CONTINUOUS-POPULATION APPROACH IN FOREST INVENTORY — DEALING WITH THE INTRICATE RELATION BETWEEN POINTS AND TREES

Minna Pulkkinen [1]

[1] *Laboratoire d'inventaire forestier (IGN / ENSG), France, minna.pulkkinen@ign.fr*

Résumé. Les populations d'arbres sont difficiles à échantillonner, car il est impossible dans la pratique de créer une base d'échantillonnage dans une grande région d'intérêt. C'est pourquoi, dans les inventaires forestiers, les arbres sont échantillonnés indirectement, en plaçant des placettes de superficie fixe à des points situés de manière aléatoire dans la région d'intérêt. Cela implique la prise en compte simultanée de deux populations — la population d'échantillonnage continue constituée de points et la population cible discrète constituée d'arbres — et la nécessité d'établir une relation entre elles. Dans cet article de synthèse, nous passons en revue les différentes approches présentées dans la littérature pour traiter de cette dualité, en nous efforçant d'expliciter le raisonnement statistique qui sous-tend chaque approche, en incluant les hypothèses qui ne sont souvent pas explicitement énoncées.

Mots-clés. Zone d'inclusion d'un arbre, densité d'attribut, densité d'inclusion d'un point, approche d'intégration de Monte Carlo, méthode de partage de poids généralisée.

Abstract. Tree populations are difficult to sample, since creating a sampling frame of trees within any larger region of interest is unfeasible in practice. Therefore, in forest inventories, trees are sampled indirectly, by placing fixed-area plots at randomly located points within the region of interest. This implies a simultaneous consideration of two populations — the continuous sampling population of points and the discrete target population of trees — and a need to establish a link between them. In this review paper, we go through the different approaches presented in literature to dealing with this duality, and dissect the statistical reasoning behind each approach, including the assumptions that are often not explicitly stated.

Keywords. Inclusion zone of a tree, attribute density, inclusion density of a point, Monte Carlo integration approach, generalised weight share method.

1 Introduction

Tree populations are difficult to sample: they are finite but often large, spatially irregularly arranged, and changing in time, which makes it in practice unfeasible to create a sampling frame (i.e., a list of all the trees and their locations within a region of interest), from which a sample of trees could be directly drawn. Instead, trees are sampled indirectly by means of plots.

2 Traditional plot sampling

Traditionally, indirect sampling has been carried out by placing a fixed-area plot at a sampling point that is assumed to be located uniformly at random in the region and by including in the sample the trees located within the plot (see, e.g., de Vries, 1986; Schreuder et al., 1993; Gregoire and Valentine, 2008, Chs. 7 and 8). The *total of the attribute of interest in the tree population* is then estimated with the Horvitz-Thompson estimator (Horvitz and Thompson, 1952) as a weighted sum of the attribute values measured on the sampled trees, with the inverses of the inclusion probabilities of the trees as weights. The *inclusion probability of a tree* is obtained with the idea first presented by Grosenbaugh (1952) and later referred to by Mandallaz (2008, Ch. 4) as the *duality principle*: a tree is included in the sample if and only if the sampling point is located in the inclusion zone of the tree. This zone thus consists of all the points in the region from which the tree could be sampled with the sampling method used. With a fixed-area plot, the zone is (the within-region part of) the plot placed at the tree pith on the ground level. With the assumption of the uniform random location of the sampling point, the tree inclusion probability then becomes, by geometric probability, the ratio of the area of the tree inclusion zone to the area of the region. Sampling is repeated at n independent sampling points, which results in n independent and identically distributed Horvitz-Thompson estimates. The population total is finally unbiasedly estimated with the mean of the n estimates, and the variance of this mean is unbiasedly estimated by the sample variance of the n estimates divided by n.

It is important to note that the trees included into the n samples obtained at n points are not lumped together for the estimation; if they were, in the computation of tree inclusion probabilities we should take into account the fact that trees can be included in the sample more than once (cf. Eriksson, 1995, “Tree design” in pp. 898–901). Further, the variance is usually not estimated using the unbiased estimator of the variance of the Horvitz-Thompson estimator, because it involves the joint inclusion probabilities of all the pairs of the trees in the sample, the computation of which would require that the locations of the sampled trees be known.

In practice, it is usually the *spatial mean of the attribute of interest in the tree population* (i.e., the population total of the attribute divided by the region area) that is estimated on each plot or at each angle-count sampling point. The region area, present as the denominator in tree inclusion probabilities, now cancels out, and the estimate of spatial mean becomes a weighted sum of the tree attribute values with the inverses of the tree inclusion zone areas as weights; with a fixed-area plot, the estimate is thus the sum of the attribute values of the trees within the plot divided by the plot area. The population total is then estimated by multiplying the mean of the n spatial mean estimates with the region area.

3 Continuous-population approach

In the traditional approach described above, the presence of two populations — the continuous *sampling population of points* and the discrete *target population of trees* within the region of interest — and the link between them have not always been explicit. Mandallaz (1991, 2008) was the first to present a *continuous-population approach*, where the two populations are explicitly linked by *attribute density* (also referred to as local density), a two-dimensional function defined at every point of the region of interest and *with the property that its integral over the region is equal to the tree population total of the attribute in the region.* It is the values of the attribute density that are observed at n sampling points drawn from the continuous population of the region (yet the drawing need not necessarily be done uniformly at random or independently for each point). With the sample of n attribute density values, the tree population total (i.e., the integral of the attribute density over the region) is then estimated. Because of the evident association to Monte Carlo integration, the continuous-population approach is also called *Monte Carlo integration approach* (Gregoire and Valentine, 2008, Ch. 10).

One way to define an attribute density that satisfies the integration condition is technically the same as the estimation of the spatial mean explained above: at any point within the region, the attribute density is defined as a weighted sum of the attribute values of the trees located within the plot placed at that point, with the inverses of the tree inclusion zone areas as weights (Stevens and Urquhart, 2001, combination of Eqs. 2 and 5; Mandallaz, 2008, combination of Eqs. 4.4 and 4.5; Gregoire and Valentine, 2008, Eq. 10.6; Grafström et al., 2017, Eq. 1). However, this is not a statistical estimate, and hence the ratios of the inclusion zone areas to the region area (which are reduced to mere inclusion zone areas), should not be interpreted as tree inclusion probabilities (cf. Eriksson, 1995, "An infinite population approach" in pp. 901–903). Different plot sizes and shapes result in different attribute density surfaces.

In the continuous-population approach, the estimation is based on the attribute density values observed at n randomly located sampling points. In the case where the sampling points are independent and located uniformly at random in the region, the point estimator of the tree population total, its variance and the unbiased estimator of the variance become the same as in the traditional approach above (Eriksson, 1995, Eqs. 7–9; Mandallaz, 2008, Theorem 4.2.1; Gregoire and Valentine, 2008, Eqs. 10.11, 10.13 and 10.14). More generally, the estimation relies on *Cordy's (1993) extension of the Horvitz-Thompson theorem to continuous populations* (Stevens and Urquhart, 2001; Grafström et al., 2017), based on the *inclusion densities of the points* derived from the marginal density functions of the points (obtained by integration from the joint probability density function of the points).

In the continuous-population approach, sampled trees (target population elements) are "hidden" in the attribute density observations at points (sampling population elements). This may be quite confusing, as usually in survey sampling population total estimators are expressed as weighted sums of the attribute values of the sampled target population units, with each element having its own design or estimation weight. Recently, Chauvet et al. (2023) presented a new method that gives an alternative way of looking at continuous-population sampling in forest inventory and enables us to present the population total estimators explicitly in terms of tree attributes and tree estimation weights (Bouriaud et al. 2024). The method is an extension of the *weight-share method* (Deville and Lavallée, 2006), originally developed for sampling a discrete population for which no sampling frame is available, with the help of another discrete population for which a sampling frame is available and which is in a known way linked to the first population. Relying on Cordy's (1993) continuous-population extension of the Horvitz-Thompson theorem, Chauvet et al. (2023) *generalised the weight-share method for sampling a discrete population from a continuous one.*

Bibliography

Bouriaud, O., Brion, P., Chauvet, G., Duong, T. and Pulkkinen, M. (2024), The weight share method in forest inventories: refining the relation between points and trees. *Canadian Journal of Forest Research* (accepted).

Chauvet, G., Bouriaud, O. and Brion, P. (2023), An extension of the weight share method when using a continuous sampling frame, *Survey Methodology*, 49(1), pp. 139–162.

Cordy, C.B. (1993), An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe, *Statistics & Probability Letters*, 18, pp. 353–362.

Deville, J.-C., and Lavallée, P. (2006), Indirect sampling: The foundations of the generalized weight share method, *Survey Methodology*, 32(2), pp. 165–176.

Eriksson, M. (1995), Design-based approaches to horizontal-point sampling, *Forest Science*, 41(4), pp. 890–907.

Grafström, A., Schnell, S., Saarela, S., Hubbell, S.P. and Condit, R. (2017), The continuous population approach to forest inventories, *Environmetrics*, 28, p. e2480.

Gregoire, T.G. and Valentine, H.T. (2008), *Sampling strategies for natural resources and the environment*, Chapman & Hall/CRC, Boca Raton, Florida, 474 p.

Grosenbaugh, L.R. (1952), Plotless timber estimates — new, fast, easy, *Journal of Forestry*, 50(1), pp. 32–37.

Horvitz, D.G. and Thompson, D.J. (1952), A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, 47(260), pp. 663–685.

Mandallaz, D. (1991), *A unified approach to sampling theory for forest inventory based on infinite population and superpopulation models*, PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 257 p.

Mandallaz, D. (2008), *Sampling techniques for forest inventories*, Chapman & Hall/CRC, Boca Raton, Florida, 256 p.

Schreuder, H.T., Gregoire, T.G. and Wood, G.B. (1993), *Sampling methods for multiresource forest inventory*, John Wiley & Sons, New York, 446 p.

Vries, P.G. de (1986), *Sampling theory for forest inventory: A teach-yourself course*, Springer-Verlag, Heidelberg, 399 p.