Daniel Mandallaz

The Anticipated Variance: a Tool for the Optimization of Forest Inventories

Chair of Forest Inventory and Planning
Swiss Federal Institute of Technology (ETH), Zürich
1997

Daniel Mandallaz

The Anticipated Variance: a Tool for the Optimization of Forest Inventories

Chair of Forest Inventory and Planning
Swiss Federal Institute of Technology (ETH), Zürich
1997

Published by:
Chair of Forest Inventory and Planning
Department of Forest and Wood Sciences
CH-8092 Zürich

Herausgeber:
Professur für Forsteinrichtung und Waldwachstum
Departement Wald-und Holzforschung
ETH Zentrum
CH-8092 Zürich

Aknowledgements

I would like to express my thanks to Professor P. Bachmann, Chair of Forest Inventory and Planning at ETH Zürich, for his remote but vital support as well as for the working environment he succeeded to create. Many thanks are also due to my colleagues, Dr. Ronghua Ye and A. Lanz, for their never ending questions: they cannot be blamed for what remains unclear in this work.

A Lea Marine

Abstract

This work presents optimal sampling schemes for forest inventory. The sampling procedures are optimal in the sense that they minimize the anticipated variance for given costs or conversely. The anticipated variance is defined as the average of the design-based variance under a simple stochastic model for the location of the trees. This location model, the local Poisson Forest, assumes that trees are uniformly and independently distributed within a given number of strata. It is then possible to derive analytically the optimal inclusion rules for all possible combinations of sampling schemes involving one-phase or two-phase procedures, likewise one-stage or two-stage, and this for both simple random sampling as well as cluster random sampling. The optimal inclusion rules are either probability proportional to size, in one-stage procedures, or a combination of probability proportional to prediction and probability proportional to error, in two-stage procedures. Optimal feasible approximations of the exact optimal sampling schemes are also given as well as the relative efficiencies of all optimal schemes.

Zusammenfassung

Diese Arbeit stellt optimale Stichprobenpläne für die Waldinventur dar. Optimal bedeutet, dass die antizipierte Varianz bei vorgegebenen Kosten minimiert wird, oder umgekehrt. Die antizipierte Varianz is das Mittel der klassischen Stichprobenvarianz unter einem stochastischem Modell, welches die räumliche Lage der Bäume erzeugt. In diesem räumlichen Modell, das lokale Poisson Modell, sind die Bäume unabhängig und uniform innerhalb Straten verteilt. Die optimalen Aufnahmewahrscheinlichkeiten können analytisch abgeleitet werden, und zwar für alle Kombinationen von einphasigen, zweiphasigen, einstufigen und zweistufigen Verfahren, sowohl für einfache als auch für Traktstichproben. Die optimalen Pläne beruhen auf Aufnahmewahrscheinlichkeiten, welche, bei einstufigen Verfahren, direkt proportional zur Zielgrösse sind, oder, bei zweistufigen Verfahren, proportional zu einer Prognose der Zielgrösse und zum Prognosefehler. Optimale praktisch durchführbare Approximationen der exakt optimalen Verfahren werden ebenfalls gegeben, wie auch deren relativen Effizienz.

Résumé

Ce travail présente des plans d'échantillonnage optimaux pour l'inventaire des forêts, au sens que la variance anticipée est minimale pour un coût donné, ou inversément. La variance anticipée est la moyenne de la variance sous le plan de sondage par rapport à un modèle stochastique pour la distribution spatiale des arbres. Ce modèle, le modèle poissonien local, suppose que les arbre sont répartis indépendamment et uniformément à l'intérieur de strates. Il est alors possible de calculer analytiquement les probabilité d'inclusion optimales pour toutes les combinaisons d'inventaires à une ou deux phases, un ou deux degrés, ainsi que pour les échantillonnage simples ou en satellites. Les plans optimaux conduisent à des probabilités d'inclusion proportionnelles, soit directement à la grandeur cible, dans les plans à un degré, soit, dans les plans à deux degrés, conjointement proportionnelles à une prévision de la grandeur cible et à l'erreur de cette prévision. Il est possible de donner les approximations optimales et réalisables des plans optimaux exacts et de calculer les efficacités relatives.

Contents

Chapter 1

Introduction

1.1 Historical background and scope

Inventories are the bases for forest management planning, which itself can be defined as the optimal utilization, under given constraints, of forest resources. Management requires therefore to collect, summarize and interpret information, i.e. to perform statistical work. The development and improvement of forest management, which began towards the end of the Middle Ages, is consequently strongly dependent on the parallel evolution of inventory techniques and statistical methodology, in particular sampling techniques, to the degree that today forest inventory is simply unthinkable without them.

Over the last 80 years, the number of techniques, their complexity, the demand for more and better information and finally the mere complexity of the required investigations seem all to have followed an exponential growth rate. This has led to a large number of specialized procedures, scattered over an immense literature, and also some uncertainties at the conceptual level. There was therefore a need to clarify and summarize the fundamental statistical concepts and tools required to perform inventories. The books of H.T. Schreuder et al [20] and of de Vries [26] were in this respect highly welcome. At a more theoretical level D. Mandallaz [10], considered design-based and model-based inference in a framework tailored to forest inventory, as well as geostatistical techniques [12].

If there is now a wide choice of forest inventory techniques to provide the information required by forest management planning, the difficulty has moved to a higher level: how do we plan forest inventories? More precisely, which techniques should we use and when to provide the required information in the most cost efficient way. Of course, this problem has received much attention in classical sampling theory, as used primarily in socio-economical studies, from J. Neyman [17], F. Yates [27], W.G. Cochran [3] and C.E. Saerndal's et al monumental work [21], and also in the forest sampling literature but in a less sytematic way and at a rather elementary mathematical level. In view of the tighter economical constraints imposed on government spending in general and forest services in particular, forest inventorists should whenever possible choose the most efficient sampling techniques available.

The purpose of this technical report is to present a simple and yet fairly general statistical framework for the optimization of forest inventories, which rests upon the concept of anticipated variance. This concept has played a key role in modern sampling theory, but so far seems to have been ignored from forest inventorists. The main idea is to minimize the anticipated variance for given expected costs or to minimize the expected costs for a given anticipated variance. The anticipated

variance is taken with respect to a simple stochastic model for the location of the trees: it is assumed that the forest is partitioned in strata, in which the trees are independently and uniformly distributed. This model allows for relatively simple analytical rules with a clear intuitive and plausible background. The resulting optimal rules are always a combination of probability proportional to either size, prediction or error. The prior information required for designing an optimal scheme is surprisingly simple. Preliminary results based on the anticipated variance were briefly outlined by the author in his inaugural lecture [13]; this work gives now a detailed and rigourous treatment.

It can be expected that the mathematically optimal sampling schemes are probably good solutions, and certainly better than sampling schemes departing strongly from them. It must be emphasized, however, that the optimality is only asymptotic and therefore approximately reached only for forest inventories carried out on large surface areas. Validation of the theory by simulations and illustration by cases studies are currently under way and will be presented in a further report.

This work deals with **sampling schemes with known inclusion probabilities, having one or two phases and one or two stages, either with simple random sampling or cluster random sampling**. It addresses primarily forest inventories performed in one single time point.

Given time, the reader with an intermediate background in probability theory and calculus should be able to fully understand most of the concepts and results which are presented in an almost completely self-contained manner. For a first reading we recommend to concentrate on the chapters dealing with simple random sampling, since cluster random sampling is technically more difficult, though the main concepts remain the same. We assume that the reader is familiar with the standard notation of elementary set theory.

1.2 Terminology and formulation of the inventory problem

We now proceed to define the terminology and notation which will be used throughout this work.

We consider a forest area F , which is assumed to be a subset of the Euclidean plane $\Re^2$. In practice this implies that F is already the suitable projection of a real forest, e.g. parallel projection whenever the earth curvature can be neglected; we shall not deal with the approximate practical solutions, like slope correction, to achieve this. The surface area of F is denoted by $\lambda(F)$, usually in ha.

We consider a well defined population $\mathcal{P}$ of N trees lying in F; the trees are identified by their labels $i = 1, 2 \ldots N$. The position vectors of the trees are denoted by $u_i \in \Re^2 \quad , i = 1 \ldots N$. To simplify the notation we shall write $i \in G$ instead of $u_i \in G \subseteq F$, the surface area of an arbitrary set G is always denoted by $\lambda(G)$.

The response variables of interest measured or observed at a given time point on each tree in $\mathcal{P}$ are denoted by $Y_i^{(m)}, m = 1, \ldots p$, and are assumed to be error-free. Whenever confusion can be excluded from the context we shall drop the upper index (m). The response $Y_i^{(m)}$ can be a real or integer number in the usual sense, e.g. the timber volume, the basal area, $Y_i \equiv 1$ for the number of stem, or a set of binary $0, 1$ variables coding categorical variables like species, state of health etc., or any variable defined with the previous ones. We take here obviously the abstract view that trees are dimensionless points in the plane in which the response variables $Y_i^{(m)}$ are defined. Though we shall only deal, as already mentioned, with inventories performed in one time point, it is instructive to note that the response variable, say $Y_i^{(2)}$, can be the change of another variable, say $Y_i^{(1)}$ over a given time period: this

is particularly useful in the context of continuous forest inventories with permanent plots.

Given any set $G \subseteq F$ the objectives of forest inventory, in the restricted sense, is to get information on densities, totals or ratios, defined according to:

$$\begin{aligned} \bar{Y}_G^{(m)} &= \frac{1}{\lambda(G)} \sum_{i \in G} Y_i^{(m)} \\ Y_G^{(m)} &= \sum_{i \in G} Y_i^{(m)} \\ R_{l,m} &= \frac{\bar{Y}_G^{(l)}}{\bar{Y}_G^{(m)}} = \frac{Y_G^{(l)}}{Y_G^{(m)}} \end{aligned} \tag{1.1}$$

possibly for many different sets G and in several time points.

If the set G is small a full census is feasible and is often to be preferred to a sample survey. However, in most instances a full census is not feasible for several G's, and almost never for the complete forest area F. Hence, one must use sampling techniques, which generally perform well in F but not necessarily so for small $G \subseteq F$, which is known as the problem of small area estimation.

In this work we shall have primarily in mind the optimization of forest inventory for large F. Actually the results have a simple analytical form only when $\lambda(F) \to \infty$. We shall restrict the discussion to the optimal estimation of densities, because it essentially implies the optimal estimation of total and also because ratios are usually not the quantities of primary interest. For an easier intuitive understanding it is best to have in mind the overall timber volume per unit area as the main objective of the inventory.

Chapter 2

Definition of Sampling Schemes

2.1 Generalities

In this chapter we shall present the most important concepts and tools of forest sampling in a framework general enough to cover the majority of the situations encountered in practice. The terminology being unfortunately not uniform the reader should carefully look at the various definitions used in the literature.

Most modern forest inventories are so called combined forest inventories, i.e. combining information obtained directly in the forest, by means of terrestrial plots, and auxiliary information obtained by remote sensing, i.e. aerial photographs or satellites images, or any other sources, like thematic maps or previous inventories. In short and in this work, the auxiliary information is collected in the **first phase**, usually with a very large sample (formally even infinite when thematic maps are used). The **second phase** collects the terrestrial information on a **subsample of the first phase sample**.The terrestrial information itself is collected either by **one-stage** procedures, in which trees are selected to obtain directly the response variable of interest, or by **two-stage** procedures, in which the **first-stage** trees are selected to obtain an approximation of the response variable and a subsample of the first-stage trees, the **second-stage** trees, is drawn to obtain the exact response. For each of the above 4 possibilities one can use **simple random sampling**, in which the information is collected in single points uniformly and independently distributed in the forest area, or **cluster random sampling**, in which the information is collected in clusters (set of point with a fixed geometrical structure), whose origins are drawn by simple random sampling; the number of points per cluster falling into the forest area is also a random variable. The theory of cluster sampling presented here differs from the existing literature, we believe it is both simpler and better.

In practice random sampling is rarely if ever used, and most inventories rely on systematic grids, eventually with random start or orientation. Most forest inventorists treat points lying on systematic grids as if they were random. This is to a large extent acceptable for point estimates, less so for variance estimates, particularly in small area estimation. When the size of the domain under study is much larger than the range of the spatial correlation, geostatistical variance estimates are generally close to their design-based counterparts, see [12] for theoretical and empirical evidence. Unfortunately the optimization of sampling schemes in the geostatistical framework is, because of mathematical complexity, up to now by and large an open problem. Since this work addresses primarily the optimization of large forest inventories, it can be reasonably assumed that the resulting guidelines

are also valid for systematic grids and therefore for practical work.

The next sections of this chapter will give rigorous definitions and results in a modern framework, inspired primarily from [10], which in turn adapted to forest inventory the pioneer work of C.E. Saerndal, now available in the excellent book [21]. It differs from the standard books of H.T Schreuder et al [20] and de Vries [26], which for the most part rest upon the classical sampling theory for finite populations as formulated in the famous and paradigmatic book of W.G. Cochran [3].

2.2 One-phase one-stage simple random sampling scheme

Throughout this work we shall use the symbols P, E, V, COV for probability, expectation, variance and covariance. To have a better intuitive understanding of the general procedure let us consider the most simple case first.

We draw a random point x uniformly in F, which means that for any set $B \subseteq \Re^2$ the probability that the point x falls into B is given by

$$P\left(x \in B\right) = \frac{\lambda\left(B \cap F\right)}{\lambda(F)} \tag{2.1}$$

Trees are selected if they are within the circle $K_r(x)$ with fixed radius r centered in x. We now define the random indicator variables:

$$I_i(x) = \begin{cases} 1 \text{ if } i \in K_r(x) \\ 0 \text{ if } i \notin K_r(x) \end{cases}$$

and the N circles $K_i(r)$ with constant radius r centered on the trees. By symmetry the ith tree is in the circle $K_r(x)$ if and only if the random point x is in the ith circle $K_i(r)$, hence we have

$$I_i(x) = 1 \Leftrightarrow x \in K_i(r)$$

The inclusion probability of the ith tree is consequently given by

$$\pi_i = P(I_i(x) = 1) = E_x\left(I_i(x)\right) = \frac{\lambda\left(K_i(r) \cap F\right)}{\lambda(F)}$$

Up to boundary effects at the forest edges this inclusion probability is constant. For a given variable Y, and neglecting boundary effects, it is natural to define the **local density** $Y(x)$ at the point x as the sum of the Y_i over the trees selected, divided by the constant surface area of the circle $K_r(x)$ that is to set

$$Y(x) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{I_i(x) Y_i}{\pi_i}$$

This dual consideration leads immediately to a far reaching generalization.We can assign to each tree its circle K_i, whose radius depends on the label i and therefore eventually on the $Y_i^{(m)}$ and u_i; for instance it might depend on diameter and species. As a matter of facts, circles are convenient but not compulsory: squares, rectangles or any shape with fixed orientation will do. The famous angle count technique with limit angle α assigns to each tree with diameter (in cm) at breast height D_i a circle with radius (in m)

$$\frac{D_i}{2\sqrt{k}}, \quad \text{where} \quad k = 10^4 \sin^2(\frac{\alpha}{2})$$

In field work the tree is included in the sample if its apparent diameter viewed from the point x (angle α_i) is larger than the limit angle α. The widely used concentric circles techniques can be viewed as a discrete approximation of the angle count: trees are included in the sample if they lie in the corresponding circles centered on the point x and their diameters are larger than the corresponding thresholds. Usually the radii of the circles take $1, 2$, eventually 3 but rarely if ever more values.

We therefore define in a general manner the inclusion indicator variables as

$$I_i(x) = \begin{cases} 1 \text{ if } x \in K_i \\ 0 \text{ if } x \notin K_i \end{cases} \tag{2.2}$$

and the **inclusion probabilities** for a single tree as well as for any pair of trees according to:

$$\begin{aligned} \pi_i &= P(I_i(x) = 1) = E_x\left(I_i(x)\right) = \frac{\lambda\left(K_i \cap F\right)}{\lambda(F)} \\ \pi_{ij} &= P(I_i(x)I_j(x) = 1) = E_x\left(I_i(x)I_j(x)\right) = \frac{\lambda\left(K_i \cap K_j \cap F\right)}{\lambda(F)} \end{aligned} \tag{2.3}$$

In this context the local density is the random variable defined by:

$$Y(x) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{I_i(x)Y_i}{\pi_i} \tag{2.4}$$

Note that $Y(x)$ is, but for the constant $\lambda(F)$, the Horwitz-Thompson estimate [3, 10, 20, 21].

Remark:

- The surface area of the forest $\lambda(F)$ appears only formally in formula 2.4. For effective calculation we need only to know $\lambda(F)\pi_i$, which by 2.3 is equal to the surface area of the circle, eventually corrected for boundary effects, i.e. $\lambda(K_i \cap F)$. We prefer to work with probabilities instead of surface areas because of the clear probabilistic background of the problem and the link with the mathematical statistical literature.

Given the π_i the local density $Y(x)$ is a function $Y(.)\ x \in F \mapsto Y(x)$ which by construction satisfies

$$E_x\left(Y(x)\right) = \frac{1}{\lambda(F)} \int_F Y(x)dx = \frac{1}{\lambda(F)} \sum_{i=1}^{N} Y_i = \bar{Y}_F \tag{2.5}$$

In other words, the spatial average of the local density is equal to the true density of the response variable. **The above formulation transforms the problem of estimating a sum over a finite population $\mathcal{P}$ of trees into the problem of estimating the integral of a function over a domain; in other words, one can view $Y(x), x \in F$ as an infinite population $\mathcal{Y}$.** The infinite population framework is, in my opinion, simpler and better suited for forest inventory than the other approaches based on various finite populations; this is particularly true for cluster sampling as defined in section 2.3 and unavoidable for model-dependent and geostatistical techniques, see [10, 12]; further justifications have been given by Eriksson in [5, 6]. Given the inclusion circles K_i, the function $Y(.)$ is well defined and suffices to construct all the statistical quantities required for the estimation of $\bar{Y}_F$, and the trees can vanish behind the scene. On the other hand, if the question is which function $Y(.)$ should be used and when, then we have to go back to the K_i

and therefore to the trees, which we shall do in chapters 3, 4 and 5. It is unfortunate that standard sampling theory has been, up to a certain degree, misused, and still is, as a corset for forest inventory: finite populations are simply not well tailored for the intrinsically geometrical nature of the problem.

The crucial difference between the **design-based** and the **model-dependent** approaches is the following: in the design-based approach $Y(x)$ is a random variable because the point x is random, while the forest is fixed, i.e. $N, Y_i^{(m)}, u_i$ are fixed, whereas in the model-dependent framework, including geostatistics, x is given and the actual forest is considered as the realization of a complex stochastic process; $Y(x)$ is then random because the $N, Y_i^{(m)}, u_i$ are random.

One important advantage of the sampling schemes discussed in this work and of the Horwitz-Thompson estimator 2.4 is that the inclusion probabilities are known for the units (trees) drawn, and that we do not have to know them for the units not drawn. Other sampling schemes do not have this property, for instance all those relying on nearest neighbors methods (one would need actually a full census to known the inclusion probabilities, which completely defeats the point of forest sampling). In my opinion these techniques, though popular, should only be used in very special situations, for instance when the interest is focused on rare events.

Up to now the individual inclusion probabilities π_i are completely arbitrary, whereas the π_{ij} depend on the π_i and by 2.3 on the geometry of the forest. The number of trees drawn from the point x is the random variable:

$$n(x) = \sum_{i=1}^{N} I_i(x) \qquad E_x(n(x)) = \sum_{i=1}^{N} \pi_i \tag{2.6}$$

For future use we give the following identities which follow directly from the definitions

$$\begin{aligned} E_x(n^2(x)) &= \sum_{i=1,j=1}^{N} \pi_{ij} \ , \pi_{ii} := \pi_i \\ \sum_{j=1,j\neq i}^{N} \pi_{ij} &= \pi_i \left(E_x(n(x)|I_i(x) = 1) - 1 \right) \end{aligned} \tag{2.7}$$

We now consider a set s_2 of n_2 points drawn uniformly and independently of each other in the forest area F. The **one-phase one-stage estimate for simple random sampling** is defined by:

$$\widehat{Y} = \frac{1}{n_2} \sum_{x \in s_2} Y(x) \tag{2.8}$$

By 2.3 the design-based variance is easily found to be

$$V(\widehat{Y}) = \frac{1}{n_2 \lambda^2(F)} \Big\{ \sum_{i=1}^{N} \frac{Y_i^2(1-\pi_i)}{\pi_i} + \sum_{i\neq j}^{N} \frac{Y_i Y_j(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} \Big\} = \frac{1}{n_2} V_s \tag{2.9}$$

where

$$V_s = \frac{1}{\lambda(F)} \int_F \left(Y(x) - \ \bar{Y} \right)^2 dx$$

is the variance of the local density under simple random sampling. In the above equation and thereafter all sums written as

$$\sum_{i\neq j}^{K} a_i a_j$$

are always to be understood as

$$\sum_{1\leq i,j\leq K} a_i a_j - \sum_{i=1}^{K} a_i^2$$

unless explicitly stated otherwise.

Since the $Y(x)$ are identically and independently distributed we obtain by first principles immediately the following unbiased estimate of variance

$$\widehat{V}(\widehat{Y}) = \frac{1}{n_2}\frac{1}{n_2-1}\sum_{x\in s_2}\left(Y(x)-\widehat{Y}\right)^2 \tag{2.10}$$

In the definition of the local density $Y(x)$ the inclusion probabilities π_i do already take into account boundary effects. An alternative approach [10] is to draw the random point x uniformly in a domain $\widetilde{F} \supset F$ such that $\forall i \quad K_i \subset \widetilde{F}$. There are then no boundary effects, but we get instead

$$E_{x\in\widetilde{F}}\left(Y(x)\right) = \frac{1}{\lambda(\widetilde{F})}\int_{\widetilde{F}} Y(x)dx = \frac{1}{\lambda(\widetilde{F})}\sum_{i=1}^{N} Y_i$$

The theoretical variance 2.9 depends via the π_{ij} on the spatial structure of the forest: i.e. all other things being equal, displacing the trees will change the variance, which, as we shall see, is the source of great difficulties, and is one of the main reasons to introduce the concept of anticipated variance.

For completeness we briefly outline the standard set-up of sampling theory and its relation to forest sampling.
Given a population $\mathcal{P}$ of N individuals $u_i, i \in \{1, 2\ldots, N\}$ a sampling design p is a probability function defined on all 2^N subsets of $\mathcal{P}$. Obviously, in practice, $p(s) = 0$ for most subsets s and the empty set has probability zero in this set-up, i.e. $p(\emptyset) = 0$. In forest sampling, however, it may well happen that from a point x no tree is drawn. Formally we have drawn the empty set $\emptyset$, and this happens with probability $p(\emptyset) = P\{I_i(x) = 0, \forall i\}$, in such a case $Y(x) = 0$. Note also that in standard sampling theory N is known (most sampling schemes used rests upon a complete list of the individuals), whereas it is generally unknown, but fixed, in forest sampling. Let us denote by $C(i)$ the set of all samples s containing the unit u_i, and by $C(i,j)$ the set of all samples s containing the pair $\{u_i, u_j\}$. Since only one sample s is drawn, the inclusion probabilities can also be defined as

$$\pi_i = \sum_{s\in C(i)} p(s)$$

and

$$\pi_{ij} = \sum_{s\in C(i,j)} p(s)$$

Most sampling design used in standard sampling theory are **non-informative**, i.e. $p(s)$ does not depend explicitly on the response variable Y_i for $i \in s$ if $p(s) > 0$ (more generally, given a set of auxiliary variables $Z_i^{(k)}$ then, conditionally on the values taken by the $Z_i^{(k)}$, the I_i and Y_i are independent, see [2]). This concept is important when expectations with respect to model and design probabilities are required (for non-informative designs the order of the expectations is irrelevant). In this sense the angle count method described above is an example of an informative design, at the tree level, for the basal area. In the infinite population set-up, however, we are sampling the function $Y(x)$ with uniform sampling and the design is non-informative at the plot level (the design-probability density function $\frac{dx}{\lambda(F)}$ does not

depend on the value $Y(x)$). In contrast to classical sampling theory where one draws a unique and sufficiently large sample of units for which the π_i, π_{ij} are usually known prior to sampling (list sampling), forest sampling draws a sufficiently large number of independents points $x \in s_2$, each of them selecting a relatively small random number of trees , for which the π_i can be exactly determined a posteriori; π_{ij} could be in principle calculated; this is however not necessary since, as we have seen, one can obtain an unbiased estimate of the theoretical variance by first principles.

Having defined the main concepts we can now proceed further to cluster sampling.

2.3 One-phase one-stage cluster random sampling scheme

A cluster of (nominal) size M is determined by a fixed set of M vectors $e_l \in \Re^2, \quad l = 1, \ldots M$. Without loss of generality we shall assume that one of the e_l, say e_1 is the null vector.

A correct definition of cluster sampling requires some technical details (which are crucial when designing simulations!). For any set $A \subset \Re^2$ let us denote by A_l the set $A + e_l = \{x | \exists a \in A, x = a + e_l\}$. We make the key assumption that the set A is large enough to ensure $F \subset A_l \quad l = 1, \ldots M$. For instance one could take any set A containing the set $\cup_{l=1}^{M} \{F - e_l\}$.

We now draw a random point x uniformly in A. The points $x_l = x + e_l$ are obviously uniformly distributed in A_l. With the above convention the origin x of the cluster is always the point x_1. We first note that given $x_l \in F$, x_l is uniformly distributed in F. Indeed, for any set $B \in \Re^2$ we have, as $F \subset A_l$ and $\lambda(A_l) = \lambda(A)\ \forall l$

$$P\left(x_l \in B | x_l \in F\right) = \frac{P\left(x_l \in B \cap F\right)}{P\left(x_l \in F\right)} = \frac{\frac{\lambda(B \cap F \cap A_l)}{\lambda(A_l)}}{\frac{\lambda(F \cap A_l)}{\lambda(A_l)}} = \frac{\lambda\left(B \cap F\right)}{\lambda\left(F\right)} \tag{2.11}$$

by 2.5 and the above we also get $E_x\left(Y(x_l) | x_l \in F\right) = \bar{Y}$. We introduce the indicator variable of the set F (the definition is obviously valid for any set) as

$$I_F(x) = \begin{cases} 1 \text{ if } x \in F \\ 0 \text{ if } x \notin F \end{cases}$$

The number of points per cluster falling into the forest area is the random variable defined by

$$M(x) = \sum_{l=1}^{M} I_F(x_l) \tag{2.12}$$

whereas the local density defined at the cluster level is given by

$$Y_c(x) = \frac{\sum_{l=1}^{M} I_F(x_l) Y(x_l)}{M(x)} \tag{2.13}$$

By using 2.11 and 2.5 we get the following important relations:

$$E_x M(x) = \sum_{l=1}^{M} P\left(x_l \in F\right) = \sum_{l=1}^{M} \frac{\lambda\left(F \cap A_l\right)}{\lambda\left(A_l\right)} = M \frac{\lambda(F)}{\lambda\left(A\right)} \tag{2.14}$$

$$E_x\left(\sum_{l=1}^{M} I_F(x_l)Y(x_l)\right) = \sum_{l=1}^{M} P\left(I_F(x_l)=1\right) E_{x_l}\left(Y(x_l)|x_l \in F\right)$$
$$= \sum_{l=1}^{M} \frac{\lambda\left(F \cap A_l\right)}{\lambda\left(A_l\right)}\ \bar{Y} = M\frac{\lambda(F)}{\lambda\left(A\right)}\ \bar{Y} \tag{2.15}$$

If we now draw n_2 points $x \in s_2$ independently and uniformly in the set A we generate n_2 clusters of effective sizes $M(x)$. The **one-phase one-stage estimate for cluster random sampling** is defined by

$$\widehat{Y}_c = \frac{\sum_{x\in s_2} M(x)Y_c(x)}{\sum_{x\in s_2} M(x)} \tag{2.16}$$

Note that 2.16 is formally the average of all $Y(x_l)$ ignoring the cluster structure and that it is the ratio of two random variables, which explains why cluster sampling is technically slightly more complicated than random sampling. Dividing numerator and denominator of 2.16 by n_2 we have by 2.15 and 2.14

$$\lim_{n_2\to\infty} E_{x\in A}(\widehat{Y}_c) = \frac{E_x\left(\sum_{l=1}^{M} I_F(x_l)Y(x_l)\right)}{E_x M(x)} = \ \bar{Y}$$

At this point it is important to realize that $Y_c(x)$ does not yield in general an unbiased estimate, not even asymptotically, i.e.

$$\lim_{n_2\to\infty} E_{x\in A}\frac{1}{n_2}\sum_{x\in s_2} Y_c(x) \neq \ \bar{Y}$$

More precisely one can show [10] that the cluster sampling point estimate 2.16 is asymptotically unbiased in the sense that

$$E_{x\in A}\widehat{Y}_c = \bar{Y} + O\left(n_2^{-1}\right) \tag{2.17}$$

and that its theoretical asymptotic variance is given by

$$V(\widehat{Y}_c) = \frac{1}{n_2}\frac{E_x M^2(x)(Y_c(x) - \ \bar{Y})^2}{E_x^2 M(x)} + O\left(n_2^{-2}\right) \tag{2.18}$$

According to [10] the variance can be estimated asymptotically with a bias of order $O(n_2^{-2})$ by

$$\widehat{V}(\widehat{Y}_c) = \frac{1}{n_2(n_2-1)}\sum_{x\in s_2}\left(\frac{M(x)}{\bar{M}_2}\right)^2\left(Y_c(x) - \widehat{Y}_c\right)^2 \tag{2.19}$$

where $\bar{M}_2 = \frac{1}{n_2}\sum_{x\in s_2} M(x)$ is the average number of points per cluster falling into the forest area. Note that void clusters, i.e. $M(x) = 0$, do not contribute to the point estimate, nor to the variance.

For a better intuitive understanding we quote a result given in [10], which expresses the variance under cluster sampling as a function of the variance under simple random sampling V_s, the intra-cluster correlation coefficient ρ and a term describing the topological characteristics of the forest, namely:

$$V(\widehat{Y}_c) = \frac{1}{n_2 E_x M(x)} V_s\left(1 + \rho\left(E_x M(x) - 1\right) + \rho\frac{V_x M(x)}{E_x M(x)}\right) + O(n_2^{-2}) \tag{2.20}$$

where

$$\rho = \frac{E_x\sum_{l\neq m}^{M} I_F(x_l)I_F(x_m)\left(Y(x_l) - \ \bar{Y}\right)\left(Y(x_m) - \ \bar{Y}\right)}{V_s E_x M(x)\left(M(x)-1\right)} \tag{2.21}$$

One can estimate ρ by equating a sample version of 2.20 and 2.19.

In general the intra-cluster correlation coefficient ρ is positive and the variance under cluster sampling is larger than under simple random sampling, obviously while keeping constant the function $Y(.)$ as well as the total number of points, i.e. $n_{2,simple} = n_{2,cluster} E_x M(x)$. The inflation factor is given by 2.20. Formula 2.20 generalizes a classical result, in which the cluster size is kept constant [3], which is of course not suitable for forest inventory. The reason for using cluster sampling is the reduction of transport costs, a matter we shall discuss later. Finally, let us note that one can view the points of a systematic grid falling into the forest area as a single very large cluster, which explains why one cannot estimate the design-based variance with systematic grids as used in forest inventory. This also implies that density estimates are only asymptotically unbiased, in contrast to estimates of totals, see [10] for more details; in practice the finite sample bias seems to be negligible.

2.4 One-phase two-stage simple random sampling

In many applications the costs of measuring the response variable Y_i are high, for instance a good determination of the volume may require to measure the diameter at breast height, the diameter at 7m above the ground and the height of the tree in order to use a three-way yield table. On the other hand, one could use a coarser but cheaper approximation of the volume based only on the diameter at breast height. It appears therefore natural to get the three measurements only on a subsample of the trees. We now formalize this simple idea. For each point $x \in s_2$ trees are drawn with probabilities π_i. The set of selected trees is denoted by $s_2(x)$. On each of the selected trees $i \in s_2(x)$ one gets an approximation Y_i^* of the exact value Y_i. In the finite set $s_2(x)$ one draws a subsample $s_3(x) \subset s_2(x)$ of trees. For each tree $i \in s_3(x)$ one measures the exact variable Y_i. Let us now define the second stage indicator variable:

$$J_i(x) = \begin{cases} 1 \text{ if } i \in s_3(x) \\ 0 \text{ if } i \notin s_3(x) \end{cases} \tag{2.22}$$

Note that by construction $I_i(x)J_i(x) = J_i(x)$. Before going further we must introduce some notation for expectations and variances . In our general context the subindex 1 refers to the first phase (which collects the auxiliary information in the large sample s_1 and which will be thoroughly defined later on), the subindex 2 refers to the second phase, which collects the terrestrial information on the first stage trees, and finally the subindex 3, which refers to the second stage trees. We need three sub-indexes since in general three random selections are involved. According to standard notation in probability theory we shall use the notation $E_{2,3}(.), V_{2,3}(.), E_{3|2}(.), V_{3|2}(.)$ for the overall expectation and variance under the random selections $(2,3)$ and for the conditional expectation and variance of the second stage procedure given the second-phase first-stage selection. We recall the following important rules for the calculation of expected value and variance of an arbitrary random variable Z depending on the random selection $(2,3)$.

$$\begin{aligned} E_{2,3}(Z) &= E_2\left(E_{3|2}(Z)\right) \\ V_{2,3}(Z) &= E_2\left(V_{3|2}(Z)\right) + V_2\left(E_{3|2}(Z)\right) \end{aligned} \tag{2.23}$$

Hence we have

$$\begin{aligned} E_{2,3}(J_i(x)) &= E_2 E_{3|2}(J_i(x)I_i(x)) \\ &= E_2 I_i(x) E_{3|2}(J_i(x)|I_i(x)) \\ &= P(J_i(x) = 1 | I_i(x) = 1) P(I_i(x) = 1) := p_i \pi_i \end{aligned} \tag{2.24}$$

where we have introduced the **second-stage** conditional inclusion probability $p_i = P(J_i(x) = 1|I_i(x) = 1)$ which so far depends on the label i and is therefore completely arbitrary.

We now assume that trees in $s_2(x)$ are sampled **independently of each other**, so that $p_{ij} = P(J_i(x)J_j(x) = 1|I_i(x)I_j(x) = 1) = p_i p_j$, i.e. we have binomial sampling with possibly unequal probability at the second stage. The advantage of the proposed scheme is that the field crew collects the required information on the first stage trees one by one, enters the data in the portable computer, which according to a random number tells then immediately whether to take further measurements or not. Other schemes are of course possible, but not so easy to implement (one needs a list of all first stage trees at the point x and possibly also other points) and not necessarily better.

To construct a good point estimate we need **the residual** $R_i = Y_i - Y_i^*$ which is known only for trees $i \in s_3(x)$. The generalized local density $Y^*(x)$ is defined according to

$$\begin{aligned} Y^*(x) &= \frac{1}{\lambda(F)} \left(\sum_{i=1}^{N} \frac{I_i(x) Y_i^*}{\pi_i} + \sum_{i=1}^{N} \frac{I_i(x) J_i(x) R_i}{\pi_i p_i} \right) \\ &= \frac{1}{\lambda(F)} \left(\sum_{i \in s_2(x)} \frac{Y_i^*}{\pi_i} + \sum_{i \in s_3(x)} \frac{R_i}{\pi_i p_i} \right) \end{aligned} \tag{2.25}$$

2.25 is an adaptation of an estimate first proposed by Saerndal [22] in classical sampling theory.

We assume here that the prediction Y_i^* of Y_i is based on an **external model**, i.e. that the corresponding model is not adjusted with the data collected by the inventory sample. In practice, one may have to use an **internal model**, where model fitting is performed with the same inventory data; this renders the exact calculation of bias and variance almost impossible in general. However, there is some theoretical and empirical evidence that internal model can be treated as external model in large samples, see [21, 10].

Since $Y_i = Y_i^* + R_i$ and $E_{3|2} J_i(x) = p_i$ we have

$$E_{2,3}(Y^*(x)) = E_2(Y(x)) = \bar{Y} \tag{2.26}$$

so that 2.25 is also unbiased. Since the second stage trees are drawn independently of each other we have

$$\begin{aligned} V(x) = V_{3|2}(Y^*(x)) &= V_{3|2} \left(\frac{1}{\lambda(F)} \sum_{i \in s_3(x)} \frac{R_i}{\pi_i p_i} \right) \\ &= \frac{1}{\lambda^2(F)} \left(\sum_{i \in s_2(x)} \frac{R_i^2 (1 - p_i)}{\pi_i^2 p_i} \right) \\ E_2 V_{3|2}(Y^*(x)) = E_x V(x) &= \frac{1}{\lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{2.27}$$

The one-phase two-stage estimate for simple random sampling is defined by:

$$\widehat{Y}^* = \frac{1}{n_2} \sum_{x \in s_2} Y^*(x) \tag{2.28}$$

Its theoretical variance is by 2.23 easily found to be

$$V(\widehat{Y}^*) = \frac{1}{n_2} V_x(Y(x)) + \frac{1}{n_2} E_x V(x) \tag{2.29}$$

and using 2.9, 2.27 this is equivalent to

$$\begin{aligned} V(\widehat{Y}^*) = &\frac{1}{n_2\lambda^2(F)}\Big\{\sum_{i=1}^{N}\frac{Y_i^2(1-\pi_i)}{\pi_i} + \sum_{i\neq j}^{N}\frac{Y_iY_j(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j}\Big\} \\ &+ \frac{1}{n_2\lambda^2(F)}\Big\{\sum_{i=1}^{N}\frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N}\frac{R_i^2}{\pi_i}\Big\} \end{aligned} \tag{2.30}$$

It is in some sense remarkable that the usual variance estimate can also be used in this case, more precisely one has

$$\begin{aligned} \widehat{V}(\widehat{Y}^*) &= \frac{1}{n_2(n_2-1)}\sum_{x\in s_2}(Y^*(x)-\widehat{Y}^*)^2 \\ E_x(\widehat{V}(\widehat{Y}^*)) &= V(\widehat{Y}^*) \end{aligned} \tag{2.31}$$

Proof:

$$\begin{aligned} n_2(n_2-1)\widehat{V}(\widehat{Y}^*) &= \sum_{x\in s_2}Y^*(x)^2 - \frac{\left(\sum_{x\in s_2}Y^*(x)\right)^2}{n_2} \\ &= \sum_{x\in s_2}Y^*(x)^2 - \frac{1}{n_2}\left(\sum_{x\in s_2}Y^*(x)^2 + \sum_{x\neq y\in s_2}Y^*(x)Y^*(y)\right) \end{aligned}$$

Taking the expectation according to 2.23 and using 2.26 and the independence of $Y^*(x), Y^*(y)$ for $x \neq y$ we obtain

$$\begin{aligned} n_2(n_2-1)E_{2,3}(\widehat{V}(\widehat{Y}^*)) &= \frac{n_2-1}{n_2}E_2\left(\sum_{x\in s_2}(V_{3|2}Y^*(x)+Y^2(x))\right) \\ &\quad - \frac{1}{n_2}E_2\sum_{x\neq y\in s_2}Y(x)Y(y) \end{aligned}$$

and therefore

$$\begin{aligned} n_2(n_2-1)E_{2,3}(\widehat{V}(\widehat{Y}^*)) &= (n_2-1)\left(E_xV(x)+V_xY(x)+\bar{Y}^2\right) - \frac{n_2(n_2-1)}{n_2}\bar{Y}^2 \\ &= (n_2-1)\left(E_xV(x)+V_xY(x)\right) \end{aligned}$$

and the result.

In some instances one may want to know the variance of the exact density $Y(x)$ when only $Y^*(x)$ is available. To this end let us note first that

$$\widehat{V}(x) = \frac{1}{\lambda^2(F)}\sum_{i\in s_3(x)}\frac{R_i^2(1-p_i)}{\pi_i^2p_i^2}$$

is an unbiased estimate of $V(x)$, in the sense that

$$E_{3|2}\widehat{V}(x) = V(x)$$

hence by 2.29 we get an unbiased estimate of $V(\widehat{Y})$ through

$$\widehat{V}(Y(x)) = n_2\widehat{V}(\widehat{Y}^*) - \frac{1}{n_2}\sum_{x\in s_2}\widehat{V}(x) \tag{2.32}$$

The one-phase two-stage sampling scheme has been implemented as defined here in the second Swiss National Inventory. Based on data from the first Swiss National Inventory, which selected roughly 120'000 first stage trees with 10'000 points and 40'000 second stage trees with constant probability $p_i \equiv \frac{1}{3}$, it was found that one could reduce the number of second stage trees to 10'000 while ensuring the same accuracy, and even decreasing the bias in small areas, simply by optimizing the p_i, a topic we shall discuss later, and using the point estimate 2.25, which explicitly takes into account the residuals. Quoting Einstein: there is nothing more practical than a good theory!

2.5 One-phase two-stage cluster random sampling

We use the same concepts and notation as in section 2.3. Assuming that the **second-stage trees are drawn independently of each other in each point of each cluster**, we can generalize 2.13 in a straightforward way, i.e.

$$Y_c^*(x) = \frac{\sum_{l=1}^{M} I_F(x_l) Y^*(x_l)}{M(x)} \tag{2.33}$$

where $Y^*(x_l)$ is the generalized local density at point $x_l = x + e_l$. Note that

$$E_{3|2} Y_c^*(x) = Y_c(x)$$

The **one-phase two-stage point estimate for cluster sampling** is then defined in perfect analogy to 2.16 as:

$$\widehat{Y}_c^* = \frac{\sum_{x \in s_2} M(x) Y_c^*(x)}{\sum_{x \in s_2} M(x)} \tag{2.34}$$

Because of 2.17 and the above remark, it is clear that 2.34 is asymptotically unbiased, i.e.

$$E_{x \in A} \widehat{Y}_c^* = \bar{Y} + O\left(n_2^{-1}\right) \tag{2.35}$$

To calculate the theoretical variance we use the decomposition 2.23. The second term is simply the variance of the one-phase one-stage estimate. The first is given by

$$\begin{aligned}
E_2 V_{3|2}(\widehat{Y}_c^*) &= E_2 \left(V_{3|2} \frac{\sum_{x \in s_2} \sum_{l=1}^{M} I_F(x_l) Y^*(x_l)}{\sum_{x \in s_2} M(x)} \right) \\
&= E_2 \left(\frac{\sum_{x \in s_2} \sum_{l=1}^{M} I_F(x_l) V(x_l)}{(\sum_{x \in s_2} M(x))^2} \right) \\
&= \frac{\sum_{x \in s_2} \sum_{l=1}^{M} E_2(V(x_l) | x_l \in F) P(x_l \in F)}{n_2^2 E_2^2 M(x)} + O(n_2^{-2}) \\
&= \frac{n_2 E_{x \in F} V(x) E_{x \in A} M(x)}{n_2^2 E_{x \in A}^2 M(x)} + O(n_2^{-2}) \\
&= \frac{E_{x \in F} V(x)}{n_2 E_{x \in A} M(x)} + O(n_2^{-2})
\end{aligned}$$

where we have used 2.11. Hence, by using the above result, 2.23 and 2.18 the asymptotic theoretical variance of the one-phase two-stage point estimate under cluster sampling is given by

$$V(\widehat{Y}_c^*) = \frac{1}{n_2} \frac{E_{x \in A} M^2(x) (Y_c(x) - \bar{Y})^2}{E_{x \in A}^2 M(x)} + \frac{E_{x \in F} V(x)}{n_2 E_{x \in A} M(x)} + O(n_2^{-2}) \tag{2.36}$$

In analogy with the result 2.31 for one-phase two-stage simple random sampling one can construct an asymptotically unbiased estimate of variance according to

$$\begin{aligned} \widehat{V}(\widehat{Y}_c^*) &= \frac{1}{n_2(n_2-1)} \sum_{x \in s_2} \left(\frac{M(x)}{\bar{M}_2} \right)^2 \left(Y_c^*(x) - \widehat{Y_c^*} \right)^2 \\ E_{2,3}(\widehat{V}(\widehat{Y}_c^*)) &= V(\widehat{Y}_c^*) + O(n_2^{-2}) \end{aligned} \tag{2.37}$$

We give only the flavour of the proof, which can be made rigourous, by using the weak law of large numbers (convergence in probability), which implies here convergence in the mean as all random variables are bounded in practice.

Proof:

Asymptotically, one can replace $\bar{M}_2$ and $\widehat{Y}_c^*$ by their true values, likewise $n_2 - 1$ by n_2; multiplying then both sides of 2.37 by $n_2^2 \bar{M}_2^2$ it remains to show that

$$\begin{aligned} E_{2,3} \sum_{x \in s_2} M^2(x)(Y_c(x)^* - \bar{Y})^2 = n_2 E_{x \in A} M^2(x)(Y_c(x) - \bar{Y})^2 \\ + n_2 E_{x \in A} M(x) E_{x \in F} V(x) \end{aligned}$$

Setting $V_c(x) = V_{3|2} Y_c^*(x)$ and using the decomposition 2.23 we see first that the left hand side is equal to

$$n_2 E_{2,3} \left(M^2(x) Y_c^*(x)^2 - 2M^2(x) Y_c^*(x) \bar{Y} + M^2(x)\ \bar{Y}^2 \right)$$

and then equal to

$$n_2 (E_2 M^2(x)(Y_c^2(x) + V_c(x)) + n_2(-2M^2(x) Y_c(x) \bar{Y} + M^2(x)\ \bar{Y}^2)$$

to finally obtain

$$n_2 E_{x \in A} M^2(x)(Y_c(x) - \bar{Y})^2 + n_2 E_{x \in A} \sum_{l=1}^{M} I_F(x_l) V(x_l)$$

Using 2.11 the second term can be rewritten as

$$n_2 \sum_{l=1}^{M} E_{x \in A} V(x_l | x_l \in F) P(x_l \in F) = n_2 E_{x \in F} V(x) E_{x \in A} M(x)$$

which ends the proof.

From a theoretical design-based point of view this is essentially all there is to know about pure terrestrial inventories performed at one time point. We now turn to combined forest inventories.

2.6 Two-phase one-stage simple random sampling

The **first phase** draws a large sample s_1 of n_1 points $x_i \in s_1$ independently and uniformly distributed in the forest area F. In each of these points auxiliary information is collected, very often of purely qualitative nature, for instance after interpretation of aerial photographs. The **second phase** draws a small sample $s_2 \subset s_1$ of n_2 points according to **equal probability sampling without replacement**. In each point $x \in s_2$ the terrestrial inventory provides the local density $Y(x)$. For points $x \in s_1 \setminus s_2$, i.e. in the large but not in the small sample, only the auxiliary

information is available, which nevertheless allows one to make a prediction $\widehat{Y}(x)$ of the true local density $Y(x)$ in the forest. Strictly speaking, we shall assume that this prediction is given by an **external model**, i.e. not adjusted with the data of the actual inventory. Let us give the probably most important example: the stand structure is determined by interpretation of aerial photographs in s_1, which by means of pre-existing yield tables allows the inventorist to make a reasonable prediction of timber volume per ha or number of stems per ha. However, if an external model is not available, an **internal model** has to be fitted first. Usually this is done in the following way: the auxiliary information at the point x is coded into a vector $Z(x) \in \Re^m$ and the prediction is obtained via a linear model, i.e. $\widehat{Y}(x) = Z(x)^t\beta$ (the upper index t meaning transposition of the vector). The estimation of the unknown parameter vector β can be done in several ways, in particular by ignoring completely sampling theory and using standard statistical tools, like analysis of variance or multiple regression. Alternatively, one can estimate β within the framework of sampling theory. There is some evidence that the choice of the estimation procedures is of secondary importance and that internal models can be treated as external models if n_2 is sufficiently large; see [10] for more details. Predictions based on models, particularly external models, should not, however, be blindly trusted and it is intuitively clear that deviations between model and reality should be taken into account. This is done by considering the **residual** $R(x) = Y(x) - \widehat{Y}(x)$. **The two-phase one-stage estimate for simple random sampling** is defined by:

$$\widehat{Y}_{reg} = \frac{1}{n_1}\sum_{x\in s_1}\widehat{Y}(x) + \frac{1}{n_2}\sum_{x\in s_2}R(x) \tag{2.38}$$

Before going further let us transpose the rules 2.23 to the present set-up

$$\begin{aligned} E_{1,2}(Z) &= E_1\left(E_{2|1}(Z)\right) \\ V_{1,2}(Z) &= E_1\left(V_{2|1}(Z)\right) + V_1\left(E_{2|1}(Z)\right) \end{aligned} \tag{2.39}$$

Given s_1 the properties of 2.38 are governed by standard sampling theory of finite population, see [3, 20], hence one immediately has

$$E_{2|1}\frac{1}{n_2}\sum_{x\in s_2}R(x) = \frac{1}{n_1}\sum_{x\in s_1}R(x)$$

Since $Y(x) = \widehat{Y}(x) + R(x)$ we have by 2.39

$$E_{1,2}(\widehat{Y}_{reg}) = E_1\frac{1}{n_1}\sum_{x\in s_1}Y(x) = \bar{Y}$$

and therefore unbiasedness. Furthermore

$$V_{2|1}\frac{1}{n_2}\sum_{x\in s_2}R(x) = \left(1-\frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{1}{n_1-1}\sum_{x\in s_1}(R(x)-\bar{R}_1)^2$$

where we set $\bar{R}_i = \frac{1}{n_i}\sum_{x\in s_i}R(x)$ for $i = 1, 2$. Consequently we have

$$E_1V_{2|1}(\widehat{Y}_{reg}) = \left(1-\frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{1}{\lambda(F)}\int_F(R(x)-\bar{R})^2dx$$

where

$$\bar{R} = \frac{1}{\lambda(F)}\int_F R(x)dx$$

According to 2.39 we can state the main result

$$E_{1,2}(\widehat{Y}_{reg}) = \bar{Y}$$

$$V_{1,2}(\widehat{Y}_{reg}) = \left(1 - \frac{n_2}{n_1}\right) \frac{1}{n_2} \frac{1}{\lambda(F)} \int_F (R(x) - \bar{R})^2 dx + \frac{1}{n_1 \lambda(F)} \int_F (Y(x) - \bar{Y})^2 dx \tag{2.40}$$

It is clear that the variance can be estimated by

$$\widehat{V}_{1,2}(\widehat{Y}_{reg}) = \left(1 - \frac{n_2}{n_1}\right) \frac{1}{n_2} \frac{1}{n_2 - 1} \sum_{x \in s_2} (R(x) - \bar{R}_2)^2 + \frac{1}{n_1} \frac{1}{n_2 - 1} \sum_{x \in s_2} (Y(x) - \bar{Y}_2)^2 \tag{2.41}$$

$$E_{1,2}\widehat{V}_{1,2}(\widehat{Y}_{reg}) = V_{1,2}(\widehat{Y}_{reg})$$

Remarks:

- The lower index *reg* indicates that most internal prediction models rely on regression techniques, which contain the post-stratified estimates as a particular case, see sections 2.9 and 2.10.
- We have implicitly assumed that points $x \in s_2$ are error-free, in the sense that the prediction $\widehat{Y}(x)$ and the observation $Y(x)$ corresponds exactly to the same point; this, of course, is only approximately true in practice.
- In practice the vector $Z(x)$ and hence the prediction $\widehat{Y}(x)$ is often based on stand or stratum characteristics. The point x determines the stratum. If the point is near stand boundaries it may be wise to redefine $Y(x)$ by treating the area outside the stand of x as non-forest area. We shall come back to this difficulty in chapter 3 (revised Horwitz-Thompson estimator).

2.7 Two-phase one-stage cluster random sampling

This is a straightforward generalization of the previous section. The first phase draws a large sample of n_1 clusters whose origins $x \in s_1$ are uniformly and independently distributed in $A \supset F$; the geometry of the clusters is determined as usual by the M vectors $e_l \in \Re^2$. In each point $x_l = x + e_l$ of a given cluster one collects the auxiliary information required to make a prediction $\widehat{Y}(x_l)$. The second phase draws a subsample of n_2 clusters out of the n_1, with origin $x \in s_2 \subset s_1$ according to equal probability sampling without replacement. For any given cluster with $x \in s_2$ the local density $Y(x_l)$ is determined in each point x_l of the cluster. We use the same notation as in one phase cluster sampling.

The two-phase one-stage estimate for cluster random sampling is defined by:

$$\widehat{Y}_{c,reg} = \frac{\sum_{x \in s_1} M(x) \widehat{Y}_c(x)}{\sum_{x \in s_1} M(x)} + \frac{\sum_{x \in s_2} M(x) R_c(x)}{\sum_{x \in s_2} M(x)} \tag{2.42}$$

where the residuals at the cluster level are defined according to

$$R_c(x) = \frac{\sum_{l=1}^{M} I_F(x_l)(Y(x_l) - \widehat{Y}(x_l))}{\sum_{l=1}^{M} I_F(x_l)}$$

Using 2.39 and the results obtained for one-phase cluster sampling as well as two-phase one-stage simple random sampling it is not difficult to show (details are given in [10]) that

$$E_{1,2}\widehat{Y}_{c,reg} = \bar{Y} + O(n_2^{-1})$$
$$V_{1,2}\widehat{Y}_{c,reg} = \frac{1}{n_1}\frac{E_x M^2(x)(Y_c(x) - \bar{Y})^2}{E_x^2 M(x)} + \left(1 - \frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{E_x M^2(x)(R_c(x) - \bar{R})^2}{E_x^2 M(x)} + O(n_2^{-2}) \tag{2.43}$$

Likewise, an asymptotically unbiased estimate of the variance is given by

$$\widehat{V}_{1,2}\widehat{Y}_{c,reg} = \frac{1}{n_1}\frac{1}{n_2 - 1}\sum_{x\in s_2}\left(\frac{M(x)}{\bar{M}_2}\right)^2 (Y_c(x) - \widehat{Y}_2)^2 + \left(1 - \frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{1}{n_2 - 1}\sum_{x\in s_2}\left(\frac{M(x)}{\bar{M}_2}\right)^2 (R_c(x) - \widehat{R}_2)^2 \tag{2.44}$$

where

$$\bar{M}_2 = \frac{\sum_{x\in s_2} M(x)}{n_2}, \widehat{R}_2 = \frac{\sum_{x\in s_2} M(x)R_c(x)}{\sum_{x\in s_2} M(x)}$$

We emphasize the fact that so far we have not assumed that the mean value of the residuals is zero, which is particularly relevant when using external models.

2.8 Two-phase two-stage simple random sampling

In this procedure all three random selections $(1, 2, 3)$ are involved. The idea is exactly as in two-phase one-stage simple random sampling, but for the fact that the unknown true local density $Y(x)$ at the point x is replaced by its estimate, i.e. the generalized local density $Y^*(x)$ defined in 2.25.

The two-phase two-stage estimate for simple random sampling is defined by:

$$\widehat{Y}^*_{reg} = \frac{1}{n_1}\sum_{x\in s_1}\widehat{Y}(x) + \frac{1}{n_2}\sum_{x\in s_2} R^*(x) \tag{2.45}$$

where we have set $R^*(x) = Y^*(x) - \widehat{Y}(x)$.

Using 2.23 and 2.39 we have

$$E_{1,2,3}\widehat{Y}^*_{reg} = E_1 E_{2|1} E_{3|1,2}\widehat{Y}^*_{reg} = \bar{Y}$$

so that the estimate is design-unbiased. Likewise, we obtain

$$V_{2,3|1}\widehat{Y}^*_{reg} = E_{2|1}V_{3|1,2}\widehat{Y}^*_{reg} + V_{2|1}E_{3|1,2}\widehat{Y}^*_{reg}$$

We obtain after some elementary algebra and using the notation of sections 2.4 and 2.6 the main result

$$E_{1,2,3}\widehat{Y}^*_{reg} = \bar{Y}$$
$$V_{1,2,3}(\widehat{Y}^*_{reg}) = \frac{1}{n_1}V_x Y(x) + \left(1 - \frac{n_2}{n_1}\right)\frac{1}{n_2}V_x R(x) + \frac{1}{n_2}E_x V(x) \tag{2.46}$$

In this formula the true residual $R(x) = Y(x) - \widehat{Y}(x)$ is of course not observable. The overall variance is therefore essentially the sum of, first, the variance of the exact local density as if we had observed it in the large sample, second, of the residual variance of the true residuals, due to replacing at the plot level observations by predictions, and, third, of the second-stage variance, due to replacing at the tree level exact measurements by predictions: a very intuitive result indeed.

Remarks

- One can rewrite the variance in 2.46 in terms of the following coefficients of determination

$$R^{*2} = \frac{V_x(\widehat{Y}(x))}{V_x(Y^*(x))} < R^2 = \frac{V_x(\widehat{Y}(x))}{V_x(Y(x))}$$

  Provided that residuals and predictions are uncorrelated (this is the case for adequate external models and, according to section 2.10, also for internal linear models) one has by simple algebra

$$V(\widehat{Y}^*_{reg}) = V(\widehat{Y}(x)\left(\frac{1-R^{*2}}{R^{*2}}\frac{1}{n_2} + \frac{1}{n_1}\right)$$

$$V(\widehat{Y}^*_{reg}) = V(\widehat{Y}(x)\left(\frac{1-R^{2}}{R^{2}}\frac{1}{n_2} + \frac{1}{n_1}\right) + \frac{1}{n_2}E_xV(x)$$

  One should therefore look for prediction models with high coefficients of determination (R-squared values) in order to substantially reduce the variance with two-phase sampling.

Using exactly the same arguments as in the proof of 2.31 it is easy to see that

$$E_{1,2,3}\frac{1}{n_2-1}\sum_{x\in s_2}(\widehat{Y}^*(x) - \bar{Y}^*_2)^2 = E_xV(x) + V_xY(x)$$

and that

$$E_{1,2,3}\frac{1}{n_2-1}\sum_{x\in s_2}(\widehat{R}^*(x) - \bar{R}^*_2)^2 = E_xV(x) + V_xR(x)$$

so that again the usual variance estimate holds, i.e.

$$\begin{aligned}\widehat{V}(\widehat{Y}^*_{reg}) = &\left(1-\frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{1}{n_2-1}\sum_{x\in s_2}(R^*(x) - \bar{R}^*_2)^2 \\ &+ \frac{1}{n_1}\frac{1}{n_2-1}\sum_{x\in s_2}(Y^*(x) - \bar{Y}^*_2)^2 \qquad (2.47)\end{aligned}$$

$$E_{1,2,3}\widehat{V}(\widehat{Y}^*_{reg}) = V_{1,2,3}(\widehat{Y}^*_{reg})$$

We shall see in section 2.10 on internal linear models that the above formulae hold when predictions and residuals are obtained via standard least squares regression of the $Y^*(x)$ on explanatory variables.

To illustrate the general theory we consider the important special case of post-stratification. We omit most of the tedious but elementary algebraic manipulations. The forest area F is partitioned into L disjoined strata $F = \cup_{k=1}^{L} F_k$. The ideal prediction $\widehat{Y}(x)$ for $x \in F_k$ would be the true mean $\bar{Y}_k$ of the kth stratum. Since it is unknown, we estimate it by the corresponding internal linear model (section 2.10 gives the necessary tools to do that), which in this case is a simple one-way analysis

of variance. The result is intuitively obvious and $\widehat{Y}(x)$ is simply the empirical mean of the stratum, that is $\widehat{Y}(x) = \bar{Y}^*_{2,k}$ for $x \in F_k$, where

$$\bar{Y}^*_{2,k} = \frac{1}{n_{2,k}} \sum_{x \in F_k \cap s_2} Y^*(x)$$

$n_{2,k}$ is the number of points in s_2 falling into the kth stratum. In this case, like for almost all internal linear models (see section 2.10), the residuals add up to 0, $\sum_{x \in F_k \cap s_2} R^*(x) = 0$, and we finally get

$$\widehat{Y}^*_{reg} = \frac{1}{n_1} \sum_{k=1}^{L} n_{1,k} \bar{Y}^*_{2,k}$$

where $n_{1,k}$ is the number of points of the large sample falling into the kth stratum. Obviously, $\widehat{Y}^*_{reg}$ is the standard post-stratified estimate since $\frac{n_{1,k}}{n_1}\lambda(F)$ is the estimate of the surface area of the kth stratum. The usual variance estimate within the kth stratum is defined as

$$\widehat{V}^*_k = \frac{1}{n_{2,k} - 1} \sum_{x \in F_k \cap s_2} (Y^*(x) - \bar{Y}^*_{2,k})^2$$

Simple algebra shows that in this case formula 2.47 yields

$$\widehat{V}(\widehat{Y}^*_{reg}) = \frac{1}{n_2} \sum_{k=1}^{L} \left(\frac{n_{2,k} - 1}{n_2 - 1} \right) \widehat{V}^*_k + \frac{1}{n_1} \sum_{k=1}^{L} \left(\frac{n_{2,k}}{n_2 - 1} \right) (\bar{Y}^*_{2,k} - \bar{Y}^*_2)^2 \qquad (2.48)$$

where we have set

$$\bar{Y}^*_2 = \frac{1}{n_2} \sum_{x \in s_2} Y^*(x)$$

We have a simple analytical expression for the point estimate so that we can calculate directly the variance. All there is to do is to use repeatedly 2.23 and 2.39 (conditioning on given $n_{2,k}$) and the following straightforward relations

$$p_k = \frac{\lambda(F_k)}{\lambda(F)}$$

$$V(n_{1,k}) = n_1 p_k (1 - p_k)$$

$$COV(n_{1,k}, n_{1,l}) = -n_1 p_k p_l$$

$$E(n^{-1}_{2,k} | n_{1,k}) = \left(n_{1,k} \frac{n_2}{n_1} \right)^{-1} + O(n_2^{-2})$$

$$V_k = V(Y(x) | x \in F_k)$$

Then one has

$$V(\widehat{Y}^*_{reg}) = \frac{1}{n_2} \sum_{k=1}^{L} p_k V_k + \frac{1}{n_1} \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 + \frac{1}{n_2} E_x V(x) + O(n_2^{-2}) \qquad (2.49)$$

The variance of the unstratified estimate $\widehat{Y}^* = \frac{1}{n_2} \sum_{x \in s_2} Y^*(x)$ is given approximately by

$$V(\widehat{Y}^*) \approx \frac{1}{n_2} \sum_{j=1}^{L} p_j V_j + \frac{1}{n_2} \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 + \frac{1}{n_2} E_x V(x)$$

This follows at once by using the decomposition

$$\int_F (Y(x) - \bar{Y})^2 dx = \sum_{j=1}^{L} \int_{F_j} (Y(x) - \bar{Y}_j + \bar{Y}_j - \bar{Y})^2 dx$$

and neglecting edge effects at the strata boundaries ($Y(x)$ is not exactly unbiased within each F_k since a point $x \in F_k$ may sample trees from an adjacent stratum, see sections 3.2, 3.3). Nevertheless, we see that post-stratification can reduce substantially the variance when the between-strata variance, i.e. the term $\sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2$, is large.

Let us now calculate the expected value of the variance estimate 2.48. Conditioning first on given $n_{2,k}$, by using then $E_{2,3}(V_k^*|n_{2,k}) = V_k + E_{x \in F_k} V(x)$ (which follows by the same arguments as in the proof of 2.31), $n_{2,k} n_2^{-1} \to p_k$, and finally by replacing in the last squared term the random variables by their expected values one gets

$$E_{1,2,3} \widehat{V}(\widehat{Y}^*_{reg}) \approx \frac{1}{n_2} \sum_{j=1}^{L} p_j V_j + \frac{1}{n_1} \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 + \frac{1}{n_2} E_x V(x) \tag{2.50}$$

so that, as announced, formula 2.48 and consequently also 2.47 yield indeed asymptotically unbiased estimates of variance. In this case, we have therefore shown that we can treat the internal model $\widehat{Y}(x) = \bar{Y}^*_{2,k}$ for $x \in F_k$ as the external model $\widehat{Y}(x) = \bar{Y}_k$, $x \in F_k$. This result holds for a large class of linear models (see [12] for details).

More generally, if we assume that residuals and predicted values are uncorrelated, which is the case for internal linear models adjusted by ordinary least squares (see section 2.10), then $V_x(Y(x)) = V_x(R(x)) + V_x(\widehat{Y}(x))$ and we can rewrite 2.46 as

$$V(\widehat{Y}^*_{reg}) = \frac{1}{n_2} V_x(Y(x)) + \frac{1}{n_2} E_x V(x) - \frac{1}{n_2} \left(1 - \frac{n_2}{n_1}\right) V_x(\widehat{Y}(x)) \tag{2.51}$$

For post-stratification one has to interpret, in the above formula, the predictions $\widehat{Y}(x)$ as constant and equal to the true mean of the stratum, so that $V_x(\widehat{Y}(x)) = \sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2$. Asymptotically, one can therefore do as if the predictions were exact when calculating the variance (but not for the point estimate since we must use the residuals in order to ensure unbiasedness); this requires, of course, that the prediction model yields asymptotically the true means in each stratum, a condition which will play an important role when calculating the anticipated variance.

2.9 Two-phase two-stage cluster random sampling

This sampling scheme is an immediate combination of sections 2.7 and 2.8. The first phase yield the predictions $\widehat{Y}(x_l), \widehat{Y}_c(x)$ and the second phase the generalized densities $Y^*(x_l), Y_c^*(x)$.

The two-phase two-stage estimate for cluster random sampling is defined by:

$$\widehat{Y}^*_{c,reg} = \frac{\sum_{x \in s_1} M(x) \widehat{Y}_c(x)}{\sum_{x \in s_1} M(x)} + \frac{\sum_{x \in s_2} M(x) R_c^*(x)}{\sum_{x \in s_2} M(x)} \tag{2.52}$$

where the generalized residuals at the cluster level are defined in the usual way as

$$R_c^*(x) = \frac{\sum_{l=1}^{M} I_F(x_l)(Y^*(x_l) - \widehat{Y}(x_l))}{\sum_{l=1}^{M} I_F(x_l)}$$

Using 2.23, 2.39 and the results obtained in sections 2.3, 2.5, 2.7, 2.8 one can show that

$$E_{x\in A}\widehat{Y}^*_{c,reg} = \bar{Y} + O(n_2^{-1})$$

$$\begin{aligned} V(\widehat{Y}^*_{c,reg}) = &\left(1-\frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{E_{x\in A}M^2(x)(R_c(x)-\bar{R})^2}{E^2_{x\in A}M(x)} \\ &+\frac{1}{n_1}\frac{E_{x\in A}M^2(x)(Y_c(x)-\bar{Y})^2}{E^2_{x\in A}M(x)}+\frac{1}{n_2}\frac{E_{x\in F}V(x)}{E_{x\in A}M(x)}+O(n_2^{-2}) \end{aligned} \tag{2.53}$$

Again tedious but elementary algebra (like in the proof of 2.37) shows that the usual estimate gives the correct answer, i.e.

$$\begin{aligned} \widehat{V}(\widehat{Y}^*_{c,reg}) = &\left(1-\frac{n_2}{n_1}\right)\frac{1}{n_2}\frac{1}{n_2-1}\sum_{x\in s_2}\left(\frac{M(x)}{\bar{M}_2}\right)^2(R^*_c(x)-\widehat{R}^*_2)^2 \\ &+\frac{1}{n_1}\frac{1}{n_2-1}\sum_{x\in s_2}\left(\frac{M(x)}{\bar{M}_2}\right)^2(Y^*_c(x)-\widehat{Y}^*_2)^2 \end{aligned}$$

$$E_{1,2,3}\widehat{V}(\widehat{Y}^*_{c,reg}) = V(\widehat{Y}^*_{c,reg}) + O(n_2^{-2}) \tag{2.54}$$

If the model is unbiased and if the predictions and residuals are uncorrelated then, as in simple random sampling, it is possible to rewrite 2.53 as

$$\begin{aligned} V(\widehat{Y}^*_{c,reg}) \approx &\frac{1}{n_2}\frac{E_{x\in A}M^2(x)(Y_c(x)-\bar{Y})^2}{E^2_{x\in A}M(x)} \\ &-\frac{1}{n_2}\left(1-\frac{n_2}{n_1}\right)\frac{E_{x\in A}M^2(x)(\widehat{Y}_c(x)-\bar{Y})^2}{E^2_{x\in A}M(x)}+\frac{1}{n_2}\frac{E_{x\in F}V(x)}{E_{x\in A}M(x)} \end{aligned} \tag{2.55}$$

The above equations holds for the definitions of predictions and residuals as given in the next section on internal linear models, which outlines some aspects particularly useful in applications. Again, as in simple random sampling, one can consider asymptotically internal model as external models.

2.10 Internal linear models in two-phase sampling

At the point level the **theoretical predictions** are obtained via a design-based linear model (see [10]) of the form

$$\widehat{Y}_\beta(x) = \bar{Y} + \beta^t(Z(x)-\bar{Z}), \beta \in \Re^p, Z(x) \in \Re^p$$

At the cluster level we set consequently

$$\widehat{Y}_{\beta,c}(x) = \bar{Y} + \beta^t(Z_c(x)-\bar{Z}), \beta \in \Re^p, Z(x) \in \Re^p$$

These predictions are purely theoretical since $\bar{y}$, $\bar{Z}$ are unknown and β is, for the time being, arbitrary. These definitions ensure by 2.14 and 2.15 that

$$\frac{E_{x\in A}M(x)\widehat{Y}_{\beta,c}(x)}{E_{x\in A}M(x)} = \bar{Y}\ \forall\beta$$

The theoretical residuals are defined as

$$R_{\beta,c}(x) = Y_c(x) - \widehat{Y}_{\beta,c}(x)$$

By construction the theoretical residuals have zero mean, more precisely

$$\bar{R}_\beta = \frac{E_{x\in A} M(x) R_c(x)}{E_{x\in A} M(x)} = 0$$

We shall assume that the model has an intercept term, which is nearly always the case in practice. This means that one component of $Z(x)$, say the first $Z_1(x)$, is constant equal to 1. We therefore partition the vectors as $Z(x)^t = (1, Z^*(x)^t)$ and $\beta^t = (\beta_1, \beta^{*t})$. The optimal choice of the unknown parameter vector β is determined by minimizing the residual variance term, i.e.

$$\min_\beta E_x M^2(x)(Y_c(x) - \bar{Y} - \beta^t(Z_c(x) - \bar{Z}))^2$$

Differentiating with respect to β leads to the normal equations for the optimal theoretical choice β_0

$$E_x M^2(x)(Z_c(x) - \bar{Z})(Z_c(x) - \bar{Z})^t \beta_0 = E_x M^2(x)(Y_c(x) - \bar{Y})(Z_c(x) - \bar{Z})$$

It follows by simple algebra from the normal equations that the optimal theoretical residuals and predictions are uncorrelated in the sense that

$$E_{x\in A} M^2(x)(R_{\beta_0,c}(x) - \bar{R})(\widehat{Y}_{\beta_o,c}(x) - \bar{Y}) = 0$$

The (p,p) matrix on the left hand side is singular for models with an intercept term, since in this case the first row and the first column are identically zero. An elegant solution is to use generalized inverses, see [10], which unfortunately is not always a suitable procedure when using standard statistical software packages. We therefore give here an alternative approach. Rewriting the normal equations in terms of the reduced vectors $Z_c^*(x), \beta^*$ we see that the general solution of the original normal equations is given by:

$$\beta_0^t = (\beta_1, \beta_0^{*t})$$
$$E_x M^2(x)(Z_c^*(x) - \bar{Z^*})(Z_c^*(x) - \bar{Z^*})^t \beta_0^* = E_x M^2(x)(Y_c(x) - \bar{Y})(Z_c^*(x) - \bar{Z^*})$$

where β_1 is arbitrary.

In practice the theoretical normal equations are obviously not available and we solve instead their sample versions, i.e.

$$\sum_{x\in s_2} M^2(x)(Z_c^*(x) - \widehat{Z}_2^*)(Z_c^*(x) - \widehat{Z}_2^*)^t \widehat{\beta}_0^* = \sum_{x\in s_2} M^2(x)(Y_c(x) - \widehat{Y}_2)(Z_c^*(x) - \widehat{Z}_2^*)$$

In other words, $\widehat{\beta}_0^*$ is obtained by linear regression with weights $M^2(x)$ of the centered response variable $Y_c(x) - \widehat{Y}_2$ on the centered explanatory variables without the intercept term, i.e. on $Z_c^*(x) - \widehat{Z}_2^*$.
The empirical predictions are given by

$$\widehat{Y}_c(x) = \widehat{Y}_2 + \widehat{\beta}_0^{*t}(Z_c^*(x) - \widehat{Z}_2^*)$$

Most software packages will give directly the predictions, say $P_c(x)$, of $\widehat{Y}_c(x) - \widehat{Y}_2$, so that one can also write $\widehat{Y}_c(x) = \widehat{Y}_2 + P_c(x)$.
The empirical residuals are then given by

$$R_c(x) = Y_c(x) - \widehat{Y}_c(x) = (Y_c(x) - \widehat{Y}_2) - \widehat{\beta}^{*t}(Z_c^*(x) - \widehat{Z}_2^*)$$

which by construction satisfy

$$\widehat{R}_2 = \frac{\sum_{x\in s_2} M(x) R_c(x)}{\sum_{x\in s_2} M(x)} = 0$$

Hence, the point estimate is finally given by

$$\widehat{Y}_{c,reg} = \widehat{Y}_2 + \widehat{\beta}_0^{*t}(\widehat{Z}_1^* - \widehat{Z}_2^*) = \widehat{Y}_2 + \widehat{\beta}_0^t(\widehat{Z}_1 - \widehat{Z}_2)$$

where

$$\widehat{\beta}_0^t = (\beta_1, \widehat{\beta}_0^{*t})$$

and β_1 is arbitrary. Note that the first component of $\widehat{Z}_1 - \widehat{Z}_2$ is zero and that all statistically relevant quantities are independent of the arbitrary choice of β_1.
In short the optimal design-based point estimate and its estimated variance can be obtained by standard regression procedures.

It is also worthwhile to note that direct regression of the $Y_c(x)$ on the $Z_c(x)$ with weight $M^2(x)$ leads to residuals satisfying $\sum_{x\in s_2} M^2(x)R_c(x) = 0$ instead of $\sum_{x\in s_2} M(x)R_c(x) = 0$ which intuitively is more appealing. In simple cluster sampling $M(x) \equiv 1$, so that one could in this case also use ordinary least squares. Further modifications are possible by estimating the matrix

$$E_x M^2(x)(Z_c^*(x) - \bar{Z}^*)(Z_c^*(x) - \bar{Z}^*)^t$$

in the large sample, see [10] for details.

It is clear by 2.54 that all the previous results are also valid in two-phase two-stage sampling, it suffices to replace everywhere $Y_c(x)$ by $Y_c^*(x)$.
The reader can verify that the post-stratified estimate discussed in section 2.8 can be obtained by setting $Z_1(x) \equiv 1$, $Z_i(x) = 1$ if $x \in F_{i-1}$ and $Z_i(x) = 0$ if $x \notin F_{i-1}$ for $i = 1, 2 \ldots L$.

In the so called model-based approach, the parameter β is estimated by direct regression of $Y_c(x)$ on $Z_c(x)$ with weights $M(x)$ instead of $M^2(x)$. The resulting estimate $\widehat{\beta}_m$ leads to the model-based regression estimate

$$\widehat{Y}_{c,reg,m} = \widehat{Y}_2 + \widehat{\beta}_m^t(\widehat{Z}_1 - \widehat{Z}_2)$$

Its estimated variance can be approximated by the usual formula with the residuals defined as

$$R_c(x) = Y_c(x) - \widehat{\beta}_m^t Z_c(x)$$

The interested reader can consult [10] for rigorous results. Finally, one can also use standard unweighted least squares of the $Y(x_l)$ on the $Z(x_l)$, that is, by ignoring the cluster structure of the data, and then define in the usual way the predictions and residuals at the cluster level. As already mentioned there is some theoretical and empirical evidence that the various point estimates as well as their variance estimates will usually be close to each other.

In the next section we briefly mention how one could further generalize two-phase sampling schemes.

2.11 Generalized two-phase sampling schemes

Instead of selecting the points $x \in s_2 \subset s_1$ by equal probability sampling, one could choose these points independently of each other according to a probability function $p(x)$. We consider only two-phase one-stage simple random sampling. The adaptation to the other schemes is obvious.

The generalized two-phase one-stage estimate is defined as

$$\widehat{Y}_{g,reg} = \frac{1}{n_1}\left(\sum_{x\in s_1} \widehat{Y}(x) + \sum_{x\in s_2} \frac{R(x)}{p(x)}\right) \tag{2.56}$$

It is easily seen that, under the condition $\bar{R} = 0$

$$E_{x\in F}\widehat{Y}_{g,reg} = \bar{Y}$$
$$V(\widehat{Y}_{g,reg}) = \frac{1}{n_1}V_{x\in F}Y(x) + \frac{1}{n_1\lambda(F)}\int_F \frac{R^2(x)(1-p(x))}{p(x)}dx \tag{2.57}$$

It can be shown that in some sense generalized two-phase sampling generalizes standard two-phase sampling in much the same way as optimal stratification generalizes stratification with proportional allocation. It turns out that the gain in efficiency is usually small. The interested reader can find more details in [11].

2.12 Remarks on simulation procedures

Let us start with simple random sampling first. Drawing a random point uniformly distributed in F is not directly possible if F has a complicated shape. Instead one draws random points uniformly distributed in a convenient larger set $B \supset F$, say B rectangular, and check whether the points lie in F or not; there are efficient algorithms to do that when the boundary of F is given by a set of polygons. We have seen that given $x \in F$, x is uniformly distributed. Hence, to simulate n_2 points independently uniformly distributed in F we simulate a random number of points T until n_2 points are in F. T has a negative binomial distribution.

The idea is the same for cluster sampling but for some intricacies which, for completeness, we briefly outline. We know that we have to draw the origin x of the cluster in a domain A such that $A_l = A + e_l \supset F\ \forall l$; this can be difficult since A has usually a complicated shape which depends on the particular geometry of the cluster and of the forest (one would need, for the same forest, different maps for different cluster geometries, which is too costly). It is much easier to draw a random point in a set $B \supset A$ of simple shape (it is straightforward by looking at the forest map to find such a rectangular set B fulfilling the conditions for a wide range of cluster geometries), and simulate T points x uniformly in B until n_2 clusters with $M(x) > 0$ are obtained. In general the sets A and $G = \{x|M(x) > 0\}$ are not equal and $G + e_l \not\supset F\ \forall l$, so that caution is needed. n_2 is fixed and T is a random variable whose probability distribution function is the negative binomial, which means that:

$$P(T = t) = \binom{t-1}{t-n_2} p^{n_2}(1-p)^{t-n_2}$$

where $p = P(M(x) > 0) = \frac{\lambda(G)}{\lambda(B)}$. For this reason we shall call this sampling scheme "negative binomial sampling".

By taking the expectations in 2.16, 2.18 and 2.19 for given $T = t$ and after conditioning on the events $\{M(x) > 0\}$ and $\{M(x) = 0\}$ it can be checked that all formulae remain valid; this is straightforward for the point estimate since the $P(M(x) > 0)$ cancel out. For the variance note that formulae 2.18 is certainly valid in B given $T = t$ and that it can be rewritten as

$$V(\widehat{Y}_c|T = t) = \frac{1}{t}\frac{E_{x\in B}\{M^2(x)(Y_c(x) - \bar{Y})^2|M(x) > 0\}}{E^2_{x\in B}(M(x)|M(x) > 0)P(M(x) > 0)}$$

Now, $P(M(x) > 0)$ is unknown, but the maximum likelihood estimate of p is easily found to be $\widehat{p} = \frac{n_2}{t}$. It is also clear that $\bar{M}_2$ and

$$\frac{1}{n_2 - 1}\sum_{x\in s_2\cap G} M^2(x)(Y_c(x) - \widehat{Y}_c)^2$$

are unbiased estimates of $E_{x\in B}\{M(x)|M(x)>0\}$ and

$$E_{x\in B}\{M^2(x)(Y_c(x)-\bar{Y})^2|M(x)>0\}$$

respectively. Since $t\widehat{p}=n_2$ we see that 2.19 is an unbiased estimated of the variance. Hence, we can use all the formulae derived under the assumption "x uniform in A", which is more convenient from the mathematical point of view, also under the negative binomial sampling procedure, which is much easier to implement. Obviously one can use the same procedure in two-phase and two-stage sampling.

Simulating systematic grids with random origin and eventually also random orientation is straightforward. One retains only points falling into F and clusters hitting F, i.e. $M(x)>0$. Obviously n_2 is then always a random variable.

So far we have considered sampling schemes with a given design, more precisely for given inclusion probabilities π_i. We have not discussed the choice of the point estimates, nor the choice of the inclusion probabilities. We have already mentioned the fact that the pairwise inclusion probabilities π_{ij} depend on the relative position of the trees in the forest, which, loosely speaking, makes each forest in some sense unique, and render the task of optimization so difficult. In the next chapter we shall see that the Horwitz-Thompson point estimates we have proposed are not a bad choice after all, and that it is possible to get rid of the troublesome π_{ij} by considering the so called anticipated variance.

Chapter 3

Criteria for optimal strategies

3.1 General Background

Statistical inference for finite populations seemed to be an accomplished body of knowledge after the famous 1934 paper of Neyman [17], in much the same way as Newtonian mechanics: it was only a matter of time to work out the details for particular cases, and main stream mathematical statisticians preferred to concentrate their efforts on more challenging tasks. Doubts began to arise after the fundamental 1952 paper of Horwitz and Thompson [9] on variable probability sampling, in which the concept of linear estimation did not appear to be as straightforward as thought. The real foundations crisis was yet to come. In 1955 Godambe showed the non-existence of a uniformly least variance estimate [7], an astounding result, which led mathematical statisticians to investigate the intricacies of sampling theory and provide new approaches. The whole subject is rather subtle and, almost 50 years later, still controversial. We refer the interested reader to [25] for an overview and to [2] for a thorough exposition. However, due to the theoretical importance of the subject we briefly give the gist of the idea according to an enlightening argument of Basu [1].

Consider a finite population $\mathcal{P}$ of N individuals identified by their labels $u_i, i \in \{1, 2, \ldots, N\}$ and a real function (attribute) $Y : \mathcal{P} \mapsto \Re$, for convenience set $Y_i = Y(u_i)$, so that the function Y is completely specified by the vector $Y = (Y_1, Y_2, \ldots, Y_N)$. Let $\mathcal{Y}$ denote the set of all possible Y defined on $\mathcal{P}$ (for instance, imagine all the possible numerical quantities associated to the trees of a forest). We consider a given survey design p, i.e. a given probability function $p(s)$, $s \in S$, where S is the set of all subsets of $\mathcal{P}$. The Horwitz-Thompson estimator of the total $T_Y = \sum_{i=1}^{N} Y_i$ based on the sample s is defined as $\widehat{Y}(s) = \sum_{i \in s} \frac{Y_i}{\pi_i}$ (the inclusion probabilities π_i have been defined in chapter two). Let $\widehat{T}_Y(s)$ be any unbiased estimator of the total T_Y based on the sample s, that is $\sum_{s \in S} p(s) T_Y(s) = T_Y$ for all Y. Consider an arbitrarily chosen function Y_o say, with total T_{Y_o}. Let $\widehat{T}_{Y_o}(s)$ be the unbiased estimator of T_{Y_o}. The following new estimator is obviously also unbiased

$$\widehat{T}^*_Y(s) = \widehat{T}_Y(s) + T_{Y_o} - \widehat{T}_{Y_o}(s)$$

Now when $Y = Y_o$, that is when the attribute being sampled and the arbitrarily chosen attribute are identical, then

$$\widehat{T}^*_Y(s) \equiv T_{Y_o}$$

and $\widehat{T}_Y^*(s)$ has zero variance since it is constant. Hence, for an estimator to be uniformly best for all $Y \in \mathcal{Y}$ it must have zero variance everywhere since Y_o was chosen arbitrarily. This is impossible (unless the sampling design is a full census), so that no uniformly best unbiased estimator exists.

The problem of a lack of best estimators arises because of the generality of Neyman's design-based formulation of the solution to the inference problem. Inferences are made with respect to the sampling distribution for any function Y, regardless of its structure on the population $\mathcal{P}$. For instance, Y_i may be the weight of individuals belonging to a population consisting of mice, elephants and trees. But this is too much freedom for a satisfactory theory of inference and no optimum properties can be found for all functions defined on a population.

One important consequence of this non-existence theorem is that no empirical comparison can ever be conclusive, since in any particular case somebody may be able to construct a better estimator.

If the very best is not available in this world, we may still have hope to find estimators which are not that bad after all. Fortunately, Godambe and Yoshi showed in 1965, see [8], that the Horwitz-Thompson estimator, that is 2.4 in the forestry context, is admissible in the class of all unbiased estimates, meaning that no other estimator in this class has a smaller variance for all functions Y (and strictly smaller for at least one Y). Furthermore, the Horwitz-Thompson estimator remains admissible in some restricted class of functions, for instance like binary attributes $Y(u) \in \{0, 1\}$. Note also, that the non-existence of a uniformly least variance unbiased estimator follows from the admissibility of the Horwitz-Thompson estimator. It can also be shown that the Horwitz-Thompson estimator is essentially the only hyper-admissible estimator in the class of all unbiased estimators, which means that it is admissible for all possible subpopulations (for each Y, a subpopulation is defined by setting some of Y_i to zero, which can be done in 2^N ways); however, this concept is rather controversial (again see [2] for details).

The previous arguments are taken from standard sampling theory and ignore completely the intrinsic geometrical structure of forest sampling. For instance, it would be interesting to known whether the Horwitz-Thompson estimator is admissible when instead of considering a set of attributes we consider a set of locations of the trees. By sampling in $A \supset F$ we can eliminate boundary effects and define a sampling scheme with given inclusion probabilities π_i; however the π_{ij} are not under control and different locations thus yield different sampling schemes in the classical sense (i.e. the set of all possible subsets of trees which can be sampled from a point x will depend on the particular location of the trees, even if the random point x is always uniformly distributed in F).

Let us consider the same forest in two different locations within the same area F, that is the vector Y is the same, but the corresponding trees have locations $u_i^{(k)}, k = 1, 2$ for the spatial pattern (1) and (2). As we have seen in chapter two this generates two different sampling schemes p_1 and p_2 with inclusion probabilities $\pi_i^{(k)}, \pi_{ij}^{(k)}$ (the two sets of samples $s \subset \mathcal{P}$ with $p_1(s) \neq 0, p_2(s) \neq 0$ are different). We consider the Horwitz-Thompson estimator $Y(x)$ under the second pattern and an arbitrary unbiased estimator $T_Y(x)$ under the first pattern. The following lemma is important (for a proof see lemma 3.2 p. 60 of [2]):
if

$$\sum_{i=1}^{N} \pi_i^{(1)} = \sum_{i=1}^{N} \pi_i^{(2)}$$

and

$$V_{p_1}(T_Y(x)) \leq V_{p_2}(Y(x)) \; \forall Y \in \Re^N$$

then

$$\pi_i^{(1)} = \pi_i^{(2)}, \pi_{ij}^{(1)} = \pi_{ij}^{(2)}$$

Hence, if we can find an unbiased estimator for pattern (1) which has a smaller variance for all functions Y than the Horwitz-Thompson estimator under pattern (2) and under the natural condition that the expected number of trees drawn is the same, then it follows that the inclusion probabilities are the same (in practice we would have kept the π_i constant anyway). From this one derives (see [2] p. 62) the admissibility of any Horwitz-Thompson based strategy $(p, Y(x))$, in the sense that it is impossible to find a sampling scheme (i.e. pattern in the the present context) p^* and an unbiased estimator $T_Y(x)$ such that $V_{p^*}(T_Y(x)) \leq V_p(Y(x))\ \forall Y$ with strict equality for at least one Y. When sampling is performed by using arbitrary circles it is straightforward to give a geometrical interpretation of the above lemma, namely that two pattern having the same π_{ij} are essentially the same with respect to the spatial structure of the forest. Trees isolated in one pattern, i.e. i is isolated if $\pi_{ij} = 0, j \neq i$, are isolated in the second, trees forming a chain in one pattern, e.g. $\pi_{12}\pi_{23}\pi_{34} \neq 0$, form a chain of the same length in the second pattern, and, last not least, trees forming a triangle in one pattern, e.g. $\pi_{12}\pi_{13}\pi_{23} \neq 0$ form a congruent triangle in the second pattern. With some geometric insight it is clear that the two forests have essentially the same spatial structure up to local or global translations and rotations.

We can summarize the above findings by saying that if we know nothing about the response functions of potential interest and nothing about the spatial structure of the forest, then using the Horwitz-Thompson estimator is not the worst thing to do. From a practical point of view this result is of limited use, but this is essentially all what can be said within the Neyman's framework. Another, and probably the most important reason for using the Horwitz-Thompson estimate in forest inventory is its simplicity and clear physical interpretation as a local density. Beside, it is extremely difficult to exhibit other estimators without using auxiliary information. It is well known in standard sampling theory that , in practice, the Horwitz-Thompson estimators performs well if the π_i are roughly proportional to the Y_i, and that it can perform poorly otherwise. As we shall see, the optimal forest sampling schemes belong to the first category.

Out of the many approaches developed in standard sampling (nearly always within the class of non-informative designs, see section 2.2) to obtain meaningful optimality criteria, the super-population model has received much attention. The basic idea is to consider the vector Y as the realization of a stochastic process (usually assuming a linear model with error structure known up to a scale constant) and to calculate the expected value, under the super-population model, of the design-based variance of the estimator, which is called **the anticipated variance**; see [25, 2] for the general theory and [13] for some straightforward applications to forest inventory. According to this approach one would consider a given forest as the realization of a complex stochastic process in the plane generating the locations u_i of the trees and the values Y_i of the variable of interest. The adequate mathematical tool for such a modelization is the so called theory of marked point process, see e.g. [4] for a excellent overview and Penttinen [16] for an application to forestry. The drawbacks of this sophisticated approach are at least threefold: first the mathematical complexity, second the difficult task of model fitting (which requires extensive data generally not available without thorough pilot studies), third it addresses more the modelization of forest structures than the down to earth problem of forest inventory, i.e. the estimation of some simple quantities for a given forest. Our objective being to provide simple and yet realistic guidelines the super-population model retained is very simple. The forest is partitioned into strata within which the locations u_i of the trees are assumed to be independently uniformly distributed

while, in contrast to standard sampling theory, the Y_i are kept constant. We shall call this model **the local Poisson forest**. This, of course, is a crude approximation of any real forest, since concurrence mechanisms are not taken into account, and is certainly wrong in special cases like regular plantations (though thinning procedures may render a regular network to look fairly random after some time) or forest consisting of more or less isolated clusters of trees. However, this is certainly a good starting point for approximation. Another justification is that this simple model yields interesting results under systematic sampling (see section 3.5). A further justification, as we shall see, is that everything can be easily calculated and the resulting optimal schemes have a clear intuitive appeal.

The first step of our program is therefore to calculate the anticipated variance of the Horwitz-Thompson estimator under the model of a local Poisson forest. The second step is to determine the optimal choice of inclusion probabilities by looking at the resulting anticipated variance as well as the resulting costs. However, before doing that, we have to give rigorous definitions and analyse some technical difficulties.

3.2 Definition of the anticipated variance

We first consider the case of the **global Poisson Forest**, that is, we assume that the coordinates $u_i, i = 1, \ldots, N$ are uniformly and independently distributed in the forest area F; we shall write $u_i(\omega)$ to emphasize the fact that the coordinates depend on the random event ω, which generates the uniform random pattern of the forest. This terminology is due to the fact that if the trees in F are generated by a spatially homogenous Poisson process then, conditionally on the number of trees, the locations of the trees are independently and uniformly distributed in F. The indicator functions 2.2 now depend also on ω, which we write as $I_i(x, \omega)$. The circle centered at the origin with radius r_i is denoted by K_i, whereas the same circle centered at x is denoted by $K_i(x)$. Thus, the circle associated with the i-th tree under the random location ω is denoted by $K_i(u_i(\omega))$. We need the following notation

$$\pi_i(\omega) = E_{x|\omega} I_i(x, \omega),\ \pi_{ij}(\omega) = E_{x|\omega} I_i(x, \omega) I_j(x, \omega),\ \pi_i(x) = E_{\omega|x} I_i(x, \omega)$$

Since $u_i(\omega)$ and x are uniformly distributed in F we have

$$E_\omega \pi_i(\omega) = E_x \pi_i(x) = \tilde{\pi}_i$$

For a given pattern ω, the Horwitz-Thompson estimator is then

$$Y(x, \omega) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{Y_i}{\pi_i(\omega)} I_i(x, \omega)$$

Clearly, $Y(x, \omega)$ is design unbiased for each pattern ω.
The overall variance can be obtained via 2.23 and is given by

$$V_{x,\omega} Y(x, \omega) = E_\omega V_{x|\omega} Y(x, \omega) + V_\omega E_{x|\omega} Y(x, \omega) = E_\omega V_{x|\omega} Y(x, \omega)$$

The term $E_\omega V_{x|\omega} Y(x, \omega)$ is called the **anticipated variance**; it is the average variance, under all uniform random spatial pattern, of the design-based variance. Note that keeping the Y_i constant allows one to interchange the order of expectations $E_{x|\omega}, E_{\omega|x}$ even when the design is informative (as we shall see the exact optimal designs of forest inventory are informative at the tree level).

According to 2.9, we need to calculate $E_\omega \pi_i(\omega)$ and $E_\omega \pi_{ij}(\omega)$ to obtain a first order approximation of the anticipated variance.

Let us now consider the **local Poisson model**. We assume that the forest F can be partitioned in L strata $F_k, k = 1, \ldots, L$, i.e. $F = \bigcup_{k=1}^{L} F_k$, $\forall k \neq l, F_k \bigcap F_l = \emptyset$, and that the locations vectors $u_i(\omega)$ are independently uniformly distributed in F_k for trees belonging to the k-th stratum. This model is a good approximation for instance in the presence of a clear stand structure according to development stages, in contrast to selection forests or virgin forests, for which the global Poisson model appears to be more appropriate. Note that we do not assume the strata to be simply connected sets (i.e. a stratum can be spread over several domains of the plane). We need the following notation

$$\bar{Y}_k = \frac{1}{\lambda(F_k)} \sum_{i \in F_k} Y_i$$

$$p_k = \frac{\lambda(F_k)}{\lambda(F)}$$

To obtain unbiased estimates for each stratum and to simplify the mathematics we redefine, for a given pattern ω, the indicator variables as

$$I_i(x, \omega) = \begin{cases} 1 \text{ if } x \in K_i(u_i(\omega)) \cap F_k \text{ and } i \in F_k \\ 0 \text{ otherwise} \end{cases} \tag{3.1}$$

That is, when the point x falls in F_k **only the trees of the same k-th stratum are included**. The conditional inclusion are defined as

$$\pi_i^{(k)}(\omega) = \begin{cases} \frac{\lambda(K_i(u_i(\omega)) \cap F_k)}{\lambda(F_k)} \text{ if } i \in F_k \text{ and } x \in F_k \\ 0 \text{ otherwise} \end{cases} \tag{3.2}$$

likewise for the pairwise inclusion probabilities

$$\pi_{ij}^{(k)}(\omega) = \begin{cases} \frac{\lambda(K_i(u_i(\omega)) \cap K_j(u_j(\omega)) \cap F_k)}{\lambda(F_k)} \text{ if } i \in F_k, j \in F_k \text{ and } x \in F_k \\ 0 \text{ otherwise} \end{cases} \tag{3.3}$$

The unconditional inclusion probabilities $\pi_i(\omega), \pi_{ij}(\omega)$ satisfy by definition the relations

$$\begin{aligned} \pi_i(\omega) &= \pi_i^{(k)}(\omega) p_k \\ \pi_{ij}(\omega) &= \pi_{ij}^{(k)}(\omega) p_k \end{aligned} \tag{3.4}$$

The **revised Horwitz-Thompson estimator** is defined as usual as

$$Y(x, \omega) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{Y_i}{\pi_i(\omega)} I_i(x, \omega) \tag{3.5}$$

By construction the revised estimator has **the important property that if $x \in F_k$ then**

$$Y(x, \omega) = \frac{1}{\lambda(F_k)} \sum_{i \in F_k} \frac{Y_i}{\pi_i^{(k)}} I_i(x, \omega)$$

which implies at once $E_x\{Y(x, \omega) | x \in F_k\} = \bar{Y}_k$. The revised estimator yields therefore unbiased estimates for each stratum F_k. The revised estimate is also unbiased for F since

$$E_{x|\omega} Y(x) = \sum_{j=1}^{L} E_{x|\omega}(Y(x, \omega) | x \in F_j) P(x \in F_j) = \sum_{j=1}^{L} p_j \bar{Y}_j = \bar{Y}$$

Remarks

- The adjective "revised" is somewhat superfluous since 3.4 is by definition an Horwitz-Thompson estimate. The sampling scheme has been "revised" to allow for unbiased estimates in each stratum, which can be a practical requirement depending on the context, and to simplify the mathematics.

- The inclusion probabilities of the standard and revised sampling schemes are the same for interior trees, they differ only at strata boundaries. It is clear that the revised procedure leads to a smaller number of sampled trees per point.

- The revised scheme simplifies the mathematics of all procedures under the local Poisson model. Both schemes are the same under the global Poisson model. In practice, one would use the revised estimate only in two-phase sampling.

- Since we are primarily interested in the local Poisson model we shall use from now on the revised estimate, with the same notation as for the standard estimate.

To calculate the anticipated variance let us note first that the true means $\bar{Y}_k, \bar{Y}$ do not depend on ω and that we can decompose the design-based variance as

$$\begin{aligned} V_{x|\omega}Y(x,\omega) &= \frac{1}{\lambda(F)}\int_F (Y(x,\omega)-\bar{Y})^2 dx = \frac{1}{\lambda(F)}\sum_{j=1}^{L}\int_{F_j}(Y(x)-\bar{Y}_j+\bar{Y}_j-\bar{Y})^2 dx \\ &= \sum_{j=1}^{L} p_j V_{x|\omega}(Y(x,\omega)|x\in F_j) + \sum_{j=1}^{L} p_j(\bar{Y}_j-\bar{Y})^2 \end{aligned} \tag{3.6}$$

Hence, we can calculate the anticipated variance under the local Poisson model if we can do it under the global Poisson model which applies within each stratum. To this end, we have to analyse the conditions under which the impact of boundary effects can be neglected. The main idea is very simple: the troublesome terms of the variance are the $\pi_{ij}(\omega)$. By taking the anticipated variance the expectation $E_\omega \pi_{ij}(\omega)$ can be rewritten first as $E_\omega E_{x|\omega} I_i(x,\omega) I_j(x,\omega)$, then, by changing the order of expectations, as $E_x E_{\omega|x} I_i(x,\omega) I_j(x,\omega)$ and therefore, since the trees are independently distributed in F, also as $E_x \pi_i(x)\pi_j(x) \approx \tilde{\pi}_i \tilde{\pi}_j$ if one neglects the boundary effects. Hence, the anticipated variance will depend only on the individual inclusion probabilities and no longer on the pairwise inclusion probabilities, i.e. on the spatial structure of the forest. Intuitively speaking, this is legitimate if the inclusion circles are small with respect to the forest area. **Let us emphasize the fact that this is not the classical problem of adjustment at the forest edge**, for instance by the well-known reflection method of Schmid-Haas or other methods (see [23]). In this work we have assumed from the very beginning that one has the correct inclusion probabilities, i.e. one knows precisely the $\lambda(F \cap K_i(u_i(\omega)))$ and the $\lambda(F \cap K_i(u_i(\omega)) \cap K_j(u_j(\omega)))$ for all trees included in the sample. This, of course, assumes that the forest boundary and the location of the trees are recorded, which is becoming a standard procedure with the on-growing use of geographical information systems. The next section gives a modern and qualitatively rigorous mathematical treatment for the impact of boundary effects on the anticipated variance. It is clear that exact and universal results (i.e. valid for any forest, no matter how complicated its shape is) are illusory.

3.3 The mathematics of boundary effects

We have seen that it suffices to consider the case of an **homogenous Poisson forest**, that is to assume that the coordinates $u_i(\omega), i = 1, \ldots, N$ are uniformly and independently distributed in the forest area F; The **dilatation** of the forest area F by the circle K_i is the Minkowski sum

$$F \oplus K_i = \{x | K_i(x) \cap F \neq \emptyset\} = \bigcup_{x \in F} K_i(x) \tag{3.7}$$

whereas the **erosion** of the forest area F by the circle K_i is defined as

$$F \ominus K_i = \{x | K_i(x) \subset F\} \tag{3.8}$$

Dilatation and erosion play a key role in integral geometry and mathematical morphology, see for instance [24]. One obviously has the inclusions

$$F \ominus K_i \subset F \subset F \oplus K_i$$

We shall see that boundary effects are negligible if the sets $F \ominus K_i$ and $F \oplus K_i$ are close to the set F for all circles K_i. We need a simple but important result of integral geometry, which states that

$$\int_{\Re^2} \lambda(F \cap K_i(x))dx = \int_{F \oplus K_i} \lambda(F \cap K_i(x))dx = \lambda(F)\lambda(K_i) \tag{3.9}$$

Proof:

$$\int_{\Re^2} \lambda(F \cap K_i(x))dx = \int_{\Re^2} dx \int_{\Re^2} I_F(y) I_{K_i(x)}(y)dy = \int_{\Re^2} dx \int_{\Re^2} I_F(y) I_{K_i(y)}(x)dy$$

which is equal to

$$\int_{\Re^2} dy I_F(y) \int_{\Re^2} I_{K_i(y)}(x)dx = \lambda(K_i) \int_{\Re^2} I_F(y)dy = \lambda(F)\lambda(K_i)$$

This result can be found e.g. in [15] under the condition that F is convex. This, however, is not necessary since the above proof requires only Fubinis's theorem to interchange the order of integration, and the obvious fact that x is in $K_i(y)$ if and only if y is in $K_i(x)$. The first equality in 3.8 follows from the definition of the dilatation.
Using the above result twice by beginning with $F \cap K_i(x)$ we get at once

$$\begin{aligned} \int_{\Re^2 \times \Re^2} \lambda(F \cap K_i(x) \cap K_j(y))dxdy &= \int_{F \oplus K_i \times F \oplus K_j} \lambda(F \cap K_i(x) \cap K_j(y))dxdy \\ &= \lambda(F)\lambda(K_i)\lambda(K_j) \end{aligned} \tag{3.10}$$

To calculate the average inclusion probabilities $\tilde{\pi}_i$ we shall assume that the forest boundary ∂F is given by a set of polygons. We consider now a point $x \in F$ near the boundary and we assume that $K_i(x) \cap \partial F \neq \emptyset$ **in one edge only**. These conditions are fulfilled if the boundary consists of fairly long edges and not to many vertices (corners), narrow forest strips are also excluded. Let us denote by $\xi = \xi(x)$ the distance of the point $x \in F \setminus F \ominus K_i$ to the nearest edge. Then one can use the well known formula

$$\lambda(F \cap K_i(x)) = \pi r_i^2 - \left(r_i^2 \arccos(\frac{\xi}{r_i}) - \xi\sqrt{r_i^2 - \xi^2} \right)$$

Straightforward integration yields then

$$\frac{1}{r_i}\int_o^{r_i} \lambda(F \cap K_i(x))d\xi = (1 - \frac{2}{3\pi})\lambda(K_i)$$

Hence, up to boundary effects in the vertices, we have

$$\gamma_{1,i} = \frac{1}{\lambda(K_i)} E_{x \in F \setminus (F \ominus K_i)} \lambda(F \cap K_i(x)) \approx \left(1 - \frac{2}{3\pi}\right)$$

With the same arguments one obtains

$$\gamma_{2,i} = \frac{1}{\lambda(K_i)} E_{x \in (F \oplus K_i) \setminus F} \lambda(F \cap K_i(x)) \approx \frac{2}{3\pi}$$

Up to boundary effects at the vertices of the boundary polygons we have

$$\frac{\gamma_{1,i}}{\gamma_{2,i}} = \frac{3\pi}{2} - 1 \tag{3.11}$$

According to 3.8 we can write

$$E_{x \in F \oplus K_i} \lambda(F \cap K_i(x)) = \lambda(K_i) \frac{\lambda(F)}{\lambda(F \oplus K_i)}$$

which is equal to

$$\lambda(K_i) \left(\frac{\lambda(F \ominus K_i)}{\lambda(F \oplus K_i)} + \frac{\lambda(F \setminus (F \ominus K_i))}{\lambda(F \oplus K_i)} \gamma_{1,i} + \frac{\lambda((F \oplus K_i) \setminus F)}{\lambda(F \oplus K_i)} \gamma_{2,i} \right)$$

Using 3.10 we obtain

$$\gamma_{1,i} = \frac{1 - \frac{2}{3\pi}}{1 + \frac{2}{3\pi}(\alpha_i - 1)}$$

where we have set

$$\alpha_i = \frac{\lambda(F \oplus K_i) - \lambda(F)}{\lambda(F) - \lambda(F \ominus K_i)}$$

For non-pathological forest shapes α_i will usually be greater than 1 and will tend to 1 if the surface area $\lambda(F)$ tends to infinity while the circle K_i remains bounded. Of course, things can go wrong with fractal-like forest, for which it is quite possible that any circle $K_i(x)$ will always intersect the forest and the non-forest area. Therefore, we can say that $\gamma_{1,i}$ will usually be smaller than $1 - \frac{2}{3\pi}$ but very close to it. For completeness we give some results pertaining to the dilated and eroded forest areas.

First, when F is convex the famous Steiner's formula (see [19]) states that

$$\lambda(F \oplus K_i) = \lambda(F) + L(F)r_i + \pi r_i^2$$

where $L(F)$ is the length of the boundary of F (i.e. the perimeter). More generally, for non-convex F satisfying the following regularity conditions: $\exists \tilde{F}\ F = \tilde{F} \oplus K, K = \bigcup_i K_i$ (K is simply the largest circle), and F is closed and open with respect to K, i.e. $F = (F \ominus K) \oplus K = (F \oplus K) \ominus K$, then one has the following generalized Steiner's formulae (see [24] p.143, where the term $2L(F)r$ is a misprint and should read $L(F)r$)

$$\lambda(F \oplus K_i) = \lambda(F) + L(F)r_i + \pi r_i^2 \chi(F)$$

$$\lambda(F \ominus K_i) = \lambda(F) - L(F)r_i + \pi r_i^2 \chi(F)$$

where $\chi(F)$ is the Euler-Poincaré characteristic of the forest area F, which is defined as the number of components minus the number of holes; for instance $\chi(F)$ is equal

to 1 when F is convex, and $\chi(F)$ is 0 when F is a circular ring. The technical condition requiring F to be open and close with respect to K is violated at the vertices of a polygons. Corners have to be smoothed. The reader can verify that the above Steiner's formulae hold exactly when F is a large circle or a large circular ring; as a counter-example, consider a square for which it is easy to see that πr_i^2 must be replaced by $4r_i^2$ in the second formula. In any case, the above formulae show that for well behaved shapes of F, the coefficient α_i is very close to 1 (exactly 1 if the Steiner's formulae hold).
Summarizing the previous results we can state that

$$E_{x\in F}\lambda(F\cap K_i(x)) = \frac{\lambda(F\ominus K_i)}{\lambda(F)}\lambda(K_i) + \frac{\lambda(F)-\lambda(F\ominus K_i)}{\lambda(F)}\gamma_{1,i}\lambda(K_i)$$

which is exactly equal to

$$\lambda(K_i)\,(1-(1-\gamma_{1,i})\epsilon_i)$$

where we have set $\epsilon_i = \frac{\lambda(F)-\lambda(F\ominus K_i)}{\lambda(F)}$. By the Steiner's formulae ϵ_i is of order $r_i\frac{L(F)}{\lambda(F)}$. Neglecting the boundary effects at the vertices we can use the previous result on straight line boundaries to get

$$\lambda(K_i)\left(1-\frac{2}{3\pi}\epsilon_i\right)$$

and we can state that under regularity conditions the average inclusion probabilities satisfy the following relations

$$\frac{\lambda(K_i)}{\lambda(F\oplus K_i)} \leq \tilde{\pi}_i \approx \frac{\lambda(K_i)}{\lambda(F)}\left(1-\frac{2}{3\pi}\epsilon_i\right) \leq \frac{\lambda(K_i)}{\lambda(F)} \tag{3.12}$$

where $\epsilon_i \to 0$ as $\frac{\lambda(K_i)}{\lambda(F)} \to 0$. Note that the first inequality is intuitively obvious (increasing the boundary area by dilatation will decrease the average intersection area $K_i(x)\cap F$), though it may be violated in pathological situations. Furthermore, one also has

$$\pi_i(\omega) \equiv \frac{\lambda(K_i)}{\lambda(F)} \;\; \forall u_i(\omega) \in F\ominus K_i$$

$$\pi_i(x) \equiv \frac{\lambda(K_i)}{\lambda(F)} \;\; \forall x \in F\ominus K_i$$

Boundary effects are consequently negligible when

$$\frac{\lambda(F\ominus K_i)}{\lambda(F)} \to 1$$

and the boundary is regular enough.

We now investigate the pairwise inclusion probabilities $\pi_{ij}(\omega)$. We have

$$E_\omega\,\pi_{ij}(\omega) = E_\omega\frac{1}{\lambda(F)}\lambda(F\cap K_i(u_i(\omega))\cap K_j(u_j(\omega)))$$

which is equal to

$$\frac{1}{\lambda(F)}\frac{1}{\lambda^2(F)}\int_{F\times F}\lambda(F\cap K_i(x)\cap K_j(y))dxdy$$

According to the previous arguments the last expression will, for regular F, be larger than

$$\frac{1}{\lambda(F)}\frac{1}{\lambda(F\oplus K_i)}\frac{1}{\lambda(K\oplus K_j)}\int_{F\oplus K_i\times F\oplus K_j}\lambda(F\cap K_i(x)\cap K_j(y))dxdy$$

since sampling outside F will usually decrease the surface area of the intersections. By 3.9 we get

$$E_\omega \pi_{ij}(\omega) \geq \frac{\lambda(K_i)}{\lambda(F \oplus K_i)} \frac{\lambda(K_j)}{\lambda(F \oplus K_j)}$$

On the other hand, since the trees are independently distributed, we also have

$$E_\omega \pi_{ij}(\omega) = E_x E_{\omega|x} I_i(x,\omega) I_j(x,\omega) = E_x \frac{\lambda(F \cap K_i(x))}{\lambda(F)} \frac{\lambda(F \cap K_j(x))}{\lambda(F)}$$

Hence we have

$$\frac{\lambda(K_i)}{\lambda(F \oplus K_i)} \frac{\lambda(K_j)}{\lambda(F \oplus K_j)} \leq E_\omega \pi_{ij}(\omega) \leq \frac{\lambda(K_i)}{\lambda(F)} \frac{\lambda(K_j)}{\lambda(F)}$$

so that under the same regularity conditions as for 3.11 we can state that

$$E_\omega \pi_{ij}(\omega) \approx \tilde{\pi}_i \tilde{\pi}_j \tag{3.13}$$

Since for non-pathological F the random variables $\lambda(F \cap K_i(x))$ and $\lambda(F \cap K_j(x))$ will be positively correlated we usually will have convergence from above, i.e.

$$E_\omega \pi_{ij}(\omega) \downarrow \tilde{\pi}_i \tilde{\pi}_j$$

Taking the expectation of $V_{x|\omega} Y(x,\omega)$ with respect to ω we see that, to the first order (an exact analytical treatment is impossible since we have to calculate the expectations of ratios of random variables in ω), all the cross-products terms vanish and we obtain the following result:
Under the regularity conditions ensuring that boundary effects are negligible, the first order approximation of the anticipated variance of the Horwitz-Thompson estimate under the global Poisson forest is given by

$$E_\omega V_{x|\omega} Y(x,\omega) = \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\tilde{\pi}_i} - \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} Y_i^2 \tag{3.14}$$

The first term, say $\alpha(F)$ with $0 < \alpha(F) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\lambda(K_i)} < \infty$, remains bounded as $\lambda(F) \to \infty$, whereas the second term can be written as $\frac{1}{\lambda(F)} \beta(F)$, with $\beta(F) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\lambda(K_i)}$. This second term will tend to zero for $\lambda(F) \to \infty$, since $0 < \beta(F) < \infty$. Hence, for large areas one can write

$$E_\omega V_{x|\omega} Y(x,\omega) \approx \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\tilde{\pi}_i} \tag{3.15}$$

This is the version we shall use for optimization since the neglected term does not depend on the inclusion probabilities (which are the variables to be optimized) and because we are primarily interested in forest inventories performed on large areas.

Another, perhaps mathematically more elegant, justification to use 3.13 is given by considering the pseudo Horwitz-Thompson estimator

$$\tilde{Y}(x,\omega) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{Y_i}{\tilde{\pi}_i}$$

which is model-design and design-model unbiased, i.e.

$$E_\omega E_{x|\omega} \tilde{Y}(x,\omega) = E_x E_{\omega|x} \tilde{Y}(x,\omega) = \bar{Y}$$

though it is neither exactly design nor model unbiased. It is easily verified that the overall variance of $\tilde{Y}(x,\omega)$, i.e $V_{x,\omega}\tilde{Y}(x,\omega)$, is precisely given by 3.13. The overall variance is equal to the anticipated variance for design-unbiased estimator and is the natural generalization for estimators which are only model-design unbiased. It is intuitively clear that $\tilde{Y}(x,\omega)$ is very close to $Y(x,\omega)$ when boundary effects are negligible. Yet another, more pragmatic reason, to use 3.13 is that it contains only the terms of the variance which depend solely on the individual inclusion probabilities, up to boundary effects independent of the locations of the trees, and not on the pairwise inclusion probabilities, which depend on the forest structure and which we cannot control.

For future use we now investigate the model bias at the boundary. We have

$$E_{\omega|x}Y(x,\omega) = \frac{1}{\lambda(F)}\sum_{i=1}^{N} Y_i E_{\omega|x}\frac{I_i(x,\omega)}{\pi_i(\omega)} = \frac{1}{\lambda(F)}\sum_{i=1}^{N} Y_i\psi_i(x)$$

where we have set

$$\psi_i(x) = \int_{F\cap K_i(x)} \frac{du}{\lambda(F\cap K_i(u))}$$

Clearly, if $K_i(u) \subset F\ \forall u \in F\cap K_i(x)$ then $\psi_i(x) = 1$, which is the case whenever $x \in F \ominus (K_i \oplus K_i)$. Furthermore, since $E_{x|\omega}Y(x,\omega) = \bar{Y}$ we have by interchanging the order of expectation $E_x\psi_i(x) = 1$. As a first approximation we can replace $\lambda(F\cap K_i(u))$ by $\lambda(F\cap K_i(x))$ to obtain $\psi_i(x) \approx 1$. To get further insight we consider an infinite straight line boundary. Let $\xi = \xi(x)$ denote the distance of the point $x \in F$ to the boundary line, likewise $\eta = \eta(u)$ will denote the distance of the point $u \in F\cap K_i(x)$. One has

$$\lambda(F\cap K_i(u)) = \lambda(K_i)\left(1 - \frac{1}{\pi}\arccos(\frac{\eta}{r_i}) + \frac{\eta}{\pi r_i}\sqrt{1-\left(\frac{\eta}{r_i}\right)^2}\right)$$

Note that $\xi \in [0, 2r_i]$ and $\eta \in [\xi - r_i, \xi + r_i]\cap[0,\infty]$. Since for given η the length of the chord in $K_i(u)$ parallel to the boundary is given by $2\sqrt{r_i^2-(\xi-\eta)^2}$, we have

$$\psi_i(x) = \int_{\max(0,\xi-r_i)}^{r_i+\xi} \frac{2\sqrt{r_i^2-(\xi-\eta)^2}d\eta}{\pi r_i^2\left(1-\frac{1}{\pi}\arccos(\frac{\eta}{r_i})+\frac{\eta}{\pi r_i}\sqrt{1-\left(\frac{\eta}{r_i}\right)^2}\right)}$$

Setting $\theta(x) = \frac{\xi(x)}{r_i} \in [0,2]$ and changing the variable to $w = \frac{\eta}{r_i}$ we obtain

$$\psi_i(x) = \frac{2}{\pi}\int_{\max(0,\theta(x)-1)}^{1+\theta(x)} \frac{\sqrt{1-(\theta(x)-w)^2}dw}{\phi(w)}$$

where the function $\phi(w)$ is defined as

$$\phi(w) = \begin{cases} 1 - \frac{1}{\pi}\arccos(w) + \frac{w}{\pi}\sqrt{1-w^2} & \text{if } \le 1 \\ 1 \quad \text{if } w > 1 \end{cases}$$

Numerical integration shows that $\min_{x\in\partial F}\psi_i(x) = 0.69$ at $\theta(x) = 0$ and that $\max_{x\in\partial F}\psi_i(x) = 1.13$ at $\theta(x) = 1$. However, the average of $\psi_i(x)$ in the boundary strip $\theta \in [0,2]$ is $E_{x\in\partial F}\psi_i(x) = 0.9964$, whereas it is 1.0146 in the boundary strip $[0.1, 1.9]$. Consequently, we can say that the average model bias in the boundary zone is negligible and we shall therefore write

$$E_{\omega|x}Y(x,\omega) \approx \bar{Y} \quad \forall x \in F \tag{3.16}$$

We are now ready to calculate the anticipated variance under the sampling schemes defined in chapter 2. To simplify the notation we shall write π_i instead of $\tilde{\pi}_i$, since we assume that boundary effects are negligible.

3.4 The anticipated variances

Let us first recall in a qualitative way the validity assumptions under the local Poisson model derived in the previous section : **The strata are large in comparison with the inclusion circles, the strata and forest boundaries are defined by polygons with edges much larger than the diameters of the inclusion circles and the number of vertices (corners) is not too large. Otherwise, the structure of the forest can be fairly complex, i.e. consisting of several components, convex or not, simply connected or not, i.e. zones of non-forest area can lie within the strata ("holes"). Small strata components with very jagged boundaries, i.e. fractal-like forests will violate these assumptions**.

3.4.1 One-phase one-stage simple random sampling

We consider the **local Poisson model**. According to 3.5 the variance of the revised Horwitz-Thompson estimator is given by

$$V_{x|\omega}Y(x,\omega) = \sum_{j=1}^{L} p_j V_{x|\omega}(Y(x,\omega)|x \in F_k) + \sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2$$

Using 3.13 within each strata and the definitions of the inclusion probabilities 3.2 and 3.3 we get after some simple algebra:
The anticipated variance of the revised Horwitz-Thompson estimator

$$\widehat{Y}(\omega) = \frac{1}{n_2} \sum_{x \in s_2} Y(x,\omega)$$

under the local Poisson model is given by

$$\begin{aligned} E_\omega V_{x|\omega}\widehat{Y}(\omega) = & \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_2} \sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2 \\ & - \frac{1}{n_2} \frac{1}{\lambda(F)} \sum_{j=1}^{L} \frac{1}{\lambda(F_j)} \sum_{i \in F_j} Y_i^2 \end{aligned} \tag{3.17}$$

Remarks

- The global Poisson model, i.e. $L = 1$, is a special case of the local Poisson model
- The anticipated variance is a decreasing function of the inclusion probabilities.
- By the proof of 3.14 we see that the first two terms are of order n_2^{-1} and that the last term is of smaller order when $\lambda(F) \to \infty$, namely $\frac{1}{n_2 \lambda(F)}$.
- The anticipated variance under the local Poisson model is essentially equal to the anticipated variance under the global Poisson model augmented by the between-strata variance (which is independent of ω)
- If a stratum k consists of several disjoined components with the same averages $\bar{Y}_k$, we obtain the same anticipated variance by considering each component as a stratum on its own.

Since we are primarily interested in large areas we shall use from now on the following expression for **the anticipated variance of the one-phase one-stage Horwitz-Thompson estimator under simple random sampling and local Poisson model**

$$E_\omega V_{x|\omega}\widehat{Y}(\omega) = \frac{1}{n_2\lambda^2(F)}\sum_{i=1}^{N}\frac{Y_i^2}{\pi_i} + \frac{1}{n_2}\sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2 \tag{3.18}$$

We can now proceed to calculate the anticipated variance under cluster random sampling. Though the main ideas have been already outlined, the computations are, as expected, more cumbersome.

3.4.2 One-phase one-stage cluster random sampling

We use the same terminology and notation as in section 2.3, but for the fact that we now work with the revised Horwitz-Thompson estimator $Y(x_l)$ at each point $x_l = x + e_l$ of the cluster with origin x (see 3.4). For ease of notation we shall usually omit the symbol ω standing for the random pattern of the trees according to the local Poisson model.

According to 2.18 we calculate

$$E_x M^2(x)(Y_c(x) - \bar{Y})^2 = E_x\left(\sum_{l=1}^{M} I_F(x_l)(Y(x_l) - \bar{Y})\right)^2$$

which is equal to

$$E_x\left(\sum_{l=1}^{M} I_F(x_l)(Y(x_l) - \bar{Y})^2 + \sum_{l\neq k}^{M} I_F(x_l)I_F(x_k)(Y(x_l) - \bar{Y})(Y(x_k - \bar{Y})\right)$$

and hence to

$$\sum_{l=1}^{M} P(x_l \in F)V(Y(x_l)|x_l \in F)$$
$$+\sum_{l\neq k}^{M} P(x_l \in F, x_k \in F)E_x\{(Y(x_l) - \bar{Y})(Y(x_k) - \bar{Y})|x_l \in F, x_k \in F\}$$

According to 2.11 we know that, given $x_l \in F$, x_l is uniformly distributed in F so that the first term is the anticipated variance under simple random sampling, which is given by 3.17. The extra-term is due to cluster sampling and is more tricky. We need to calculate

$$E_\omega E_x\{(Y(x_l) - \bar{Y})(Y(x_k) - \bar{Y})|x_l \in F, x_k \in F\}$$

First we decompose the event

$$\{x_k \in F\} \cap \{x_l \in F\}$$

into the disjoined events

$$\left(\bigcup_{j=1}^{L}\{x_k \in F_j\} \cap \{x_l \in F_j\}\right) \bigcup \left(\bigcup_{i\neq j}^{L}\{x_k \in F_i\} \cap \{x_l \in F_j\}\right)$$

and use the decomposition rule for conditional expectation on disjoined A_s

$$E(Z|\cup_s A_s) = \frac{\sum_s E(Z|A_s)P(A_s)}{\sum_s P(A_s)}$$

Hence we must calculate the following terms

1. $E_\omega E_{x|\omega}\{(Y(x_l)-\bar{Y})(Y(x_k)-\bar{Y})|x_k \in F_i, x_l \in F_i\}$
2. $E_\omega E_{x|\omega}\{(Y(x_l)-\bar{Y})(Y(x_k)-\bar{Y})|x_k \in F_i, x_l \in F_j\}\ i \neq j$

To calculate the first expression we add $-\bar{Y}_i + \bar{Y}_i$, expand the product, interchange the order of expectation, use 3.15 so that $E_\omega Y(x) = \bar{Y}_i\ \forall x \in F_i$ to finally obtain

$$E_x E_{\omega|x} Y(x_k)Y(x_l) - \bar{Y}_i^2 + (\bar{Y}_i - \bar{Y})^2$$

At this point we make the assumption **that the same tree cannot be sampled from two different points of the cluster**, which is nearly always the case in practice, excepted may be for the angle count method and very large trees. That is, we assume that

$$I_i(x_k,\omega)I_i(x_l,\omega) = 0\ \forall i\ \forall k \neq l \tag{3.19}$$

Taking the expectation and after some algebra we see that term (1) is equal to

$$(\bar{Y}_i - \bar{Y})^2 - \frac{1}{\lambda^2(F_i)} \sum_{k \in F_i} Y_k^2$$

and therefore for large areas asymptotically equal to $(\bar{Y}_i - \bar{Y})^2$ (by the same arguments given in 3.14 applied to each stratum). Let us now tackle term (2). By adding $-\bar{Y}_i + \bar{Y}_i$ and $-\bar{Y}_j + \bar{Y}_j$ and using the same arguments as above we get

$$(\bar{Y}_i - \bar{Y})(\bar{Y}_j - \bar{Y})$$

Putting the pieces together we finally obtain the following intermediate result for the extra term of the anticipated variance

$$\sum_{i=1}^{L}\sum_{l\neq k}^{M}(\bar{Y}_i - \bar{Y})^2 P(x_k \in F_i, x_l \in F_i) + \sum_{i\neq j}^{L}\sum_{l\neq k}^{M}(\bar{Y}_i - \bar{Y})(\bar{Y}_j - \bar{Y})P(x_k \in F_i, x_l \in F_j)$$

To go further we introduce the random variables associated with the number of points of a cluster falling into a given stratum, i.e. we set

$$M_j(x) = \sum_{l=1}^{M} I_{F_j}(x_l)$$

It is then straightforward to see that the various sums of the $P(x_k \in F_i, x_l \in F_j)$ -like terms can be expressed as expectation, variances and covariances of the M_j. Using 3.17 we get the following result:

The anticipated variance of the one-phase one-stage estimate under cluster sampling and local Poisson model is given by

$$\begin{aligned} E_\omega V_{x|\omega}\widehat{Y}_c(\omega) = {} & \frac{1}{n_2 E_x M(x)} \left(\frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 \right) \\ & + \frac{1}{n_2 E_x M(x)} \sum_{j=1}^{L} \left(\frac{E_x M_j(x)(M_j(x)-1)}{E_x M(x)} \right) (\bar{Y}_j - \bar{Y})^2 \\ & + \frac{1}{n_2 E_x M(x)} \sum_{i\neq j}^{L} \left(\frac{E_x M_j(x) M_i(x)}{E_x M(x)} \right) (\bar{Y}_i - \bar{Y})(\bar{Y}_j - \bar{Y}) \end{aligned} \tag{3.20}$$

Since $\frac{E_x M_j(x)}{E_x M(x)} = p_j$ and $\sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y}) = 0$ we get after some simple algebra the following equivalent expression

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}_c(\omega) = & \frac{1}{n_2 E_x M(x)} \left(\frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} \right) \\ & + \frac{1}{n_2} V_x \left(\sum_{j=1}^{L} \frac{M_j(x)}{E_x M(x)} (\bar{Y}_j - \bar{Y}) \right) \end{aligned} \tag{3.21}$$

The above formulae look impressive at first sight, the following remarks are a help for interpretation.

Remarks

1. If we have only one single stratum (global Poisson forest) then, for the same total number of points, cluster random sampling leads formally to the same anticipated variance as simple random sampling, which is intuitively reasonable.

2. If we assume, formally, that given $M(x)$ the $M_j(x)$ have a multinomial distribution, then by conditioning on $M(x)$ and using 2.23 in 3.19 we see that in this case also cluster random sampling and simple random sampling have the same anticipated variance since the extra term is found to be $(\sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y}))^2 = 0$, because $\sum_{j=1}^{L} p_j \bar{Y}_j = \bar{Y}$. Of course, the multinomial distribution cannot hold, since the points of a cluster are not independent: if they were, we would have simple random sampling, which the above argument simply confirms. The physical interpretation of this result is the following: if the distances between the cluster points are comparable to the size of the components of the strata, then the intra-cluster correlation is small, since points of the same cluster are likely to be spread over several strata, and cluster sampling is close to random sampling.

3. The other extreme case is that all the points of a cluster will be usually in the same stratum. An approximate model is therefore to set $M_j(x) = M(x)$ with probability p_j and $M_j(x) = 0$ with probability $1 - p_j$, together with the condition that $M_i(x) M_j(x) = 0$ whenever $i \neq j$. In such a case the anticipated variance can be easily rewritten as

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}_c(\omega) \approx & \frac{1}{n_2 E_x M(x)} \left(\frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 \right) \\ & + \frac{1}{n_2 E_x M(x)} \left((E_x M(x) - 1) + \frac{V_x M(x)}{E_x M(x)} \right) \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 \end{aligned}$$

   Comparing this with formula 2.20 we can reinterpret the intra-cluster correlation coefficient as

$$\rho \approx \frac{\sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2}{E_\omega V_x Y(x, \omega)}$$

   Hence, if the strata are much larger than the dimension of the cluster, the intra-cluster correlation will increase with the between-strata variance and cluster sampling leads to a higher anticipated variance than simple random sampling. It can be expected that, in practice, this case gives the upper bound for the extra-term due to cluster sampling (mathematically it seems possible

that a specific stand structure combined with an ad-hoc cluster geometry may give higher values).

4. It is theoretically possible that the anticipated variance under cluster sampling is smaller than under simple sampling, for instance when $M_j(x)$ is always close to 1 for all j. In such a case, the strata can consist of many small connected components and the assumption of negligible boundary effects is questionable anyway, or the dimension of the cluster is much larger than the components of the strata and we are back to the second case with zero intra-cluster correlation.

For future use and also to simplify the complicated expression 3.20 we introduce the following notation

$$\beta^2 = \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 \tag{3.22}$$

β^2 is the between-strata variance, which depends only on the structure of the forest and not on the sampling scheme.

$$\alpha = E_x M(x) - 1 + \frac{V_x M(x)}{E_x M(x)} \tag{3.23}$$

α depends only on the geometry of the cluster and of the geometry of the forest area but not on the inclusion probabilities.

$$\begin{aligned} \delta = & \frac{\sum_{j=1}^{L} E_x M_j(x)(M_j(x)-1)(\bar{Y}_j - \bar{Y})^2}{E_x M(x) \alpha \beta^2} \\ & + \frac{\sum_{i \neq j}^{L} E_x (M_j(x) M_i(x))(\bar{Y}_i - \bar{Y})(\bar{Y}_j - \bar{Y})}{E_x M(x) \alpha \beta^2} \end{aligned} \tag{3.24}$$

δ depends on the geometry of the cluster and the structure of the forest, but not on the inclusion probabilities. It can be expected to lie between 0 and 1 in practice. Values near 1 indicate that most clusters are likely to have all their points within the forest area in the same stratum, whereas values near zero indicate that most clusters will be spread over several different strata. Theoretically, values below 0 and above 1 are possible, but unlikely to occur in standard cases. It is interesting to compare directly the anticipated variance obtained for simple and cluster random sampling:

- Simple random sampling

$$E_\omega V_{x|\omega} \widehat{Y}(\omega) = \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_2} \beta^2$$

- Cluster random sampling

$$E_\omega V_{x|\omega} \widehat{Y}_c(\omega) = \frac{1}{n_{2,c} E_x M(x)} \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_{2,c} E_x M(x)} (1+\theta) \beta^2$$

  where $\theta = \alpha\delta$ and, for the comparison to be meaningful, $n_2 = n_{2,c} E_x M(x)$ (so that the overall number of points is the same). The inflation factor θ can be expected to lie between 0 (when the M_j are independent multinomial) and $E_x M(x) - 1 + \frac{V_x M(x)}{E_x M(x)}$ (when most clusters lie entirely within one single stratum).

We are now able to generalize these results to the other sampling schemes.

3.4.3 One-phase two-stage simple random sampling

This is straightforward. By using 2.29, 2.30 and 3.17 we obtain immediately the following result:
The anticipated variance of the one-phase two-stage Horwitz-Thompson estimator under simple random sampling and the local Poisson model is given by

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*(\omega) = & \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_2} \sum_{k=1}^{L} p_k (\bar{Y}_k - \bar{Y})^2 \\ & + \frac{1}{n_2 \lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.25}$$

It is natural that the contribution to the variance of the second-stage procedure is unaffected by the forest structure, since it samples trees independently of each other.

3.4.4 One-phase two-stage cluster random sampling

Again, this is straightforward. By using 2.36, 2.30 and 3.19 we obtain the following result:
The anticipated variance of the one-phase two-stage Horwitz-Thompson estimator under cluster random sampling and local Poisson model is given by

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}_c^*(\omega) = & \frac{1}{n_2 E_x M(x)} \left(\frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \sum_{k=1}^{L} p_k (\bar{Y}_k - \bar{Y})^2 \right) \\ & + \frac{1}{n_2 E_x M(x)} \sum_{j=1}^{L} \left(\frac{E_x M_j(x)(M_j(x)-1)}{E_x M(x)} \right) (\bar{Y}_j - \bar{Y})^2 \\ & + \frac{1}{n_2 E_x M(x)} \sum_{i \neq j}^{L} \left(\frac{E_x (M_j(x) M_i(x))}{E_x M(x)} \right) (\bar{Y}_i - \bar{Y})(\bar{Y}_j - \bar{Y}) \\ & + \frac{1}{n_2 E_x M(x)} \frac{1}{\lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.26}$$

3.4.5 Two-phase one/two-stage simple random sampling

The calculation of the anticipated variance under two-phase sampling schemes seems to be beyond exact analytical treatment in the general case. However, by considering the special case of stratification it is possible to get at least a qualitative insight. **The main idea is to assume that the external prediction model used is a post-stratification model with the same strata as the local Poisson model for the random location of the trees**. Indeed, it does make sense to believe that the best post-stratification characteristic has something to do with the stand structure. We already know (see 2.49 and 2.51) that the design-based variance of the post-stratified estimate can be written as

$$V(\widehat{Y}_{reg}^*) = \frac{1}{n_2} \sum_{j=1}^{L} p_j V_j + \frac{1}{n_1} \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 + \frac{1}{n_2} E_x V(x)$$

or, more generally, as

$$V(\widehat{Y}^*_{reg}) = \frac{1}{n_2} V_x(Y(x)) + \frac{1}{n_2} E_x V(x) - \frac{1}{n_2}\left(1 - \frac{n_2}{n_1}\right) V_x(\widehat{Y}(x))$$

In contrast to section 2.8, we are using the revised Horwitz-Thompson estimate, which is unbiased also within each stratum. According to the asymptotic arguments given in section 2.8 one can identify $V_x(Y(x))$ with $\sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2$. To calculate the anticipated variance by the first formula it suffices to note that $\sum_{j=1}^{L} p_j V_j = V_x(Y(x)) - \sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2$ and to use 3.16 to obtain the following result:

The anticipated variance of the two-phase two-stage post-stratified estimate is given by

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*_{reg}(\omega) = & \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_1} \sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2 \\ & + \frac{1}{n_2 \lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \\ & - \frac{1}{n_2} \frac{1}{\lambda(F)} \sum_{k=1}^{L} \frac{1}{\lambda(F_k)} \sum_{i \in F_k} Y_i^2 \end{aligned} \tag{3.27}$$

For large areas we neglect as usual the last term, and we have

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*_{reg}(\omega) = & \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_1} \sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y})^2 \\ & + \frac{1}{n_2 \lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.28}$$

One sets formally $p_i \equiv 1$ in the above formula for the two-phase one-stage case, so that the last term vanishes.

The most interesting feature of this result is revealed by letting n_1 tend to infinity in the one-stage case: in practice this is the case when thematic maps are available and the surface areas of each stratum are known. Indeed, one obtains

$$\lim_{n_1 \to \infty} E_\omega V_{x|\omega} \widehat{Y}^*_{reg}(\omega) = \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i}$$

which is the anticipated variance of the one-phase one-stage estimate under the **global Poisson model**. In other words, **post-stratification filters out the heterogeneity of the the local Poisson forest by removing the between-strata variance**. This formula explains in an enlightening way why post-stratified estimates are more accurate than unstratified estimates (under the assumption, of course, that the stratification is meaningful).

By analogy and induction we can reasonably pretend that this result holds for general models by assuming that the predictions $\widehat{Y}(x)$ can be used to stratify the forest according to some threshold values. In practice, most linear models used are of the analysis of variance type since the auxiliary information is qualitative (e.g. stand map). The saturated models correspond to full stratification. Very often full stratification is not meaningful since the number of different strata can become very large, whereas a simple additive ANOVA model, possibly with a few interactions, may already provide an excellent fit. The prediction model should

yield asymptotically for each stratum F_k of the local Poisson model $\widehat{Y}(x) \equiv \bar{Y}_k$. Summarizing, and with the previous remark in mind, we can "state" the following result:

The anticipated variance of two-phase two-stage estimate under simple random sampling and local Poisson model is given by

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*_{reg}(\omega) = & \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_1} V_x(\widehat{Y}(x)) \\ & + \frac{1}{n_2 \lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.29}$$

The extension of these results to cluster sampling is intuitively clear.

3.4.6 Two-phase one/two-stage random cluster sampling

The general idea and justifications are the same as in simple random sampling. First, we replace asymptotically each prediction $\widehat{Y}(x_l)$ by the true stratum mean $\bar{Y}_j$ when $x_l \in F_j$; then we use for $E_x M^2(x)(\widehat{Y}_c(x) - \bar{Y})^2$ (which is independent of ω) the same decomposition as for $E_x M^2(x)(Y_c(x) - \bar{Y})^2$ (see the beginning of the proof of 3.19), and, as in simple sampling, we have the same substitution effect $n_2 \to n_1$ for the terms induced by the strata. This implies the following result:

The anticipated variance of two-phase two-stage post-stratified estimate under cluster random sampling and local Poisson forest is given by

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*_{c,reg}(\omega) = & \frac{1}{n_2 E_x M(x)} \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_1 E_x M(x)} \sum_{j=1}^{L} p_j (\bar{Y}_j - \bar{Y})^2 \\ & + \frac{1}{n_1 E_x M(x)} \sum_{j=1}^{L} \left(\frac{E_x M_j(x)(M_j(x) - 1)}{E_x M(x)} \right) (\bar{Y}_j - \bar{Y})^2 \\ & + \frac{1}{n_1 E_x M(x)} \sum_{i \neq j}^{L} \left(\frac{E_x M_j(x) M_i(x)}{E_x M(x)} \right) (\bar{Y}_i - \bar{Y})(\bar{Y}_j - \bar{Y}) \\ & + \frac{1}{n_2 E_x M(x)} \frac{1}{\lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.30}$$

Simple algebra as in the proof of 3.20 yields the equivalent formula

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*_{c,reg}(\omega) = & \frac{1}{n_2 E_x M(x)} \left(\frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} \right) \\ & + \frac{1}{n_1} V_x \left(\sum_{j=1}^{L} \frac{M_j(x)}{E_x M(x)} (\bar{Y}_j - \bar{Y}) \right) \\ & + \frac{1}{n_2 E_x M(x)} \frac{1}{\lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.31}$$

One sets formally $p_i \equiv 1$ in the above formula for the two-phase one-stage case, so that the last term vanishes.

We observe the same effect as for simple random sampling: post-stratification can substantially reduce the variance if the between-strata variance and hence

also the intra-cluster correlation are large. In the limiting case $\lim n_1 \to \infty$ post-stratified cluster sampling yields the same anticipated variance per sample point as simple random sampling.

Asymptotically we can replace $\widehat{Y}(x_l)$ by $\bar{Y}_j$ whenever $x_l \in F_j$, which implies $M(x)\widehat{Y}_c(x) = \sum_{j=1}^{L} \bar{Y}_j$. Consequently, and under the same conditions as for simple random sampling, we can "state" the following result:

The anticipated variance of the two-phase two-stage estimate under cluster random sampling and local Poisson model is given by

$$\begin{aligned} E_\omega V_{x|\omega} \widehat{Y}^*_{c,reg}(\omega) = & \frac{1}{n_2 E_x M(x)} \left(\frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} \right) \\ & + \frac{1}{n_1} V_x \left(\frac{M(x)(\widehat{Y}_c(x) - \bar{Y})}{E_x M(x)} \right) \\ & + \frac{1}{n_2 E_x M(x)} \frac{1}{\lambda^2(F)} \left(\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} - \sum_{i=1}^{N} \frac{R_i^2}{\pi_i} \right) \end{aligned} \tag{3.32}$$

In the next section we present an heuristic argument justifying the use of the anticipated variance also under systematic sampling.

3.5 The anticipated variance under systematic sampling

All the calculations of the anticipated variance rest upon the fundamental result 3.13, obtained for one-phase one-stage random sampling under the global Poisson forest. Formally, one can consider a systematic sample as one single huge cluster with $n_{2,c} = 1$ and $\bar{M} = n_2$, the number of points of the grid falling into the forest area. We have seen in 3.19 that, in this case, we get the same anticipated variance as in simple random sampling. This argument, however, is purely formal, since the formulae given for cluster sampling were only asymptotically valid, i.e. for a large number of clusters.

We now give a further simple heuristic argument showing that this result is likely to hold in large areas also under systematic sampling. We consider a global Poisson forest and a rectangular grid with fundamental cell F_0. To simplify the geometry we assume that $F = \bigcup_{k=1}^{n_2} F_{o,k}$ where all the $F_{o,k}$ are congruent to F_0. A systematic grid with random origin in e.g. $F_{0,1}$ and fixed orientation generates in each $F_{0,k}$ a point x_k located up to translation in the same place. Let us now overlay conceptually all the congruent cells with their trees (recall that in our model trees are dimensionless points) onto the fundamental cell F_0. Drawing a systematic grid is then equivalent to drawing **a single point $\tilde{x}$ uniformly in F_0**. To be more precise one should also identify the opposite edges of F_0 (we transform the rectangular cell F_0 into a torus) and also fold over the edges the circles K_i which hit the boundary of F_0 (since a tree near the boundary of its cell $F_{0,k}$ can be sampled from a point in an adjacent cell). We neglect the boundary effects within each $F_{0,k}$ so that the inclusion probabilities with respect to $\tilde{x}$ are now given by

$$\pi_{i,0} = \frac{\lambda(K_i)}{\lambda(F_0)} \ , \quad \pi_{ij,0} = \frac{\lambda(K_i \cap K_j)}{\lambda(F_0)}$$

Even if the assumption of a global Poisson forest does not hold exactly in F it is intuitively clear that the artificial overlaid forest defined in F_0 is, for each ω, unlikely to display any structure: local repulsion or aggregation pattern of trees will have

been destroyed to a great extent by the overlay. We can therefore reasonably expect that the overlaid trees are independently uniformly distributed in F_0, even if this is not exactly the case in F. Consider now the Horwitz-Thompson estimator on the torus defined as

$$Y_0(\tilde{x},\omega) = \frac{1}{\lambda(F_0)} \sum_{i=1}^{N} \frac{Y_i}{\pi_{i,0}} I_i(\tilde{x},\omega)$$

where the indicator variable $I_i(\tilde{x},\omega)$ is 1 when the i-th tree is sampled from one of the x_k, otherwise 0. By construction $Y_0(\tilde{x},\omega)$ is an unbiased estimate of the overlaid forest, i.e.

$$E_{\tilde{x}|\omega} Y_0(\tilde{x},\omega) = \frac{1}{\lambda(F_0)} \sum_{i=1}^{N} Y_i = n_2 \bar{Y} \quad \forall\omega$$

and its anticipated variance is given by

$$E_\omega V_{\tilde{x}|\omega}(Y_0(\tilde{x},\omega)) = \frac{1}{\lambda^2(F_0)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_{i,0}} - \frac{1}{\lambda^2(F_0)} \sum_{i=1}^{N} Y_i^2$$

It is easily verified that, by construction, the usual Horwitz-Thompson estimate $\widehat{Y}(\omega) = \frac{1}{n_2} \sum_{k=1}^{N} Y(x_k,\omega)$ and the above Horwitz-Thompson estimate $Y_0(\tilde{x},\omega)$ are linked by

$$Y_0(\tilde{x},\omega) = n_2 \widehat{Y}(\omega) = \sum_{k=1}^{n_2} Y(x_k,\omega)$$

so that $\frac{1}{n_2} Y_0(\tilde{x},\omega) = \widehat{Y}(\omega)$ and consequently

$$E_\omega V(\widehat{Y}(\omega)) = \frac{1}{n_2^2} E_\omega V_{\tilde{x}|\omega}(Y_0(\tilde{x},\omega))$$

Since $n_2\lambda(F_0) = \lambda(F)$ and $\lambda(F_0)\pi_{i,0} = \lambda(F)\pi_i$ one finally obtains

$$E_\omega V(\widehat{Y}(\omega)) = \frac{1}{n_2\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} - \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} Y_i^2$$

It is interesting to note that one has formally the same result by assuming that the trees are uniformly and independently distributed within each $F_{0,k}$ and that the x_k are independently uniformly distributed in $F_{0,k}$ (stratified random sampling). To get further insight we write the anticipated variances in terms of the

$$\alpha(F) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\lambda(K_i)}\ , \quad 0 < \alpha(F) < \infty$$

$$\beta(F) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} Y_i^2\ , \quad 0 < \beta(F) < \infty$$

The anticipated variance under simple random sampling (see 3.13) can be rewritten as

$$E_\omega V_{random}(\widehat{Y}(\omega)) = \frac{\lambda(F_0)}{\lambda(F)} \left(\alpha(F) - \frac{\beta(F)}{\lambda(F)} \right)$$

after reinterpreting $\lambda(F_0)$ as the mean sampled area per point, i.e. $\lambda(F_0) = \frac{\lambda(F)}{n_2}$. The anticipated variance under random cluster sampling with only one single large cluster of expected size n_2 is, formally, given by the same expression, i.e.

$$E_\omega V_{cluster}(\widehat{Y}(\omega)) = \frac{\lambda(F_0)}{\lambda(F)} \left(\alpha(F) - \frac{\beta(F)}{\lambda(F)} \right)$$

This result is purely formal since the formulae for cluster sampling are only valid for a large number of clusters. Note also that in 3.19 we neglected the term $-\frac{1}{\lambda^2(F)}\sum_{i=1}^{N} Y_i^2$ which can be added in the first part of the proof by using 3.13. The anticipated variance under systematic sampling is, by the previous arguments, given by

$$E_\omega V_{syst}(\widehat{Y}(\omega)) = \frac{\lambda(F_0)}{\lambda(F)}\left(\alpha(F) - \frac{\beta(F)}{\lambda(F_0)}\right)$$

The anticipated variance under systematic sampling is therefore smaller than under random sampling, in agreement with the well known fact that treating a systematic sample as a random sample usually leads to an overestimation of the variance. Now, if $\lambda(F_0) \to \infty$ and, consequently also $\lambda(F) \to \infty$, the three approaches are asymptotically equivalent: this is the case for constant sampling density over increasingly larger areas, but not when $n_2 \to \infty$ over finite areas, which, however, is unlikely to occur in practice!

Let us now look at the local Poisson forest. By considering again a systematic grid formally as a single large cluster of expected size n_2 and replacing in 3.19 the M_j by their expected values, we see that the second and third term add up to approximately $(\sum_{j=1}^{L} p_j(\bar{Y}_j - \bar{Y}))^2 = 0$. This implies that the anticipated variance is the same as under simple random sampling with n_2 points. This is legitimate if the (random) numbers of points of the grid falling into F and the F_j are all sufficiently large to be replaced by their expected values and if the bias of the point estimate based on a single very large cluster is negligible, which is likely to be the case for inventories performed on large areas. For global and local Poisson forests we can therefore conjecture that in large samples the anticipated variances under simple one-phase random and systematic sampling should be close to each other.

According to geostatistical techniques ([12]) the spatial autocorrelation range within a global Poisson forest is very short (spherical variogram with range shorter than $2 \max r_i$) and the geostatistical variance is very close to the variance under random sampling. The autocorrelation range is primarily determined by the spatial stand structure. There is empirical and theoretical evidence that in two-phase sampling schemes the autocorrelation range of the residuals is very small, whereas the autocorrelation range of the predictions is also determined by the stand structure. Since, as we have seen, two-phase sampling and the associated estimation techniques amount essentially to transform a local Poisson forest into a global Poisson forest one can conjecture that the anticipated variance formulae are more appropriate under two-phase systematic sampling than under one-phase systematic sampling.

The estimation of variance under systematic sampling is, strictly speaking, impossible within classical sampling theory (unless, of course, one has replicates of the grid, which is rarely if ever the case in forest inventory). One must either treat systematic samples as random samples (eventually with ad-hoc procedures like paired differences, but this is not satisfactory) or by using geostatistical techniques as in [12]. We see that the anticipated variance may provide a simple alternative, since it can be estimated by standard techniques with $\frac{Y_i^2}{\lambda(K_i)}$ as a new response variable.

Before going into the proper optimization task we have to discuss the costs induced by the various procedures.

Chapter 4

Modelization of costs

4.1 Generalities

The modelization of costs is a difficult and up to a certain degree illusory task, but obviously unavoidable in the context of cost-efficient inventory techniques. The ultimate objective is to design sampling schemes which minimize the variance for a given budget or which minimize the costs for a given variance. As we have seen, there is no best estimator for all forests and a sensible strategy is to consider an admissible estimator which performs well on average for a class of forests. The anticipated variance under the local Poisson model provides precisely the necessary tools to do that in the simplest class. Recall that in this model the N values Y_i of the response variable are fixed, whereas the trees are uniformly and independently distributed within the L strata. Hence, the restricted objective is to determine the inclusion probabilities π_i in order to minimize the anticipated variance of the Horwitz-Thompson under given costs or conversely to minimize the costs under a given anticipated variance.

There are of course many difficulties to define the costs involved in forest inventory and an exhaustive discussion is beyond the scope of this work (and probably also unnecessary). It is however general practice to distinguish between **fixed costs** (e.g. salaries of permanent senior staff, technical equipment like computers, devices for the interpretation of aerial photographs and image analysis, software, cars etc) and **variable costs** which depend on the inventory procedures (e.g. number of plots, number of trees measured, number of aerial photographs interpreted, travelling costs etc). Up to a certain degree the inventory technique has little impact on the fixed overhead costs, but obviously a large impact on the variable costs. We shall therefore consider in this work only the variable costs. The variable costs can be split into essentially two components: the **costs of measurements** (e.g. for installation of the field plots, measuring trees, orientation and interpretation of the aerial photographs) and **travelling costs** (time spent to access the field plots, gas-oil, lodging etc). Well conducted pilot studies and past experience allow generally for a good appraisal of the measurement costs whereas travelling costs are usually more difficult to assess for inventories of large surfaces (the shortest distances from the plots to the road network and lodging facilities are among other things rather tricky to estimate or extrapolate from existing data).

The measurement costs are essentially linear in the number of points (field plots or aerial photographs) and the expected number of trees sampled in one point.

With systematic grids the shortest overall distance required to go through all field plots is roughly proportional to the square root of the number of plots (this is obvious for a quadratic grid while neglecting boundary effects). The shape of

the grid is also important: rectangular grids yield, for the same number of points, much shorter overall distances than quadratic grids (consider the extreme case of a grid reduced to one single line). It can be shown that the square root law remains valid for uniform sampling (see [14]). We have said in the introduction that we shall consider systematic grids as random samples, which is certainly acceptable for estimation of large domains. The correct analysis of systematic grids requires the use of geostatistical methods to calculate the variance. In his pioneer work B. Matérn [14] investigated the impact of the grid shape on the accuracy and on the overall distance; he recommended to use rectangular grids with ratios "length to width" of at least three to one as long as the distance between neighboring points is larger than the range of the spatial correlation. These were however preliminary results (and, nearly forty years later, unfortunately almost the only ones!), valid for a special covariance function and an "infinite square forest without holes". From a pragmatic point of view, one can say that it is certainly worthwhile considering rectangular grids as long as the grids cover the entire forest area in a "representative way" (which is certainly not the case for a single line). In this work $\phi(n_2)$ will denote the overall travelling costs resulting from having n_2 points. We can expect $\phi(n_2)$ to be roughly proportional to $\sqrt{n_2}$, the proportionality constant depending among many other things also on the shape of the grid. With Geographical Information Systems (GIS) it is in principle possible to determine $\phi(n_2)$ numerically for a wide range of n_2 and several grid shapes. For this reason we shall give some of the results in term of $\phi(n_2)$ for given n_2. It is however important to keep in mind that all the calculations are valid only under the assumption that systematic grids can be treated as random samples, which certainly narrows down the choice of feasible grids. Furthermore, common sense and experience tell us that complex cost function resulting in complex calculation are rarely worthwhile. For this reason it is useful to approximate the function $\phi(n_2)$ by a linear function over an adequate range for n_2, for instance by least square (it is generally very simple to determine such a feasible range). The main advantage of a linearized $\phi(n_2)$ is that simple analytical results are available (for instance square root-like $\phi(n_2)$ already lead to cubic equations!) and that the dual problems **minimizing the costs for given anticipated variance versus minimizing the anticipated variance for given costs** lead to the same solutions (but for constants), which is not the case with non-linear $\phi(n_2)$.

We believe that the cost functions presented here at least capture the main features and that the reader should have in principle no difficulty to adapt the results to his own cost functions (usually the only differences are due to minor changes in the definitions of the various coefficients, but the mathematical structure remains the same). Let us now go into the details.

4.2 Cost functions

4.2.1 Simple random sampling

First, we note that by 2.6 the expected number of trees sampled at the first stage is given by $\sum_{i=1}^{N} \pi_i$ and that, with the same argument, the expected number of trees sampled at the second stage is given by $\sum_{i=1}^{N} \pi_i p_i$. We introduce the following definitions for the various costs, expressed preferably in man-time unit (e.g. 10 minutes corresponds to 5 minutes work for a team of two persons, modifications to take wage differences into account are obvious):

1. c_0 is the mean unit installation cost per second phase point. This entails for instance the time required to locate exactly the sample point (but not to access it), to describe it (important for its allocation to a given stratum), to delimit e.g. concentric circles etc.

2. c_1 is the mean unit cost per first phase point to collect the auxiliary information. This might entail for instance the mean time required to orientate and interpret aerial photographs (whether flight and development costs are included depends on the particular circumstances), or the time required to perform ocular assessment of the timber volume, etc.

3. c_2 is the mean unit cost per first stage tree to measure the exact response variable Y_i. This might entail for instance the time required to measure diameters at 1.3 and 7 meters as well as the height if Y_i is the timber volume.

4. c_{21} is the mean unit cost per first stage tree to obtain the approximate value Y_i^* of Y_i. This might entail for instance the time required to measure only the diameter at 1.3 meter.

5. c_{22} is the mean unit cost per second stage tree to perform the extra measurement required to know the exact response Y_i. This might entail for instance the time required to measure diameter at 7 meter and the height.

Remarks

- When calculating relative efficiencies of the various inventory schemes we shall make the reasonable assumption that $c_2 = c_{21} + c_{22}$
- One can obviously generalize the above frame work by allowing all the unit costs to depend on the stratum. The mathematics of optimization remain essentially the same but the resulting formulae are obviously more cumbersome.

We are therefore led to the following cost functions:

1. **One-phase one-stage simple random sampling**

$$\phi(n_2) + n_2 c_0 + n_2 c_2 \sum_{i=1}^{N} \pi_i \tag{4.1}$$

2. **One-phase two-stage simple random sampling**

$$\phi(n_2) + n_2 c_0 + n_2 c_{21} \sum_{i=1}^{N} \pi_i + n_2 c_{22} \sum_{i=1}^{N} \pi_i p_i \tag{4.2}$$

3. **Two-phase one-stage simple random sampling**

$$\phi(n_2) + n_1 c_1 + n_2 c_0 + n_2 c_2 \sum_{i=1}^{N} \pi_i \tag{4.3}$$

4. **Two-phase two-stage simple random sampling**

$$\phi(n_2) + n_1 c_1 + n_2 c_0 + n_2 c_{21} \sum_{i=1}^{N} \pi_i + n_2 c_{22} \sum_{i=1}^{N} \pi_i p_i \tag{4.4}$$

By linearizing the travelling cost we get $\phi(n_2) \approx \alpha + \beta n_2$ and the slope coefficient β can be added to the installation cost. In this case c_0 becomes $\tilde{c}_0 = c_0 + \beta$. Likewise, the available budget C is reduced by the intercept term α and we shall write $\tilde{C} = C - \alpha$. Let us now look at cluster sampling.

4.2.2 Cluster random sampling

By 2.6, 2.11 and 2.12 it is clear that all terms in c_2, c_{21}, c_{22} must be multiplied by $E_x M(x)$ to have the expected costs per cluster. This is also approximately the case for c_1 in the case of aerial photographs but for slightly different orientation costs. For the installation costs this is the case if one ignores the travelling costs between the points of the same cluster. For these reasons we shall write the installation costs as $E_x M(x) c_{c0}$ and the first-phase costs as $E_x M(x) c_{c1}$. Whether then $c_{c0} \approx c_0$ and $c_{c1} \approx c_1$ depends on the particular circumstances.

We therefore assume that the following cost functions hold under cluster random sampling:

1. **One-phase one-stage cluster random sampling**

$$\phi(n_2) + n_2 E_x M(x) c_{c0} + n_2 E_x M(x) c_2 \sum_{i=1}^{N} \pi_i \tag{4.5}$$

2. **One-phase two-stage cluster random sampling**

$$\phi(n_2) + n_2 E_x M(x) c_{c0} + n_2 E_x M(x) c_{21} \sum_{i=1}^{N} \pi_i + n_2 E_x M(x) c_{22} \sum_{i=1}^{N} \pi_i p_i \tag{4.6}$$

3. **Two-phase one-stage cluster random sampling**

$$\phi(n_2) + n_1 E_x M(x) c_{c1} + n_2 E_x M(x) c_{c0} + n_2 E_x M(x) c_2 \sum_{i=1}^{N} \pi_i \tag{4.7}$$

4. **Two-phase two-stage cluster random sampling**

$$\begin{aligned} &\phi(n_2) + n_1 E_x M(x) c_{c1} + n_2 E_x M(x) c_{c0} + n_2 E_x M(x) c_{21} \sum_{i=1}^{N} \pi_i \\ &+ n_2 E_x M(x) c_{22} \sum_{i=1}^{N} \pi_i p_i \end{aligned} \tag{4.8}$$

As under simple random sampling we shall replace c_{c0} by $\tilde{c}_{c0}$ and C by $\tilde{C}$ after linearizing the travelling costs $\phi(n_2)$.

4.3 Similarity law for travelling costs

It frequently happens in practice that one knows for an inventory performed in the area A_p the overall travelling costs C_p and hence the mean travelling cost per point $\bar{c}_p = \frac{C_p}{n_p}$. One would like to extrapolate this knowledge to another area A with the same characteristics with respect to travelling conditions.

Let us assume that the square root law is a good approximation, hence we can write

$$C_p = \alpha \sqrt{n_p} \sqrt{\lambda(A_p)}$$

for some constant α. Since $C_p = n_p \bar{c}_p$ we have $\alpha = \frac{\bar{c}_p}{\sqrt{n_p}\sqrt{\lambda(A_p)}}$. If the travelling conditions are the same we can expect that the overall travelling costs for n_2 points in the region A are given by $\phi(n_2) = \alpha \sqrt{n_2} \sqrt{\lambda(A)}$ and consequently by

$$\phi(n_2) = \bar{c}_p \sqrt{\frac{\lambda(A)}{\lambda(A_p)}} \sqrt{\frac{n_2}{n_p}}$$

which is useful to investigate the impact of travelling costs for various sampling densities on the basis of pilot studies.

We are now finally ready to tackle the main problem of optimization.

Chapter 5

Optimal sampling schemes

5.1 Overview

We shall now derive optimal sampling schemes by minimizing the anticipated variance for given costs or conversely. The strategy is the following: for given sample sizes n_1 and n_2 we determine the optimal inclusion probabilities π_i by using the Cauchy-Schwartz inequality (which will be explained below). Up to this point the travelling cost function $\phi(n_2)$ is arbitrary and one could, at least in principle, minimize numerically the anticipated variance based on the optimal π_i with respect to n_2, n_1 in any given practical case. To go further we linearize $\phi(n_2)$ and minimize analytically the anticipated variance based on the optimal π_i, which depend on n_2, n_1 by treating the sample sizes as continuous variables. To do that requires frequently the technique of the Lagrange multipliers. The reader unfamiliar with it should maybe have a quick brush-up on that topic by consulting any standard book on calculus. The technique is essentially as follows: to find the extrema of a function $f(x_1, x_2, \ldots, x_n)$ of n variables subject to K constraints $g_i(x_1, x_2, \ldots, x_n) \equiv C_i,\ i = 1, 2 \ldots K$ it is sufficient to find the extrema of the so-called Lagrange function

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_K) = f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{K} \lambda_i (g_i(x_1, \ldots, x_n) - C_i)$$

This is done by solving the system of equations

$$\frac{\partial L}{\partial x_i} = 0,\ \frac{\partial L}{\partial \lambda_k} = 0,\ i = 1, 2 \ldots n,\ k = 1, 2 \ldots K$$

Of course the sample sizes are integer and not real numbers, but using more sophisticated techniques of integer programming is not worthwhile; beside, those techniques do not provide analytical solutions, which is what we must have in order to get a qualitative insight. The theoretical results are essentially as follows: the inclusions probabilities (**P**) are always either proportional to size (**S**), prediction (**P**) or error (**E**), which we shall for short refer to as **PPS, PPP, PPE**. As we shall see **PPS** is not always feasible (the grand exception being of course the basal area with the angle count technique) but can be approximated by two-stage procedures. Formally one could almost write **PPS=PPP+PPE** since **S=P+E**. In two-phase procedures the sample size n_2 should be kept as small as possible, likewise in one-phase procedures in global Poisson forest. It is generally impossible to implement in practice the exact optimal schemes. For this reason we give also discrete optimal approximations (e.g. with 1,2,3,4 concentric circles).

As announced, and for completeness, let us now state the famous Cauchy-Schwartz inequality. For any two sets of real numbers $a_i, b_i,\ i = 1, 2 \ldots, n$ one has

$$\left(\sum_{i=1}^{n} a_i b_i \right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \tag{5.1}$$

with equality if and only if $\exists\ \alpha\ \forall i\ a_i = \alpha b_i$. This extremely useful inequality is an immediate consequence of the following algebraic equality

$$\left(\sum_{i=1}^{n} a_i b_i \right)^2 = \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i b_j - a_j b_i)^2$$

Setting $a_i = \frac{|Y_i|}{\sqrt{\pi_i}}$ and $b_i = \sqrt{\pi_i}$ in 5.1 we obtain

$$\sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} \geq \frac{\left(\sum_{i=1}^{N} |Y_i| \right)^2}{\sum_{i=1}^{N} \pi_i} \tag{5.2}$$

and similarly

$$\sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i} \geq \frac{\left(\sum_{i=1}^{N} |R_i| \right)^2}{\sum_{i=1}^{N} \pi_i p_i} \tag{5.3}$$

Since in practice $Y_i \geq 0 \forall i$ we shall write $|Y_i| = Y_i$; for the residuals we must of course keep the absolute value.

Remarks

- The optimization rests upon the obtained anticipated variances and thefore on the assumption of negligible boundary effects for sufficiently regular forest and strata shapes. This implies that for a large majority of trees the inclusion ares $\lambda(K_i \cap F_k)$ will be equal or at least roughly proportional to $\lambda(K_i)$. The resulting selection rules should therefore not be seriously affected by boundary effects. Departures from the local Poisson model and non-negligible boundary effects are more likely to affect the numerical agreement between the anticipated variance and the empirical variance for a given forest.

- If, for any two given sampling schemes and given costs, the ratio of the anticipated variances is close to the ratio of the corresponding empirical variances, that is, if the anticipated and empirical relative efficiencies are close, then decision making based on the anticipated variance is valid, even under numerical discrepancies between anticipated and empirical variances.

We are now ready to proceed with the optimization of the sampling schemes we have analyzed so far. In order to save space we shall omit most of the trivial but tedious calculations, like setting derivatives to zero, checking that we have indeed a minimum and calculating its value.

5.2 One-phase one-stage simple random sampling

We consider the local Poisson model and we want to minimize the anticipated variance for a given overall budget C. According to 3.17 and 4.1 we have to minimize

$$E_\omega V_{x|\omega} \widehat{Y}(\omega) = \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} + \frac{1}{n_2} \sum_{k=1}^{L} p_k (\bar{Y}_k - \bar{Y})^2$$

under the constraint

$$\phi(n_2) + n_2 c_0 + n_2 c_2 \sum_{i=1}^{N} \pi_i \leq C$$

It is clear that the inequality can be replaced by an equality since by increasing the budget we can obviously reduce the variance. For given n_2 the lower bound of

$$\frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i}$$

is, by 5.2, achieved for $\pi_i = \alpha |Y_i|$ and is equal to

$$\frac{1}{n_2 \lambda^2(F) \alpha} \sum_{i=1}^{N} |Y_i|$$

Replacing π_i by $\alpha |Y_i|$ in the constraint we get

$$\alpha = \frac{C - n_2 c_0 - \phi(n_2)}{n_2 c_2} \frac{1}{\sum_{i=1}^{N} |Y_i|}$$

so that we finally have the solution

$$\begin{aligned} \lambda(F)\pi_i &= \frac{C - n_2 c_0 - \phi(n_2)}{n_2 c_2} \frac{|Y_i|}{\bar{Y}} \\ \min_{\pi_i | n_2} E_\omega V_{x|\omega} \widehat{\bar{Y}}(\omega) &= \frac{c_2 \bar{Y}^2}{C - n_2 c_0 - \phi(n_2)} + \frac{1}{n_2} \beta^2 \end{aligned} \tag{5.4}$$

where we have introduced the between-strata variance

$$\beta^2 = \sum_{k=1}^{L} p_k (\bar{Y}_k - \bar{Y})^2$$

To simplify the notation we introduce the symbol MAV to denote the minimum of the anticipated variance for given n_2, i.e. we set

$$MAV(\widehat{\bar{Y}}) = min_{\pi_i | n_2} E_\omega V_{x|\omega} \widehat{\bar{Y}}(\omega)$$

Clearly 5.4 is a **PPS** scheme if $Y_i \geq 0$ which is always the case in practice. Under the global Poisson model $\beta^2 = 0$ and we have

$$\frac{\partial MAV(\widehat{\bar{Y}})}{\partial n_2} \geq 0$$

in the range $C - n_2 c_0 - \phi(n_2) \geq 0$ since $\frac{\partial \phi(n_2)}{\partial n_2} \geq 0$ for each reasonable travelling cost function (they have to increase with the number of points). Hence, for a global Poisson forest the number of points in the one-phase one-stage scheme must be kept as small as possible (mathematically $n_2 = 1$ is the best choice). What does this mean? The global Poisson forest is homogenous and, intuitively speaking, looks much the same everywhere; hence it does not make sense to waste resources, i.e. travelling expenses, to observe it everywhere. Obviously one has to down weight this mathematical result in practice. First, small n_2 generate very large plots and consequently the assumption of negligible boundary effects will be violated. Furthermore, in practice, very large plots can be awkward to handle because of numerous boundary adjustments and slope corrections; finally one needs a sufficiently large number

of points to estimate the variance and eventually to get separate estimates for subareas. **The correct interpretation of this mathematical result is therefore to choose n_2 large enough to fulfill the practical and theoretical constraints but not much larger**.

The situation is totally different in the local Poisson forest. It is easy to see that $\frac{\partial MAV(\widehat{Y})}{\partial n_2}$ can be negative for small n_2 and positive for large n_2 so that there exists a true minimum. For linearized $\phi(n_2)$ the solution is easily found to be

$$n_{2,opt} = \frac{\tilde{C}}{\tilde{c}_0 + \sqrt{\tilde{c}_0 c_2}\frac{\bar{Y}}{\beta}}$$

where $\tilde{C} = C - intercept(\phi(n_2))$ and $\tilde{c}_o = c_o + slope(\phi(n_2))$. The lower bound of the anticipated variance is then

$$\min_{n_2} MAV(\widehat{Y}) = \frac{\left(\sqrt{c_2}\bar{Y} + \sqrt{\tilde{c}_0}\beta\right)^2}{\tilde{C}}$$

We now compare the optimal scheme with the equal probability scheme $\pi_i \equiv \pi$. Trivial calculations show that the optimal n_2 for this scheme (there is nothing to optimize with the π_i) lead to the following lower bound for the anticipated variance

$$\min_{n_2, \pi_i \equiv \pi} MAV(\widehat{Y}) = \frac{\left(\sqrt{\frac{c_2 N \sum_{i=1}^N Y_i^2}{\lambda^2(F)}} + \sqrt{\tilde{c}_0}\beta\right)^2}{\tilde{C}}$$

Hence, the **PPS** scheme is more efficient than equal probability sampling if and only if

$$N\sum_{i=1}^N Y_i^2 \geq \left(\sum_{i=1}^N Y_i\right)^2$$

By Cauchy-Schwartz inequality (set $a_i \equiv 1, b_i = Y_i$) this is always the case, with equality if Y_i is constant. Hence, equal probability sampling is optimal for the overall number of stems per ha, otherwise not.

Let us now consider the **dual problem**, i.e.
minimize

$$\phi(n_2) + n_2 c_2 \sum_{i=1}^n \pi_i$$

under the constraint

$$\frac{1}{n_2\lambda^2(F)}\sum_{i=1}^N \frac{Y_i^2}{\pi_i} + \frac{1}{n_2}\beta^2 = W$$

By 5.2 in the reverse order we have

$$\sum_{i=1}^N \pi_i \geq \frac{\left(\sum_{i=1}^N Y_i\right)^2}{\sum_{i=1}^N \frac{Y_i^2}{\pi_i}}$$

the lower bound being again achieved with the **PPS** rule $\pi_i = \alpha Y_i$ and is then equal to $\alpha \sum_{i=1}^N Y_i$. The constraint yields at once

$$\alpha = \frac{1}{\lambda^2(F)}\frac{\sum_{i=1}^N Y_i}{n_2 W - \beta^2}$$

and the cost functions becomes

$$c(n_2) = \phi(n_2) + c_2 \frac{\bar{Y}^2}{W - \frac{\beta^2}{n_2}}$$

In global Poisson forests this is an increasing function of n_2 so that again n_2 should be kept as small as possible. In local Poisson forests there is a true minimum which for the linearized $\phi(n_2)$ is easily found to be given by

$$c(n_{2,opt}) = \frac{\left(\sqrt{c_2}\bar{Y} + \sqrt{\tilde{c}_0}\beta\right)^2}{W}$$

We have neglected, and we shall always do so, the term $intercept(\phi(n_2))$ in the above calculation : first, it is usually small compared with the other terms, second, when comparing two sampling schemes the intercept term is the same and can be considered as a fixed overhead cost. For illustration we consider again the equal probability sampling scheme $\pi_i \equiv \pi$ for which the lower bound of the cost at the optimal value of n_2 is given by

$$\frac{1}{W}\left(\sqrt{\frac{c_2 N \sum_{i=1}^N Y_i^2}{\lambda^2(F)}}\right)^2$$

We note the duality of the formulae for the lower bounds of the costs and anticipated variances: **they are the same but for exchanging the** C **and** W . Hence, the relative efficiencies of equal probability sampling with respect to **PPS** at the respective optimal sample sizes is the same whether we look at the ratios of the minimum of the anticipated variances or of the minimum of the expected costs. This very nice property is fairly general and characteristic of cost function linear in the sample size n_2, as we shall see, but it is no longer true for non-linear cost functions, in particular for non-linear travelling costs $\phi(n_2)$. Let us denote by $RE(1|2)$ the relative efficiency, in the above sense, of the sampling schemes 1 and 2. We have

$$RE(\mathbf{PPS}|\pi_i \equiv \pi) = \left(\frac{\sqrt{c_2}\bar{Y} + \beta\sqrt{\tilde{c}_0}}{\sqrt{c_2}\sqrt{\bar{N}\overline{Y^2}} + \beta\sqrt{\tilde{c}_0}}\right)^2 \leq 1$$

where we have set $\bar{N} = \frac{N}{\lambda(F)}$ and $\overline{Y^2} = \frac{\sum_{i=1}^N Y_i^2}{\lambda(F)}$. The equality holds by Cauchy-Schwartz if and only if the Y_i are constant.

The above result shows in particular that the angle count method is optimal for the basal area, with respect to the anticipated variance for local Poisson forests. The optimality of **PPS** procedures in standard sampling theory is usually presented as a consequence of the Yates-Grundy formula valid for fixed sample sizes. For completeness we give a generalization of this formula valid also in the inventory context. We first note that

$$\frac{1}{2}\sum_{i\neq j}^N (\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 = \sum_{i\neq j}^N \pi_j \frac{Y_i^2}{\pi_i} - \sum_{i\neq j}^N Y_i Y_j + \sum_{i\neq j}^N \frac{Y_i Y_j}{\pi_i \pi_j}\pi_{ij} - \sum_{i\neq j}^N \frac{Y_i^2}{\pi_i^2}\pi_{ij}$$

Using 2.7 and comparing with 2.9 we get after some algebra

$$V(Y(x)) = \frac{1}{\lambda^2(F)} \frac{1}{2} \sum_{i \neq j}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2 + \frac{1}{\lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i} \left(E_x\{n(x)|I_i = 1\} - E_x n(x) \right) \tag{5.5}$$

The first term is the Yates-Grundy formula and the second is the correction due to the fact that the number of trees sampled is a random variable. Hence, using exact **PPS** gives zero variance if the sample size is fixed, which is the justification given in standard sampling. Of course, in practice, one can at best have an approximate **PPS** and the variance can be expected to be small. In forest sampling exact **PPS** is only possible for the basal area and using the angle count method; however, in this case, the second term does not vanish. For a global Poisson forest it is intuitively clear that $E_\omega E_x\{n(x)|I_i(x,\omega) = 1\} \approx 1 + E_x n(x)$ so that 5.5 is, under **PPS** essentially the anticipated variance, which is a nice result. A formula similar to 5.5 has been given by Ramakrishnan in 1975 (see [18]).

In practice, as already mentioned, exact **PPS** is usually impossible to implement. For this reason, we shall now present a discrete optimal approximation thereof. The idea stems of course from the widely used concentric circles technique, which now receives, to my knowledge for the first time, a rigorous mathematical justification.

5.3 Discrete approximation of PPS

PPS means that $\pi_i = \alpha Y_i$. By a discrete approximation we mean that the identity function $i : Y_i \mapsto Y_i$ is replaced by a step function $f : Y_i \mapsto f(Y_i)$ taking only finitely many values. For instance $2, 3$, rarely if ever more with concentric circles. We consider the N values Y_i, not all necessarily different (e.g. diameters are rounded to cm), to be partitioned in K intervals C_l, $l = 1, 2 \ldots, K$. N_l is the number of Y_i in C_l. We define the discrete approximation as

$$f(Y_i) = E_Y\{Y_i | Y_i \in C_l\} = \frac{\sum_{Y_i \in C_l} Y_i}{N_l} = \mu_l \tag{5.6}$$

in the sense that if Y_i is in the l-th interval, then the approximation $f(Y_i)$ is the conditional expectation μ_l in this class (E_Y denotes the expectation with respect to the finite population of Y_i values). Hence, we have by definition

$$f(Y_i) \equiv \mu_l \, \forall Y_i \in C_l$$

and by construction

$$\sum_{i=1}^{N} f(Y_i) = \sum_{i=1}^{N} Y_i$$

By Cauchy-Schwartz inequality we get

$$\left(\sum_{i=1}^{N} Y_i \right)^2 = \left(\sum_{i=1}^{N} \frac{Y_i}{\sqrt{f(Y_i)}} \sqrt{f(Y_i)} \right)^2 \leq \sum_{i=1}^{N} \frac{Y_i^2}{f(Y_i)} \sum_{i=1}^{N} f(Y_i) = \sum_{i=1}^{N} \frac{Y_i^2}{f(Y_i)} \sum_{i=1}^{N} Y_i$$

and therefore

$$\sum_{i=1}^{N} \frac{Y_i^2}{f(Y_i)} \geq \sum_{i=1}^{N} Y_i$$

with equality if and only if $f(Y_i) = \alpha Y_i$ which implies $\alpha = 1$ since $\sum_{i=1}^N f(Y_i) = \sum_{i=1}^N Y_i$. Hence, for any discrete approximation $f(Y_i)$ we have

$$\gamma = \frac{\sum_{i=1}^N \frac{Y_i^2}{f(Y_i)}}{\sum_{i=1}^N Y_i} \geq 1$$

the lower bound $\gamma = 1$ being achieved only by the identity $f(Y_i) = Y_i$. The problem is therefore to find, for a given number K of classes, i.e. for K discrete values, the optimal choice of the intervals C_l which minimizes γ. Considering the sum over the Y_i as N times the expectation with respect to the distribution of the Y_i we can write

$$\sum_{i=1}^N \frac{Y_i^2}{f(Y_i)} = N E_Y \frac{Y^2}{f(Y)} = N \sum_{l=1}^K P\{Y \in C_l\} E_Y \{\frac{Y^2}{f(Y)} | Y \in C_l\}$$

Since $f(Y) = \mu_l$ whenever $Y \in C_l$ this is equal to

$$N \sum_{l=1}^K P\{Y \in C_l\} \frac{E_Y\{Y^2 | Y \in C_l\}}{\mu_l}$$

Therefore the intervals must be chosen in order to minimize

$$\sum_{l=1}^K P(Y \in C_l) \frac{E_Y\{Y^2 | Y \in C_l\}}{E_Y\{Y | Y \in C_l\}} \geq E_Y(Y) \tag{5.7}$$

The lower bound being achieved only when the number of intervals is equal to the number of different Y values. The value of the discrete function in a given interval is equal to the conditional expectation of Y in this interval. To do this in practice one needs to have a rough idea of the distribution of the Y_i in the forest. Usually one has a finite number, say roughly 100, of different values, and finding the minimum over $2, 3, 4$ classes can easily be done with a computer by checking all the possibilities. The optimal approximate choice is to set $\pi_i = \alpha f(Y_i)$ where the function $f(Y)$ satisfies the conditions outlined above. The constraint and simple algebra yield the solution

$$\begin{aligned} \lambda(F)\pi_i &= \frac{C - n_2 c_0 - \phi(n_2)}{n_2 c_2} \frac{f(Y_i)}{\bar{Y}} \\ MAV(\widehat{Y}) &= \frac{c_2 \bar{Y}^2 \gamma}{C - n_2 c_0 - \phi(n_2)} + \frac{1}{n_2}\beta^2 \end{aligned} \tag{5.8}$$

For linearized $\phi(n_2)$ one obtains the optimal sample size

$$n_{2,opt} = \frac{\tilde{C}}{\tilde{c}_0 + \sqrt{\tilde{c}_0 c_2 \gamma}\frac{\bar{Y}}{\beta}}$$

and the lower bound for the anticipated variance

$$\min_{n_2} MAV(\widehat{Y}) = \frac{\left(\sqrt{c_2\gamma}\bar{Y} + \sqrt{\tilde{c}_0}\beta\right)^2}{\tilde{C}}$$

As expected the approximate **PPS** has a higher lower bound for the anticipated variance than the exact **PPS**; furthermore, the lower bound is an increasing function of the coefficient γ which, roughly speaking, tells us how far away we are from exact **PPS** (recall that $\gamma = 1$ is the minimum).

It is common practice to think in terms of the relative variance instead of the absolute variance and we therefore define the lower bound, for given n_2, of the relative anticipated variance as $MRAV = \frac{MAV}{\bar{Y}^2}$. Then

$$\min_{n_2} MRAV(\widehat{\bar{Y}}) = \frac{\left(\sqrt{c_2\gamma} + \sqrt{\tilde{c}_0}\Delta\right)^2}{\tilde{C}} \tag{5.9}$$

where we have defined the square of the between strata coefficient of variation as

$$\Delta^2 = \frac{\beta^2}{\bar{Y}^2}$$

Hence, in order to have at least a rough idea of the best achievable accuracy in one-phase one-stage sampling one needs only to know approximately: the between strata variation Δ, the distribution of the response variable (which gives γ) and, obviously, the costs parameters. Let us now generalize this result to one-phase two-stage sampling.

5.4 One-phase two-stage simple random sampling

According to 3.24 and 4.2 we have to minimize

$$E_\omega V_{x|\omega}\widehat{\bar{Y}}^*(\omega) = \frac{1}{n_2\lambda^2(F)}\sum_{i=1}^{N}\frac{Y_i^2 - R_i^2}{\pi_i} + \frac{1}{n_2}\beta^2 + \frac{1}{n_2\lambda^2(F)}\sum_{i=1}^{N}\frac{R_i^2}{\pi_i p_i}$$

under the constraint

$$\phi(n_2) + n_2 c_0 + n_2 c_{21}\sum_{i=1}^{N}\pi_i + n_2 c_{22}\sum_{i=1}^{N}\pi_i p_i = C$$

For given n_2 Cauchy-Schwartz inequality tells us immediately that the π_i must be proportional to $\sqrt{Y_i^2 - R_i^2}$ and that the $\pi_i p_i$ must be proportional to $|R_i|$. We could write down these theoretically optimal solutions but they are not of great practical use. However, they lead us to consider the following prediction model M at the tree level

$$\begin{aligned} Y_i &= Y_i^* + R_i \\ E_M R_i = 0,\ \ V_M R_i &= \sigma_i^2,\ \ COV_M(R_i, R_j) = 0,\ \forall i \neq j \end{aligned} \tag{5.10}$$

In this standard linear model the external prediction Y_i^* is fixed and uncorrelated with R_i so that we have $E_M(Y_i^2 - R_i^2) = (Y_i^*)^2$. **For this reason we shall formally set** $\sqrt{Y_i^2 - R_i^2} = Y_i^*$, which is approximately correct on average. Furthermore, since the σ_i^2 and $\frac{N}{\lambda(F)}$ are bounded we have also

$$\lim_{\lambda(F)\to\infty} E_M\left(\frac{1}{\lambda(F)}\sum_{i=1}^{N}Y_i - \frac{1}{\lambda(F)}\sum_{i=1}^{N}Y_i^*\right)^2 = 0$$

so that we formally also set

$$\frac{1}{\lambda(F)}\sum_{i=1}^{N}Y_i = \frac{1}{\lambda(F)}\sum_{i=1}^{N}Y_i^*$$

Finally, we use the discrete **PPS** approximation for the prediction Y_i^* instead of the true value Y_i and we look for the optimal solution in the class

$$\pi_i = \alpha_1 f(Y_i^*)\ ,\ \pi_i p_i = \alpha_2|R_i|$$

The Lagrange function is then given by

$$L = \frac{1}{n_2\lambda^2(F)}\left(\frac{1}{\alpha_1}\sum_{i=1}^{N}\frac{Y_i^{*2}}{f(Y_i^*)} + \frac{1}{\alpha_2}\sum_{i=1}^{N}|R_i|\right)$$
$$+\lambda\left(\alpha_1 n_2 c_{21}\sum_{i=1}^{N} f(Y_i^*) + \alpha_2 n_2 c_{22}\sum_{i=1}^{N}|R_i| - (C - n_2c_0 - \phi(n_2))\right)$$

Solving the equations for the extremum

$$\frac{\partial L}{\partial \alpha_1} = \frac{\partial L}{\partial \alpha_2} = \frac{\partial L}{\partial \lambda} = 0$$

yields after some algebra the solution

$$\begin{aligned}
\lambda(F)\pi_i &= \frac{C - n_2c_0 - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})}\frac{\sqrt{\gamma}}{n_2\sqrt{c_{21}}}f(Y_i^*)\\
\lambda(F)\pi_i p_i &= \frac{C - n_2c_0 - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})}\frac{1}{n_2\sqrt{c_{22}}}|R_i|\\
\frac{\sum_{i=1}^{N}\pi_i p_i}{\sum_{i=1}^{N}\pi_i} &= \frac{\varepsilon}{\sqrt{\gamma}}\sqrt{\frac{c_{21}}{c_{22}}}\\
MAV(\widehat{Y}^*) &= \frac{\bar{Y}^2\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)^2}{C - n_2c_0 - \phi(n_2)} + \frac{1}{n_2}\beta^2
\end{aligned} \tag{5.11}$$

where we have set

$$\varepsilon = \frac{\sum_{i=1}^{N}|R_i|}{\sum_{i=1}^{N}Y_i^*},\quad \gamma = \frac{\sum_{i=1}^{N}\frac{Y_i^{*2}}{f(Y_i^*)}}{\sum_{i=1}^{N}Y_i^*}$$

ε is the relative prediction error, which can be written as $\varepsilon \approx \frac{\overline{|R|}}{\bar{Y}}$ with $\overline{|R|} = \frac{1}{\lambda(F)}\sum_{i=1}^{N}|R_i|$. Note that the coefficient γ is now defined with respect to the Y_i^*. As expected, the minimum anticipated variance is an increasing function of γ. Formally one can obtain the optimal one-phase one-stage scheme from 5.11 by replacing Y_i^* by Y_i, c_{21} by $c_2 = c_{21} + c_{22}$ and setting $R_i = 0, \varepsilon = 0$ (the result $\pi_i = 0$ simply means that we do not have to take further measurements at the second stage since have the exact value already at the first stage). $\gamma = 1$ corresponds then to exact **PPS** or exact **PPP**. Hence, as announced, the optimal one-phase two-stage sampling scheme draws the first stage trees with a probability proportional to prediction, **PPP**, and the second stage trees with a probability proportional to error, **PPE**. Again the minimum of the anticipated variance is an increasing function of γ.

Linearizing the travelling cost function $\phi(n_2)$ we obtain

$$n_{2,opt} = \frac{\Delta\tilde{C}}{(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})\sqrt{\tilde{c}_0} + \Delta\tilde{c}_0}$$

and the relative anticipated variance at the minimum is then equal to

$$\min_{n_2} MRAV(\widehat{Y}^*) = \frac{\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \sqrt{\tilde{c}_0}\Delta\right)^2}{\tilde{C}} \tag{5.12}$$

The **relative efficiency at the optimum** of the one-phase one-stage to the one-phase two-stage is consequently

$$RE(11|12) = \left(\frac{\sqrt{(c_{21} + c_{22})\gamma} + \Delta\sqrt{\tilde{c}_0}}{(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}) + \Delta\sqrt{\tilde{c}_0}}\right)^2 \geq 1 \tag{5.13}$$

and we obtain the important result that one-phase two-stage sampling is more efficient than one-phase one-stage sampling whenever the following relation holds

$$\sqrt{\frac{c_{21}}{c_{22}}} < \frac{\gamma - \varepsilon^2}{2\varepsilon\sqrt{\gamma}}$$

that is, whenever the relative prediction error ε is small or the extra measurement costs to get the exact value of the response variable are high, which is intuitively reasonable. We consider again the dual optimization problem of minimizing the expected costs for a given anticipated variance W. Using Cauchy-Schwartz inequality in the reverse order and the same techniques as above we obtain finally the dual solution

$$\begin{aligned}
\lambda(F)\pi_i &= \frac{\bar{Y}\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)\sqrt{\gamma}\right)}{n_2 W - \beta^2} \frac{1}{\sqrt{c_{21}}} f(Y_i^*) \\
\lambda(F)\pi_i p_i &= \frac{\bar{Y}\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)}{n_2 W - \beta^2} \frac{1}{\sqrt{c_{22}}} |R_i| \\
\frac{\sum_{i=1}^N \pi_i p_i}{\sum_{i=1}^N \pi_i} &= \frac{\varepsilon}{\sqrt{\gamma}} \sqrt{\frac{c_{21}}{c_{22}}} \\
C(n_2) &= \frac{\bar{Y}^2 \left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)^2}{n_2 W - \beta^2} + n_2 c_0 + \phi(n_2)
\end{aligned} \tag{5.14}$$

Comparing with 5.11 we note that the ratio of the expected number of 2nd to 1st stage trees at the optimum is the same for both problems. The better the prediction model and the more expensive the extra-measurements, the smaller the ratio, which is intuitively clear.

Linearizing the travelling costs and neglecting the intercept term we obtain after some algebra the optimal sample size

$$n_{2,opt} = \frac{1}{W}\left(\frac{\bar{Y}\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)\beta}{\sqrt{\tilde{c}_0}} + \beta^2\right)$$

and the minimum expected costs for the anticipated variance W is

$$\min_{n_2} C(n_2|\widehat{Y}^*) = \bar{Y}^2 \frac{\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \sqrt{\tilde{c}_0}\Delta\right)^2}{W} \tag{5.15}$$

From this and comparing with 5.13 one sees again that the relative efficiency $RE_{11|12}$ is the same whether we look at the ratio of the costs or the ratio of the anticipated variances at the optimum.

In the next section we derive the optimal two-phase two-stage sampling scheme and, as a special case, the optimal two-phase one stage sampling scheme.

5.5 Two-phase one/two-stage simple random sampling

We optimize the two-phase two-stage scheme first. According to 3.27, 4.4 and the approximation $\sqrt{Y_i^2 - R_i^2} = Y_i^*$ we have to minimize

$$E_\omega V_{x|\omega} \widehat{Y}_{reg}^*(\omega) = \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^N \frac{Y_i^{*2}}{\pi_i} + \frac{1}{n_1}\beta^2 + \frac{1}{n_2 \lambda^2(F)} \sum_{i=1}^N \frac{R_i^2}{\pi_i p_i}$$

under the constraint

$$C = n_1 c_1 + \phi(n_2) + n_2 c_0 + n_2 c_{21} \sum_{i=1}^N \pi_i + n_2 c_{22} \sum_{i=1}^N \pi_i p_i$$

For given n_2 and n_1 the solutions is obviously given by 5.11 after replacing the remaining budget $C - n_2c_0 - \phi(n_2)$ by $C - n_2c_0 - n_1c_1 - \phi(n_2)$ and $\frac{1}{n_2}\beta^2$ by $\frac{1}{n_1}\beta^2$ we obtain

$$\begin{aligned}
\lambda(F)\pi_i &= \frac{C - n_1c_1 - n_2c_0 - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})} \frac{\sqrt{\gamma}}{n_2\sqrt{c_{21}}} f(Y_i^*) \\
\lambda(F)\pi_i p_i &= \frac{C - n_1c_1 - n_2c_0 - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})} \frac{1}{n_2\sqrt{c_{22}}} |R_i| \\
\frac{\sum_{i=1}^N \pi_i p_i}{\sum_{i=1}^N \pi_i} &= \frac{\varepsilon}{\sqrt{\gamma}} \sqrt{\frac{c_{21}}{c_{22}}} \\
MAV(\widehat{Y}_{reg}^*) &= \frac{\bar{Y}^2 \left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)^2}{C - n_1c_1 - n_2c_0 - \phi(n_2)} + \frac{1}{n_1}\beta^2
\end{aligned} \tag{5.16}$$

Since $\frac{\partial MAV(\widehat{Y}_{reg}^*)}{\partial n_2} \geq 0$ the number of terrestrial plots n_2 should be kept as small as possible under practical constraints. This is intuitively clear since, as we have seen, two-phase sampling filters out the inhomogeneities of the local Poisson forest. For given n_2 the optimal n_1 is easily found to be

$$n_{1,opt}(n_2) = \frac{\Delta\left(C - n_2c_0 - \phi(n_2)\right)}{\sqrt{c_1}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \Delta\sqrt{c_1})}$$

and the resulting relative anticipated variance is then equal to

$$MRAV(n_2) = \frac{\left(\sqrt{\gamma}\sqrt{c_{21}} + \varepsilon\sqrt{c_{22}} + \Delta\sqrt{c_1}\right)^2}{C - n_2c_0 - \phi(n_2)} \tag{5.17}$$

which, as expected, is an increasing function of n_2. Note that in the above formulae one could replace the between strata variance β^2 by the more general form $V(\widehat{Y}(x))$.

The non-existence of a true minimum is somewhat disturbing even if natural for post-stratified estimation under the local Poisson model. For this reason we give also a slightly different approach. By requiring the expected number of first stage trees to be fixed, i.e. $\sum_{i=1}^N \pi_i = m_1$, we can not only achieve a true minimum but also have a very good idea about the plot size, an important criterion for field work. For technical reasons it is easier to work with the total expected number of second stage trees $M_2 = n_2m_2$, $m_2 = \sum_{i=1}^N \pi_i p_i$. For given n_1 and M_2 the Cauchy-Schwartz inequality yields after some algebra the lower bound of the anticipated variance as

$$\frac{1}{n_2}\gamma\frac{\bar{Y}^2}{m_1} + \frac{1}{M_2}\overline{|R|}^2 + \frac{1}{n_1}\beta^2$$

After linearizing the travelling costs the above expression must be minimized under the constraint

$$n_1c_1 + n_2(\tilde{c}_0 + m_1c_{21}) + M_2c_{22} = C$$

The Lagrange technique yields the optimal solution n_1, n_2 and M_2 which can be

rewritten in terms of n_1, n_2 and m_2

$$
\begin{aligned}
\lambda(F)\pi_i &= m_1 \frac{f(Y_i^*)}{\overline{Y^*}} \\
\lambda(F)\pi_i p_i &= m_2 \frac{|R_i|}{\overline{|R|}} \\
n_{1,opt} &= \frac{\tilde{C}}{\left(\sqrt{\gamma}\sqrt{c_{21}+\frac{\tilde{c}_0}{m_1}}+\varepsilon\sqrt{c_{22}}+\Delta\sqrt{c_1}\right)}\frac{\Delta}{\sqrt{c_1}} \\
n_{2,opt} &= \frac{\tilde{C}}{\left(\sqrt{\gamma}\sqrt{c_{21}+\frac{\tilde{c}_0}{m_1}}+\varepsilon\sqrt{c_{22}}+\Delta\sqrt{c_1}\right)}\frac{\sqrt{\gamma}}{m_1\sqrt{c_{21}+\frac{\tilde{c}_0}{m_1}}} \\
m_2 &= \frac{\tilde{C}}{\left(\sqrt{\gamma}\sqrt{c_{21}+\frac{\tilde{c}_0}{m_1}}+\varepsilon\sqrt{c_{22}}+\Delta\sqrt{c_1}\right)}\frac{\varepsilon}{n_{2,opt}\sqrt{c_{22}}} \\
\left(\frac{n_{2,opt}}{n_{1,opt}}\right)^2 &= \frac{\gamma c_1}{m_1\Delta^2}\frac{1}{\tilde{c}_0+m_1 c_{21}} \\
\min_{n_2|m_1} MRAV(\widehat{Y}^*_{reg}) &= \frac{\left(\sqrt{\gamma}\sqrt{c_{21}+\frac{\tilde{c}_0}{m_1}}+\varepsilon\sqrt{c_{22}}+\Delta\sqrt{c_1}\right)^2}{\tilde{C}}
\end{aligned}
\tag{5.18}
$$

Hence, the optimization of the most general sampling scheme encountered so far is surprisingly simple. The best achievable lower bound of the relative anticipated variance depends on the budget, the various costs, the relative prediction error ε, the between strata coefficient of variation Δ, as well the correction factor γ for discrete **PPP**. It is easily checked that the lower bound $\min_{n_2|m_1} MRAV(\widehat{Y}^*_{reg})$ is a decreasing function of the number m_1 of first stage trees and therefore of the plot size.

Formally one can obtain the optimal one-phase one-stage scheme from 5.18 by replacing Y_i^* by Y_i, c_{21} by $c_2 = c_{21}+c_{22}$ and setting $R_i = 0, \varepsilon = 0$ (the result $\pi_i = 0$ simply means that we do not have to take further measurements at the second stage since have the exact value already at the first stage).

The relative efficiency of one-phase two-stage sampling with respect to two-phase two-stage sampling is, by 5.12 and 5.18 easily found to be

$$
RE(12|22) = \left(\frac{\sqrt{\gamma}\sqrt{c_{21}}+\varepsilon\sqrt{c_{22}}+\Delta\sqrt{\tilde{c}_0}}{\sqrt{\gamma}\sqrt{c_{21}+\frac{\tilde{c}_0}{m_1}}+\varepsilon\sqrt{c_{22}}+\Delta\sqrt{c_1}}\right)^2 \tag{5.19}
$$

Since in general $\tilde{c}_0 >> c_1$ two-phase two-stage sampling will be more efficient than one-phase two-stage sampling, i.e. $RE_{12|22} > 1$. Again, it can be shown that one obtains the same relative efficiency when minimizing the expected costs for a given anticipated variance. The generalization to cluster sampling uses the same algebraic arguments, which allows us to skip the technical details.

5.6 One-phase one/two stage cluster random sampling

We only deal with one-phase two-stage case, since one-phase one-stage can be obtained by setting formally $\varepsilon = 0, Y_i^* = Y_i, R_i = 0, c_{21} \mapsto c_2 = c_{21}+c_{22}$. Using 3.19,

3.23 (together with following remarks), 5.10 and 4.8 we have to minimize

$$E_\omega V_{x|\omega}\widehat{Y}_c^*(\omega) = \frac{1}{n_2 E_x M(x)\lambda^2(F)}\sum_{i=1}^{N}\frac{Y_i^{*2}}{\pi_i} + \frac{1}{n_2 E_x M(x)}(1+\theta)\beta^2 + \frac{1}{n_2 E_x M(x)\lambda^2(F)}\sum_{i=1}^{N}\frac{R_i^2}{\pi_i p_i}$$

under the constraint

$$\phi(n_2) + n_2 E_x M(x) c_{c0} + n_2 E_x M(x) c_{21}\sum_{i=1}^{N}\pi_i + n_2 E_x M(x) c_{22}\sum_{i=1}^{N}\pi_i p_i = C$$

Recall that $E_x M(x) c_{co}$ is the total installation cost of the cluster, including the travelling cost for visiting all the points of the cluster, and that the inflation factor θ, due to the intra-cluster correlation, can be expected to lie between 0 (when the clusters are spread over several different strata) and $E_x M(x) - 1 + \frac{V_x M(x)}{E_x M(x)}$ (when they lie entirely within the same stratum).
For given n_2 we can use 5.11 by replacing formally $n_2 \mapsto n_2 E_x M(x)$ and $\beta^2 \mapsto (1+\theta)\beta^2$ to obtain

$$\begin{aligned}
\lambda(F)\pi_i &= \frac{C - n_2 E_x M(x) c_{c0} - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})}\frac{\sqrt{\gamma}}{n_2 E_x M(x)\sqrt{c_{21}}} f(Y_i^*) \\
\lambda(F)\pi_i p_i &= \frac{C - n_2 E_x M(x) c_{c0} - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})}\frac{1}{n_2 E_x M(x)\sqrt{c_{22}}}|R_i| \\
\frac{\sum_{i=1}^{N}\pi_i p_i}{\sum_{i=1}^{N}\pi_i} &= \frac{\varepsilon}{\sqrt{\gamma}}\sqrt{\frac{c_{21}}{c_{22}}} \\
MAV(\widehat{Y}_c^*) &= \frac{\bar{Y}^2\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)^2}{C - n_2 E_x M(x) c_{c0} - \phi(n_2)} + \frac{1}{n_2 E_x M(x)}(1+\theta)\beta^2
\end{aligned} \tag{5.20}$$

The plot size is therefore inversely proportional to the number of points in the cluster. Linearizing the travelling costs and setting

$$\tilde{c}_{co} = E_x M(x) c_{co} + slope(\phi(n_2))$$

we obtain by 5.12 the optimal sample size

$$n_{2,opt} = \frac{\sqrt{1+\theta}\Delta\tilde{C}}{\sqrt{\tilde{c}_{co}}\left(\sqrt{E_x M(x)}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \sqrt{\tilde{c}_{co}}\sqrt{1+\theta}\Delta\right)}$$

and finally the lower bound of the relative anticipated variance as

$$\min_{n_2} MRAV(\widehat{Y}_c^*) = \frac{\left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \sqrt{c_{c0} + \frac{slope(\phi)}{E_x M(x)}}\sqrt{1+\theta}\Delta\right)^2}{\tilde{C}} \tag{5.21}$$

The **relative efficiency at the optimum** of the one-phase two-stage cluster sampling to the one-phase two-stage simple sampling is consequently given by

$$RE(12, c|12, s) = \left(\frac{\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \sqrt{c_{co} + \frac{slope(\phi)}{E_x M(x)}}\sqrt{1+\theta}\Delta}{\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \sqrt{\tilde{c}_0}\Delta}\right)^2 \tag{5.22}$$

Let us denote by T_c the travelling costs for visiting all points of the cluster, then one has approximately

$$c_{co} \approx c_o + \frac{T_c}{E_x M(x)}$$

In practice one can expect the total installation costs of the cluster, i.e. $E_x M(x) c_{co}$, to be somewhat smaller than $E_x M(x) c_o + T_c$. Hence, in any case, cluster sampling is more efficient than simple random sampling if

$$\left(c_0 + \frac{T_c}{E_x M(x)} + \frac{slope(\phi)}{E_x M(x)} \right) (1+\theta) < (c_o + slope(\phi)) \tag{5.23}$$

In particular, when $\theta = 0$, that is is when the cluster geometry is such that the points of the clusters are likely to be spread over different strata, then cluster sampling is likely to be more efficient than simple random sampling if the travelling costs within the cluster are not to large, i.e. if

$$T_c < (E_x M(x) - 1) slope(\phi)$$

On the other hand, if the points of the cluster tend to all fall into the same stratum then $1+\theta$ will be close to $E_x M(x) + \frac{V_x M(x)}{E_x M(x)}$ and cluster sampling is less efficient than simple random sampling. The intuitive explanation is that in this case it is better to use a single large plot than $E_x M(x)$ smaller plots in a cluster. Of course, in practice, large plots can be awkward to handle because of boundary effects and slope correction; furthermore the theoretical validity assumptions might be violated. Last not least, cluster sampling offers the possibility to perform other investigations, for instance transect sampling of vegetation, when going from one cluster point to the next. In any case, 5.22 allows one to assess the relative merits of both procedures at their optimum. Note that for a given budget, a given cluster sampling scheme may be more or less efficient than a given simple sampling scheme. The above result gives the relative efficiency of the **optimal** sampling schemes. For practical work, it is certainly wise to plot the anticipated variances and the resulting plot sizes of both simple and cluster sampling as a function of the sample size n_2 to draw meaningful conclusions.

5.7 Two-phase one/two stage cluster random sampling

Again we deal only with the two-stage procedure, since the one-stage can be obtained by setting $\varepsilon = 0, Y_i^* = Y_i, R_i = 0, c_{21} \mapsto c_2 = c_{21} + c_{22}$. According to 3.23, 3.29 and 4.8 we have to minimize

$$E_\omega V_{x|\omega} \widehat{Y}^*_{c,reg}(\omega) = \frac{1}{n_2 E_x M(x) \lambda^2(F)} \sum_{i=1}^{N} \frac{Y_i^{*2}}{\pi_i} + \frac{1}{n_1 E_x M(x)} (1+\theta) \beta^2 + \frac{1}{n_2 E_x M(x) \lambda^2(F)} \sum_{i=1}^{N} \frac{R_i^2}{\pi_i p_i}$$

under the constraint

$$C = n_1 E_x M(x) c_{c1} + \phi(n_2) + n_2 E_x M(x) c_{c0} + n_2 E_x M(x) c_{21} \sum_{i=1}^{N} \pi_i + n_2 E_x M(x) c_{22} \sum_{i=1}^{N} \pi_i p_i$$

For given n_1, n_2 we can obviously use 5.20 after subtracting the first-phase costs from the budget and we have the result

$$\begin{aligned}
\lambda(F)\pi_i &= \frac{C - n_1 E_x M(x) c_{c1} - n_2 E_x M(x) c_{c0} - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})} \frac{\sqrt{\gamma}}{n_2 E_x M(x)\sqrt{c_{21}}} f(Y_i^*) \\
\lambda(F)\pi_i p_i &= \frac{C - n_1 E_x M(x) c_{c1} - n_2 E_x M(x) c_{c0} - \phi(n_2)}{\bar{Y}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}})} \frac{1}{n_2 E_x M(x)\sqrt{c_{22}}} |R_i| \\
\frac{\sum_{i=1}^N \pi_i p_i}{\sum_{i=1}^N \pi_i} &= \frac{\varepsilon}{\sqrt{\gamma}} \sqrt{\frac{c_{21}}{c_{22}}}
\end{aligned} \tag{5.24}$$

For given n_1, n_2 the lower bound of the anticipated variance is then given by

$$\begin{aligned}
MAV(\widehat{Y}^*_{c,reg}) = &\frac{\bar{Y}^2 \left(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}}\right)^2}{C - n_1 E_x M(x) c_{c1} - n_2 E_x M(x) c_{c0} - \phi(n_2)} \\
&+ \frac{1}{n_1 E_x M(x)} (1+\theta)\beta^2
\end{aligned} \tag{5.25}$$

As for simple random sampling we see that $\frac{\partial MAV(\widehat{Y}^*_{c,reg})}{\partial n_2} > 0$. Therefore the number n_2 of terrestrial clusters should be kept as small as possible. For given n_2 the optimal n_1 is easily found to be

$$n_{1,opt}(n_2) = \frac{\Delta\sqrt{1+\theta}\,(C - n_2 E_x M(x) c_{c0} - \phi(n_2))}{E_x M(x)\sqrt{c_{c1}}(\sqrt{c_{21}\gamma} + \varepsilon\sqrt{c_{22}} + \Delta\sqrt{1+\theta}\sqrt{c_1})}$$

and the resulting lower bound of the relative anticipated variance is then equal to

$$MRAV(n_2) = \frac{\left(\sqrt{\gamma}\sqrt{c_{21}} + \varepsilon\sqrt{c_{22}} + \sqrt{1+\theta}\sqrt{c_{c1}}\Delta\right)^2}{C - n_2 E_x M(x) c_{c0} - \phi(n_2)} \tag{5.26}$$

which, as expected, is an increasing function of n_2. Comparing 5.17 with 5.26 we see that the relative efficiency of two-phase two-stage simple sampling with respect to two-phase two-stage cluster sampling **for a given total number k of terrestrial plots** is given by the expression

$$RE(22c|22s)(k) = \left(\frac{\sqrt{\gamma}\sqrt{c_{21}} + \varepsilon\sqrt{c_{22}} + \sqrt{1+\theta}\sqrt{c_{c1}}\Delta}{\sqrt{\gamma}\sqrt{c_{21}} + \varepsilon\sqrt{c_{22}} + \sqrt{c_1}\Delta}\right)^2 \frac{C - kc_o - \phi(k)}{C - kc_{co} - \phi(\frac{k}{E_x M(x)})}$$

The first term will be in most instances larger than 1 and the second smaller than 1, so that depending on the particular circumstances and the value of k cluster sampling may be more or less efficient than simple random sampling.

To obtain a true minimum we proceed as for simple random sampling by fixing the expected number of first stage trees per point to be $\sum_{i=1}^N \pi_i = m_1$. Likewise we denote by $m_2 = \sum_{i=1}^N \pi_i p_i$ the expected number of second stage trees per point. By Cauchy-Schwartz we know that we must choose the inclusion probabilities according to

$$\lambda(F)\pi_i = m_1 \frac{f(Y_i^*)}{\bar{Y}}$$

$$\lambda(F)\pi_i p_i = m_2 \frac{|R_i|}{\overline{|R|}}$$

As for simple two-phase sampling we work with $M_2 = n_2 m_2$. Simple calculations show that the anticipated variance can be rewritten as

$$\frac{1}{n_2 E_x M(x)}\gamma\frac{\bar{Y}^2}{m_1} + \frac{1}{E_x M(x) M_2}\overline{|R|}^2 + \frac{1}{n_1 E_x M(x)}(1+\theta)\beta^2$$

After linearizing the travelling costs the above expression must be minimized under the constraint

$$n_1 E_x M(x) c_{c1} + n_2 E_x M(x)\left(c_{c0} + m_1 c_{21} + \frac{slope(\phi(n_2))}{E_x M(x)}\right) + E_x M(x) M_2 c_{22} = C$$

But for the constant $E_x M(x)$, which can be factorised out, and the change $\beta^2 \mapsto (1+\theta)\beta^2$ this is the same problem as in simple random sampling so that we can apply mutatis mutandis the result 5.18 and we write down the characteristics of the optimal two-phase two-stage cluster sampling scheme under the condition that the number of first-stage trees is m_1. They read:

$$\begin{aligned}
\lambda(F)\pi_i &= m_1 \frac{f(Y_i^*)}{\overline{Y^*}} \\
\lambda(F)\pi_i p_i &= m_2 \frac{|R_i|}{\overline{|R|}} \\
n_{1,opt} &= \psi \frac{\sqrt{1+\theta}\Delta}{E_x M(x)\sqrt{c_{c1}}} \\
n_{2,opt} &= \psi \frac{\sqrt{\gamma}}{m_1 E_x M(x)\sqrt{c_{21} + \frac{c_{co}}{m_1} + \frac{slope(\phi)}{m_1 E_x M(x)}}} \\
\left(\frac{n_{2,opt}}{n_{1,opt}}\right)^2 &= \frac{\gamma c_{c1}}{m_1 \Delta^2\left(m_1 c_{21} + c_{co} + \frac{slope(\phi)}{E_x M(x)}\right)} \\
m_2 &= \psi \frac{\varepsilon}{n_{2,opt} E_x M(x)\sqrt{c_{22}}}
\end{aligned} \tag{5.27}$$

where the constant ψ is defined as

$$\psi = \frac{\tilde{C}}{\sqrt{\gamma}\sqrt{c_{21} + \frac{c_{co}}{m_1} + \frac{slope(\phi)}{m_1 E_x M(x)}} + \varepsilon\sqrt{c_{22}} + \sqrt{c_{c1}}\sqrt{1+\theta}\Delta}$$

Finally, the lower bound of the relative anticipated variance is given by

$$MRAV(\widehat{Y}^*_{c,reg}) = \frac{\left(\sqrt{\gamma}\sqrt{c_{21} + \frac{c_{co}}{m_1} + \frac{slope(\phi)}{m_1 E_x M(x)}} + \varepsilon\sqrt{c_{22}} + \sqrt{c_{c1}}\sqrt{1+\theta}\Delta\right)^2}{\tilde{C}} \tag{5.28}$$

which, as expected, is a decreasing function of m_1. Comparing 5.18 and 5.28 we see that, for a given number of first stage trees and at the optimum, the relative efficiency, **for given** m_1, of two-phase two-stage cluster sampling with respect to two-phase two-stage simple random sampling is given by

$$RE(22c|22s) = \left(\frac{\sqrt{\gamma}\sqrt{c_{21} + \frac{c_o}{m_1} + \frac{slope(\phi)+T_c}{m_1 E_x M(x)}} + \varepsilon\sqrt{c_{22}} + \sqrt{c_1}\sqrt{1+\theta}\Delta}{\sqrt{\gamma}\sqrt{c_{21} + \frac{c_o}{m_1} + \frac{slope(\phi)}{m_1}} + \varepsilon\sqrt{c_{22}} + \sqrt{c_1}\Delta}\right)^2 \tag{5.29}$$

where we have assumed that $c_{co} = c_0 + \frac{T_c}{E_x M(x)}$ and $c_{c1} = c_1$; in practice these equalities are likely to be inequalities $<$. Hence, if $RE(22c|22s)(m_1) < 1$ then cluster sampling is certainly more efficient. Again, this is likely to be the case when the intra-cluster correlation and the travelling costs within the cluster are not too large.

Final remarks

All the formulae given so far are purely mathematical results, independent of the numerical values assigned to the occurring constants. Consequently, in any given practical case, one should check their feasibility. For instance, it may theoretically happen that $\pi_i \geq 1$ or $p_i \geq 1$. In such a case, one has to set $\pi_i = 1$ and $p_i = 1$. Likewise one must have $\frac{n_2}{n_1} < 1$ for two-phase sampling to be meaningful, and so on. If the cost parameters are roughly adequate these problems should not occur. It is also important to check the size of the inclusion circles: if the optimal circles are so large that boundary effects cannot be assumed to be negligible, it is wise to increase the sample size n_2 until the validity assumptions are fulfilled. One can reasonably hope that the relative efficiencies given by the anticipated variances will correspond to the relative efficiencies given by the true variances even under moderate departures from the Poisson models. Also, in practice, if the relative efficiency of two sampling schemes is very close to 1, say 1.05 or 0.95 then it is probably wiser to decide in favor of the simpler scheme (the famous "keep it simple" argument!). Last not least, one should perform a sensitivity analysis to judge the impact of the most important cost parameters on the location and the magnitude of the optimum. Ideally, the cost or variance function at the optimum should be flat, which ensures some robustness of the results. With these remarks in mind we are now ready to draw the final conclusions.

Chapter 6

Conclusions

The anticipated variance under the local Poisson model generating the forest strata is a simple tool to optimize forest inventories. The resulting sampling rules have a clear intuitive background and it is difficult to imagine circumstances under which they would be qualitatively wrong. Generally speaking it turns out that two-phase sampling is potentially more efficient than one-phase sampling, that two-stage sampling is superior to one-stage sampling, and that cluster sampling is superior to simple sampling when the intra-cluster correlation and the travelling costs within the cluster are small. Furthermore, the inclusion probabilities of the trees must be proportional to size, **PPS**, either directly in one-stage procedure or by a combination of probability proportional to the predicted size, **PPP**, and then to the expected prediction error, **PPE** in two-stage sampling. An optimal discrete approximation of **PPS** or **PPP** can be given, which depends only on the distribution of the response variable in the tree population and not on the sampling scheme. Furthermore, all probabilities are inversely proportional to the square root of the relevant cost parameters. In two-phase sampling the number of terrestrial plots must be as small and their size as large as possible under practical and, to a lesser degree, mathematical constraints. These rules can be expected to be qualitatively correct under tree location models departing from the local Poisson model, in which the trees are uniformly distributed within strata. Under this model and after linearizing the travelling cost function it is possible to give fairly simple analytical expression for all quantities relevant for choosing the most efficient sampling scheme. The optimization requires, in simple sampling, only the approximate knowledge of the various cost parameters, the between-strata coefficient of variation and the afore mentioned distribution of the response variable; in cluster sampling, one also needs some knowledge of the impact of the relative geometry of cluster and strata on the intra-cluster correlation.

It can be expected that the numerical results are useful guidelines for forest inventories performed on large areas, even under moderate departure from the model. This, however, will be investigated empirically in future work.

Bibliography

[1] D. Basu, 1971: An essay on the logical foundations of survey sampling, Holt, Rinehart and Winston, Toronto, pp. 203-242.

[2] C.M. Cassel, C.E. Saerndal, J.H. Wretman, 1977: Foundations of inference in survey sampling, Wiley, New York.

[3] W.G. Cochran, 1977: Sampling techniques, Wiley, New York.

[4] N. Cressie, 1991: Statistics for spatial data, Wiley, New York.

[5] M. Eriksson, 1995: Compatible and time-additive change component estimators for horizontal-point-sample data, Forest Science, Vo. 41, No 4, pp. 796-822.

[6] M. Eriksson, 1995: Design-based approaches to horizontal-point-sampling, Forest Science, Vo. 41, No 4, pp. 890-907.

[7] V.P. Godambe, 1955: A unified theory of sampling from finite populations, J. Roy. Statist. Soc. Ser. B,17, 268-278.

[8] V.P. Godambe, V.M. Joshi, 1965: Admissibility and Bayes estimates in sampling finite populations, Ann. Math. Statistics, 36.

[9] D.G. Horwitz, D.J. Thompson, 1952: A generalization of sampling without replacement from a finite universe, J. Amer. Statist. Ass., 47, pp 663-685.

[10] D. Mandallaz, 1991: A unified approach to sampling theory for forest inventory based on infinite population models, PhD Thesis, Chair of forest inventory and planning, ETH Zürich.

[11] D. Mandallaz, 1991: Optimization of general double sampling schemes in infinite population models: an application to forest inventory. Proceedings IUFRO S4.11-00, London 1991, University of Greenwich, ed. K. Reynolds, G. Gertner.

[12] D. Mandallaz, 1993: Geostatistical method for double sampling scheme, 1993, Chair of forest inventory and planning, ETH Zürich.

[13] D. Mandallaz, 1995: "Le Hasard fait bien les choses": statistische Methoden für die Waldinventur, Schw. Zeitschrift für das Forstwes., 146, pp. 1015-1032.

[14] B. Matérn, 1986: Spatial variation, 2nd edition, Lecture Notes in Statistics, Springer-Verlag, Berlin.

[15] J. Mecke, R.G. Scheider, D. Stoyan, W.R.R. Weil, 1990: Stochastische Geometrie, Birkhäuser Verlag, Basel.

[16] A. Penttinen, D. Stoyan, H.M. Hentonnen, 1992: Marked point processes in forest statistics, Forest Science, 38, 4, pp. 806-824.

[17] J. Neyman, 1934: On two different aspects of the representative methods: the method of stratified sampling and the method of purposive selection, JRSS, 97, 558-606.

[18] M.K. Ramakrishnan, 1975: A generalization of the Yates-Grundy variance estimator, Sankya, serie C, pp. 204-206.

[19] L. A. Santaló, 1976: Integral Geometry and Geometric Probability, Addison-Wesley Publishing Company, Reading, Massachusset.

[20] H.T. Schreuder, T.G. Gregoire, G.B. Wood, 1993: Sampling methods for multiresource forest inventory, Wiley, New York.

[21] C.E. Särndal, B. Swensson, J.Wretman, 1992: Model assisted survey sampling, Springer series in statistics, New York.

[22] C.-E. Särndal, 1984: Inférence statistique et analyse des données sous des plans d'échantillonnage complexes, Les presses de l'Université de Montréal.

[23] P. Schmid-Haas, 1969: Stichproben am Waldrand, Mitteilungen der Schweizerischen Anstalt für das Forstwesen, 45, pp. 235-303.

[24] J. Serra, 1982: Image Analysis and Mathematical Morphology, Academic Press.

[25] T.F.M. Smith, 1976: The foundations of survey sampling, a review, J.R. Statist. Soc., A, 139, Part 2, 183-203.

[26] P.G. de Vries, 1986: Sampling theory for forest inventory, Springer-Verlag, Berlin.

[27] F. Yates, 1949: Sampling methods for census and surveys, Griffin, London.