much lower or greater than 10% and the estimate of the sampling error is small with respect to the estimated proportion, these results are of utility in reaching clear opinions about the reliability of providers. If the estimated proportion is near, more or less, to 10% and the estimate of the sampling error does not exclude the possibilities of a wrong decision, these uncertainties should be declared to providers in order to claim products of better, less uncertain quality.

Appendix: equal-probability sampling schemes for selecting boxes from a batch

As stated in sect. 3, let U be a population of N boxes partitioned into K pallets U_1, \dots, U_K , where each pallet U_k is constituted by L_k levels, $U_{k,1}, \dots, U_{k,L_k}$ each of them constituted by 4 or 6 boxes. Therefore, any box $j \in U$ is univocally identified by four spatial coordinates $x_{j,1}, x_{j,2}, x_{j,3}, x_{j,4}$, where $x_{j,1}$ ranges from 1 to K and denotes the pallet, $x_{j,2}$ ranges from 1 to L_k and denotes the level in the pallet, $x_{j,3}$ ranges from 1 to 2 and denotes the vertical position in the level, and $x_{j,4}$ ranges from 1 to 2 and denotes the horizontal position in the level (see Fig. 4).

Fig. 4 Graphical representation of a batch of $N = 32$ boxes partitioned into $K = 2$ pallets of $L_1 = 5$ and $L_2 = 3$ levels, respectively, each constituted by 4 boxes. The yellow box is the box of coordinates $(1, 4, 1, 2)$



Now, in order to accomplish the EU guidelines, the sampling scheme to select boxes should have inclusion probabilities invariably equal to n/N and, at the same time, should be able to ensure a well-spread of selected boxes throughout pallets and levels, as exemplified in Fig. 5.

The simplest way to select n boxes from a batch of N boxes with equal inclusion probabilities n/N is to adopt SRSWOR, i.e., by performing n random and without replacement selections of boxes from the batch, irrespective of their position in the pallets and levels. In this way, there are $\binom{N}{n}$ possible samples of size n of equal probability $\binom{N}{n}^{-1}$. Owing to the complete randomness of the selection, SRSWOR can give rise to unbalanced selections of boxes. For example, if applied to the population of Fig. 4, SRSWOR gives rise to $\binom{32}{5} = 233504$ possible equal probability samples of size 5. In practice, under SRSWOR the two samples represented in Fig. 5 have the same probability to be selected. Therefore, SRSWOR is likely to be inefficient in batches where nearby boxes tend to be similar in terms of defective units.

A stratified design in which the batch of N boxes is partitioned into n blocks of contiguous boxes and a box is selected from each block, probably constitutes the simplest way to weaken the possibility of selecting nearby boxes. To this purpose the partition into blocks should be suitably performed in such a way to ensure blocks constituted by neighbouring boxes, avoiding, as much as possible, the presence in the same block of boxes belonging to different pallets or different levels. Therefore, suppose a suitable partition of the batch into $n - 1$ blocks of H boxes and a residual block of $N - (n - 1)H$ boxes. Then, as stated by Särndal et al. (1992, Sect. 3.4, pp. 73-75), systematic sampling can be performed by selecting a box in the first block and systematically taking a box every H thereafter, until the end of the list. In this way, there are $N - (n - 1)H$ possible samples of size n and $nH - N$ samples of size $n - 1$, all of them selected with equal probabilities $1/H$ that also constitutes the first-order inclusion probability of each box in the batch.

For example, the population of Fig. 4 can be suitably partitioned into 4 blocks of $H = 7$ boxes and one residual block of 4 boxes, in the following way

Block 1: $(1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 2, 1), (1, 1, 2, 2), (1, 2, 1, 1), (1, 2, 1, 2), (1, 2, 2, 1)$

Block 2: $(1, 2, 2, 2), (1, 3, 1, 1), (1, 3, 1, 2), (1, 3, 2, 1), (1, 3, 2, 2), (1, 4, 1, 1), (1, 4, 1, 2)$

Block 3: $(1, 4, 2, 1), (1, 4, 2, 2), (1, 5, 1, 1), (1, 5, 1, 2), (1, 5, 2, 1), (1, 5, 2, 2), (2, 1, 1, 1)$

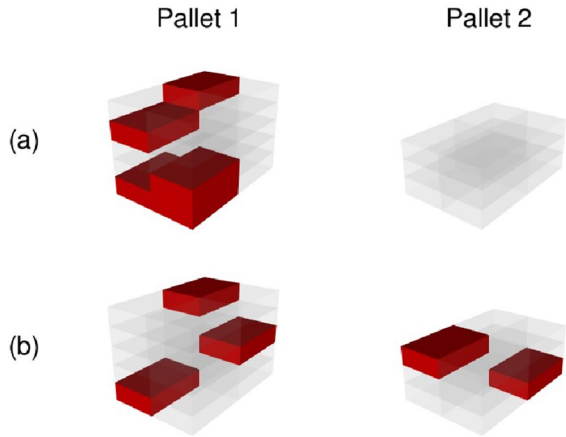
Block 4: $(2, 1, 1, 2), (2, 1, 2, 1), (2, 1, 2, 2), (2, 2, 1, 1), (2, 2, 1, 2), (2, 2, 2, 1), (2, 2, 2, 2)$

Block 5: $(2, 3, 1, 1), (2, 3, 1, 2), (2, 3, 2, 1), (2, 3, 2, 2)$

Then, by selecting one box in the first block and subsequently taking a box every 7, there are 4 possible samples of size 5, and 3 samples of size 4, each of them selected with probability $1/7$, i.e.,

Sample 1: $(1, 1, 1, 1), (1, 2, 2, 2), (1, 4, 2, 1), (2, 1, 1, 2), (2, 3, 1, 1)$

Fig. 5 Graphical representation of an unbalanced (a) and balanced (b) selection of $n = 5$ boxes (highlighted in red) from the batch of $N = 32$ boxes of Fig. 4



- Sample 2: (1, 1, 1, 2), (1, 3, 1, 1), (1, 4, 2, 2), (2, 1, 2, 1)(2, 3, 1, 2)
- Sample 3: (1, 1, 2, 1), (1, 3, 1, 2), (1, 5, 1, 1), (2, 1, 2, 2), (2, 3, 2, 1)
- Sample 4: (1, 1, 2, 2), (1, 3, 2, 1), (1, 5, 1, 2), (2, 2, 1, 1), (2, 3, 2, 2)
- Sample 5: (1, 2, 1, 1), (1, 3, 2, 2), (1, 5, 2, 1)(2, 2, 1, 2)
- Sample 6: (1, 2, 1, 2), (1, 4, 1, 1), (1, 5, 2, 2), (2, 2, 2, 1)
- Sample 7: (1, 2, 2, 1), (1, 4, 1, 2), (2, 1, 1, 1), (2, 2, 2, 2)

As the 7 samples constitute a partition of the whole batch, each box has a first-order inclusion probability of $1/7$. However, in this case it is apparent that systematic sampling cannot accomplish the EU guidelines, because the inclusion probabilities of $1/7$ differ from $5/32 = 0.15625$ that would be necessary for achieving an estimator of type (1) or (4). Moreover, the sample size is not invariably equal to 5, as would be requested by Tables 1. In practice, systematic sampling can accomplish the EU guidelines only when $H = N/n$, i.e. when the batch can be partitioned into n blocks of equal size N/n , that is not true in general.

Alternatively, the so called one-per-stratum sampling can be performed by partitioning the batch into n blocks of contiguous boxes of approximately equal sizes H_1, \dots, H_n , and then randomly and independently selecting one box in each blocks (see e.g., Breidt, 1995). In this case, there are $H_1 \dots H_n$ possible samples of equal size n that are selected with equal probability $(H_1 \times \dots \times H_n)^{-1}$ while the inclusion probability of each box in the i -th block is equal to $1/H_i$.

For example, the population of Fig. 4 can be suitably partitioned into 3 blocks of 7 boxes, one block of 6 boxes and one block of 5 boxes, in the following way

- Block 1: (1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 2, 1), (1, 1, 2, 2), (1, 2, 1, 1), (1, 2, 1, 2), (1, 2, 2, 1)
- Block 2: (1, 2, 2, 2), (1, 3, 1, 1), (1, 3, 1, 2), (1, 3, 2, 1), (1, 3, 2, 2), (1, 4, 1, 1), (1, 4, 1, 2)
- Block 3: (1, 4, 2, 1), (1, 4, 2, 2), (1, 5, 1, 1), (1, 5, 1, 2), (1, 5, 2, 1), (1, 5, 2, 2), (2, 1, 1, 1)
- Block 4: (2, 1, 1, 2), (2, 1, 2, 1), (2, 1, 2, 2), (2, 2, 1, 1), (2, 2, 1, 2), (2, 2, 2, 1)
- Block 5: (2, 2, 2, 2), (2, 3, 1, 1), (2, 3, 1, 2), (2, 3, 2, 1), (2, 3, 2, 2)

Then, by randomly selecting one box in each block there are $7^3 \times 6 \times 5 = 10290$ possible samples of size 5, and the boxes in the first 3 blocks have first-order inclusion probabilities equal to $1/7$, those in the fourth block are equal to $1/6$ and those in the fifth block are equal to $1/5$. Also in this case it is apparent that one-per-stratum sampling cannot accomplish the EU guidelines, because the inclusion probabilities of $1/7$, $1/6$ and $1/5$ differ from

$5/32 = 0.15625$ that would be necessary for achieving an estimator of type (1) or (4). Once again, one-per-stratum sampling can accomplish the EU guidelines only when the batch can be partitioned into n blocks of equal size N/n , that is not true in general.

Stated the impossibility of ensuring samples of fixed size n and equal first-order inclusion probabilities n/N by the familiar and straightforward systematic and one-per stratum sampling, more complex sampling schemes are necessary to avoid the selection of neighbouring boxes and provide well spread samples at the same time addressing the EU guidelines. To this purpose, it is necessary to point out that the box coordinates $x_{j,1}, x_{j,2}, x_{j,3}, x_{j,4}$ are sufficient to identify box positions in the batch, but they are not suitable to quantify distances between boxes. Indeed, if the familiar Euclidean criterion

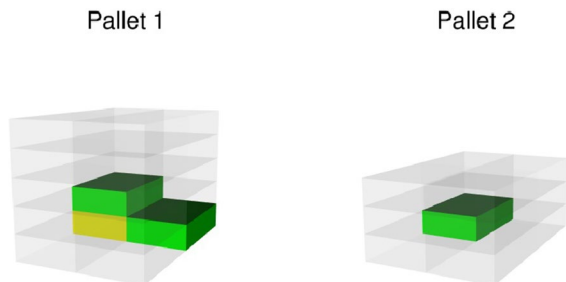
$$d_{j,h} = \sqrt{\sum_{i=1}^4 (x_{j,i} - x_{h,i})^2} \tag{A.1}$$

is adopted to determine the distance between boxes j and h , then, referring for example to the population of Fig. 4, the distance of the box (1, 1, 1, 1) to the box (1, 1, 2, 1) in the same pallet and the same level is equal to 1. However, 1 is also the distance to the box (1, 2, 1, 1) that instead is in the upper level and 1 is even the distance to the box (2, 1, 1, 1) that is in the other pallet (see Fig. 6).

To solve the problem and to assign greater distances between boxes belonging to different pallets and different levels, we multiply the pallet coordinate $x_{j,1}$ by 100 and the level coordinate $x_{j,2}$ by 10, leaving unchanged the vertical and horizontal positions $x_{j,3}$ and $x_{j,4}$. In this way, the distance $d_{j,h}$ between boxes j and h can be suitably determined by the Euclidean criterion of type (A.1) applied to the new z -coordinates $z_{j,1} = 100x_{j,1}, z_{j,2} = 10x_{j,2}, z_{j,3} = x_{j,3}, z_{j,4} = x_{j,4}$. Therefore, distances are great between boxes belonging to different pallets and if pallets are numbered in accordance with their position in the warehouse, distances increase with the physical distances of pallets in the batch, because even pallet position may influence the deterioration of fruits and vegetables. Distances also increase, even if for smaller amounts, between boxes belonging to different levels. In practice, referring to the previous example of Fig. 6, on the basis of z -coordinates it follows that the distances of the box (1, 1, 1, 1) to boxes (1, 1, 2, 1), (1, 2, 1, 1) and (2, 1, 1, 1) are equal to 1, 10 and 100, respectively.

By exploiting this distance criterion, the LPM by Grafström et al. (2012) seems to be a suitable solution. The idea pursued by the authors is to create a strong negative correlation between the inclusion of closed units at the same time preserving pre-fixed first-order inclusion probabilities, whose sum must be equal to n to ensure samples of fixed size n (see e.g. Hedayat and Sinha, 1991, p.12). LPM is performed exploiting a selection algorithm

Fig. 6 Graphical representation of boxes (1, 1, 2, 1), (1, 2, 1, 1) and (2, 1, 1, 1) (highlighted in green) that have equal Euclidean distance 1 to the box (1, 1, 1, 1) (highlighted in yellow) on the basis of the x -coordinates



referred to as the pivotal method (Deville and Tillé, 1998). The original algorithm randomly selects two units from the undecided ones and randomly updates their first-order inclusion probabilities in such a way that one is increased as much as possible and one is decreased in the same way, but preserving inalterate their sum until the sampling outcome is decided for at least one unit, in the sense that at least one inclusion probability is updated to 0 (unit excluded) or 1 (unit selected). Thus, the sampling ends in at most N steps.

In this framework, the suggestion given by Grafström et al. (2012) is to choose the first unit at random from the undecided units and to take the nearest undecided unit as the second unit, or to choose it at random among the nearest undecided units, when, as in a batch of boxes, the nearest units may be more than one. Then, if applied for selecting boxes from a batch, the scheme avoids the selection of neighbouring boxes and achieve a well-spread sample. For example, Fig. 7 shows the good spreading achieved by a sample of size 5 selected by LPM from the batch of Fig. 4.

A further suitable solution for effectively selecting boxes is the use of DBSS by Grafström and Tillé (2013) achieved by combining a generalization of the LPM with the cube method by Deville and Tillé (2004). As LPM, DBSS is tailored to sample populations of units with pre-fixed first-order inclusion probabilities excluding the selection of nearby units but at the same time ensuring nearly balanced samples with respect to M auxiliary variables. In practice, the totals of these variables estimated from the sample must be almost the same as the true totals in the population. It is worth noting that for obtaining samples of fixed size n , the first-order inclusion probabilities of units must be chosen as one of these variables, in such a way that their sum invariably equals n . The basic idea of DBSS is to generalize LPM by randomly selecting clusters of $M + 1$ nearby undecided units (see the cluster selection algorithm by Grafström and Tillé, 2013, p.125) instead of two neighbouring undecided units, as in the LPM algorithm, and to repeatedly apply the fast implementation of the so-called flight phase of cube method (e.g. Tillé, 2006, Sect. 8.6.2) on these clusters. In practice, for each selected cluster of $M + 1$ points, the flight phase consists in updating the first-order inclusion probabilities while respecting the balancing conditions until at least one inclusion probability is updated to 0 (unit excluded) or 1 (unit selected). Thus, the flight phase ends in at most $N - M$ steps with a number of undecided units smaller than $M + 1$. Obviously, when a unit is selected from a cluster, the first-order inclusion probabilities of the other units in the cluster are generally decreased in order to respect the balancing condition that their sum must be equal to n , while, for the same reason, they are generally increased if the unit is excluded. Therefore, because the updating is done for clusters of neighbouring units, the scheme, as LPM, ultimately creates a strong negative correlation between the inclusion of closed units.

Finally, at the end of the flight phase, if the number of undecided units is 0, there is no other action to perform. Otherwise the so called landing phase is necessary to complete the

Fig. 7 Graphical representation of a sample $n = 5$ boxes (highlighted in red) from the batch of $N = 32$ boxes of Fig. 4 by means of the local pivotal method

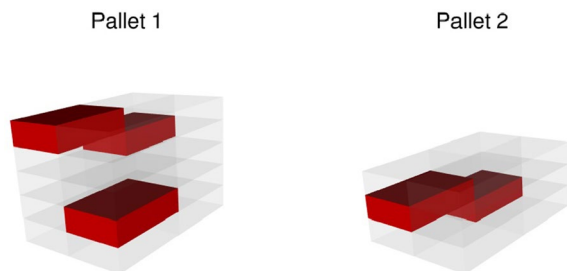
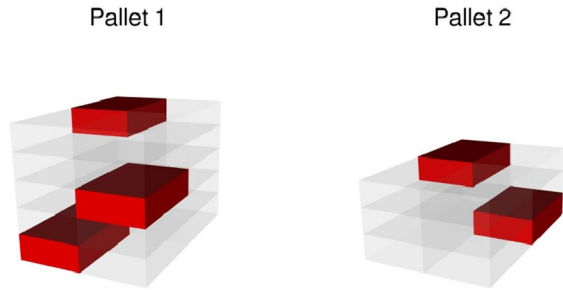


Fig. 8 Graphical representation of a sample $n = 5$ boxes (highlighted in red) from the batch of $N = 32$ boxes of Fig. 4 by means of the doubly balanced spatial sampling



sample (Deville and Tillé, 2004). In practice, if there are $0 < G < M + 1$ undecided units and g units have to be selected, the landing phase consists in enumerating all the possible sets of size g that can be selected out of the G . Then, one of these sets is selected and the g units of the set are adopted to complete the sample. In most cases, owing to rounding problems, the resulting sample is not exactly but only approximately balanced with respect to the M auxiliary variables.

If the first-order inclusion probabilities are used as the sole balancing variable, DBSS coincides with LPM, only ensuring samples of units well spread throughout the population (Grafström and Tillé 2013). If, besides the first-order inclusion probabilities, further balancing variables are adopted to quantify some spatial characteristics of the units, then it is also possible to force samples to satisfy further properties. In this scenario, a suitable option suggested by Grafström and Tillé (2013) is the use of the unit coordinates as balancing variables. In this case, DBSS not only provides well-spread samples, but samples and the population have approximately the same average coordinates.

DBSS can be applied for selecting samples of boxes from a batch with equal inclusion probabilities n/N , using the four z -coordinates as additional balancing variables, for a total of $M = 5$ balancing equations. For example, Fig. 8 shows the good spreading achieved by a sample of size 5 selected by DBSS from the batch of Fig. 4 joined with the similarity of the averages of the z -coordinates in the whole batch, i.e. 137.5, 26.25, 1.5, 1.5, with their HT estimates achieved from the selected sample, i.e. 140, 26, 1.4, 1.4, respectively.

Funding Open access funding provided by Università degli Studi di Siena within the CRUI-CARE Agreement.

Declarations

Conflicts of interest This paper is based in a study financed by Conad del Tirreno contract n. 905-2016. The usual disclaimers apply.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Breidt, F.J.: Markov chain designs for one-per-stratum sampling. *Survey Methodology*. **21**, 63–70 (1995)
- Deville, J.C., Tillé, Y.: Unequal probability sampling without replacement through a splitting method. *Biometrika*. **85**, 89–101 (1998)
- Deville, J.C., Tillé, Y.: Efficient balanced sampling: the cube method. *Biometrika*. **91**, 893–912 (2004)
- Grafström, A., Lisic, J.: Balanced and Spatially Balanced Sampling. (2019). <http://www.antongrafstrom.se/balancedsampling>
- Grafström, A., Lundström, N.L.P., Schelin, L.: Spatially Balanced Sampling through the Pivotal Method. *Biometrics*. **68**, 514–520 (2012)
- Grafström, A., Tillé, Y.: Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*. **24**, 120–131 (2013)
- Hedayat, A.S., Sinha, B.K.: *Design and Inference in Finite Population Sampling*. Wiley (1991)
- Särndal, C.E., Swensson B., Wretman J. : *Model Assisted Survey Sampling*. Springer-Verlag, New York (1992)
- Tillé, Y.: *Sampling Algorithms*. Springer-Verlag, New York (2006)
- Wolter, K.M.: *Introduction to Variance Estimation*. Springer-Verlag, New York (1985)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.