Imperial College of Science, Technology & Medicine

Department of Mathematics

EFFICIENCY OF STATISTICS OF STEREOLOGY

by

Alan Stewart Downie

Thesis submitted to the University of London for the degree of Doctor of Philosophy (Ph.D.)

January 1991

ABSTRACT

Stereology is concerned with the estimation of quantitative structural characteristics of physical objects (usually three dimensional) from data obtained in a lower dimension than that of the specimen under study. Classical methods have concentrated largely on design-based estimation, drawing heavily on the results of integral geometry and geometric probability to produce sampling schemes that give rise to unbiased estimators of structural features of a particular specimen, where the unbiasedness is with respect to the randomisation inherent in the sampling scheme. Although the results hold under very general conditions the theoretical sampling schemes often ignore the fact that independent repetitions are not feasible, because specimens cannot be "glued" back together after sectioning, and there has been very little success in assessing the efficiency of such estimators; the first part of this thesis looks at this area, paying particular attention to systematic sampling schemes, which would normally be employed in practice.

More recently model-based approaches to stereology have been developed and a brief review of random set models is presented together with a new estimator for the variance of the classical estimator of volume fraction derived from a systematic lineal analysis. We demonstrate the application of the result in a statistical framework.

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Dr. Rodney Coleman, without whom this thesis would not exist. He introduced me to the subject of stereology and has guided me through my period of research at Imperial College with skill, patience and good humour. I also owe a debt of gratitude to Dr. Richard Hunt, formerly of Exeter University, who initially kindled my interest, as an undergraduate, in Statistics and in research. Whilst at Imperial College I have been lucky enough to have many good friends all of whom I thank for their friendship; the musicians with whom I have spent many (probably *too* many) happy hours playing and in particular my colleagues in the Statistics section, Silvia, Angela, David and Chris, who have played such an important part in my life throughout this time. Thanks are also due to Roger for taking care of the printing and binding and the SERC for providing financial support.

TABLE OF CONTENTS

Abstract			2
Acknowledgements			3
Table of Contents			4
Chapter 1	**Introduction**		**6**
1.1	Sampling Methods		7
	1.1.1	Uniformly Random Sampling	7
	1.1.2	Systematic Sampling	8
	1.1.3	"Three Dimensional" Probes	9
1.2	Modelling		10
Chapter 2	**Estimation of Absolute Volume**		**12**
2.1	Introduction		12
2.2	The General Framework		12
2.3	Fundamental Principles		16
2.4	Restricted Case		23
2.5	Unbiased Sampling Methods for Particles		32
2.6	Extended Case		80
Chapter 3	**Relative Volume Estimation**		**92**
3.1	Introduction		92
3.2	Model Based Inference		92
3.3	General Framework		93
3.4	Boolean Models		99
3.5	The Covariance		105

3.6 Lineal Analysis 109

3.6.1 Covariance Function for the Exponential Model 114

3.6.2 Variance of Lineal Fraction 115

3.6.3 Modelling Intercept Lengths 117

3.6.4 Estimating Parameters 124

3.6.5 Simulated Examples 130

3.6.6 Limitations of the Method 139

3.6.7 Alternative Approaches 142

3.7 Stereology in a Wider Context 145

Chapter 4 Conclusions 149

References 151

Chapter 1 Introduction

Although there are isolated instances of stereological techniques and solutions to problems which have been in existence for as much as two hundred years or more the science of stereology as a field of study in its own right is very young. The essential characteristics of stereological problems are that they involve the extraction and prediction of information (usually quantitative) of a *geometric* nature where the observable quantities are of a lower physical dimension than that of the feature of interest. Consequently, applications which require a stereological treatment arise in an enormous range of disciplines and yet the fundamental principles involved are common to all; thus stereology provides a common ground for a wide variety of fields of study as diverse as metallurgy and neuro-surgery. Underpinning the subject is a mathematical framework, based on the results of integral geometry and geometric probability, and implicit in the nature of the subject is a need for the application of statistical techniques of design and analysis of experiments. However, the geometric aspect of stereology presents many problems not normally encountered in more standard statistical contexts and thus the contribution of statistics to stereology has been severely limited with most results being based on first order properties of estimators and with unbiasedness being used almost exclusively as a criterion for judging the estimators. In this thesis we examine the second order properties of various stereological estimators, some well known and some new, looking all the time to exploit existing techniques and methodologies of standard statistical theory.

1.1 Sampling Methods

Whilst the early history of stereology and geometric probability makes fascinating reading (see, for example Miles, 1987) we need go back no further than 1976 to find the first rigorous mathematical treatment of the subject in a general framework. The work of Miles and Davy (Miles & Davy, 1976, Davy & Miles, 1977) is a thorough treatment of classical stereology for bounded, deterministic specimens from a *design-based* point of view. By design-based we mean that the probabilistic structure of the data obtained from sections through the specimen is taken to be determined completely by the sampling scheme (that is, the method of generating the sections).

1.1.1 Uniformly random sampling

The *Isotropic Uniform Randomness* described by Miles and Davy (see §2.3) has proved to be of fundamental importance, being the invariant measure (under rigid motions) on the co-ordinate space used to describe sections, and it is a reference criterion for other types of randomness used to generate sections. Their work is concerned principally with the unbiased estimation of various structural parameters for arbitrary specimens and in particular demonstrates a theoretically correct approach to ratio estimation under very general conditions. However, because of the complete arbitrariness of the specimen assumed in this work there can be no more than a qualitative discussion of the second order properties of the estimators and their relative efficiencies. Furthermore, the sampling methods described are not realisable in practice for sample sizes greater than one because of the destruction of the specimen necessary to obtain a section. This problem may soon be overcome in some contexts with the development of microscopic techniques that allow focussing *within* a specimen but there is a more fundamental question of whether the generation of independent sections is in fact optimal; we return to this point later.

In order to go further within the design-based framework of Miles and Davy (ignoring the practical problems) we need a fuller specification of the probabilistic

structure of the data obtained from random sections. Since this structure depends on the geometry of the specimen as well as on the sampling scheme it follows that we need to make some assumptions, that is, to do some modelling. Basically, the modelling can be at two levels; either we can make assumptions about the geometry itself, which together with the sampling scheme implies the distribution of the data (even though this may not always be tractable) or we can model the distribution of the data directly; that is, we can assume a parametric form for the distribution of the obsevations that are available and then treat the problem as one of inference about the parameters of that distribution. The classic example of the former is the celebrated "corpuscle problem" of Wicksell (1925, 1926) in which particles contained within a specimen are assumed to be spherical (or ellipsoidal) whereas a (rare) example of the latter is to be found in Cruz Orive (1980) where estimation of volume fraction is treated as a regression problem with non-constant variance. This second example fails to achieve the stated goal of finding a best (ie minimum variance) linear unbiased estimator because unbiasedness is unachievable under the conditions stated, but nevertheless it suggests a new approach to an old problem with considerable appeal from a statistical point of view and perhaps could be developed further using a minimum mean squared error criterion.

1.1.2 Systematic Sampling

The motivation for systematic sampling arises from two sources, namely the impracticality of generating independent IUR sections and the fact that in most cases systematic sampling schemes are more efficient (have smaller variance) than completely random ones. References to systematic sampling can be found very early in the literature (Weibel, 1970) but not until recently has there been work on the variance of such estimators (Mattfeldt, 1987, Gundersen and Jensen, 1987). As is always the case in stereology, assumptions have to be made about the geometry of the specimen, either explicitly or implicitly in terms of the form of the covariance structure, in order to make progress but even so, only in the most exceptional circumstances could systematic sampling be less efficient than independent IUR sampling. One of the main problems

in analysing estimators based on systematic sampling is the high degree of correlation between observations but this can be overcome to an extent by adding some randomness to systematic schemes to produce schemes which will be referred to here as *stratified*. The motivation for the term is that such schemes essentially partition the specimen into *strata* and then sample randomly from within those strata. This type of sampling scheme does not appear in the stereological literature, although Matérn (1960) discusses stratified random sampling applied to area estimation by point counts. The results he presents (Matérn, 1985) for the variance of point count estimators of areas of circles, based on systematic grids, can certainly be improved on by using a stratified scheme, which stabilises and reduces the variance.

1.1.3 "Three Dimensional" Probes

Much interest has centred in the last five years on techniques based on the *disector* (Sterio, 1984), which in some sense can be regarded as a three dimensional probe and which allows the estimation of particle number under very general conditions. Previously this had only been possible under very strict shape assumptions (Wicksell's problem) which were felt to be unsupportable, particularly in biological contexts where structures tend to be highly irregular. In fact the disector still requires assumptions about the geometry of the specimen but of a nature that takes account of the experimenter's knowledge of the specific problem. The disector was originally presented within the framework of random sectioning but here we examine it from a different point of view, in which the randomisation is not induced by the sectioning but by the sampling of units from an arbitrary partition of the specimen. This makes more sense since the object of interest is in fact a finite population of particles and it also makes clear how one would realise the sampling scheme in practice. Furthermore, it allows us to construct different estimators and designs and assess their relative efficiencies.

1.2 Modelling

We have already pointed out the need for modelling in order to assess the performance of estimators but modelling can be an emotive issue in stereology, and indeed, in sampling theory generally (see Royall, 1976, for an interesting discussion). The appeal of classical, design-based methods is that unbiasedness, which is clearly a desirable property of an estimator, is easily proved and arbitrariness of the specimen does not have to be sacrificed. However, not only does this ignore the fact that a biased estimator with a small variance may be preferable to an unbiased one with large variance but also it does not allow for the inclusion of problem-specific knowledge in the design and analysis. Royall gives an excellent example of the way in which a suitable model can be used not only to improve the inference but also to optimise the design of an observational study. Of course the model is specific to the situation and the results are only valid as far as the model is valid but the point being made is that where knowledge exists that is relevant to a particular problem it makes sense to represent it in a model rather than throw it away and rely solely on a random sampling distribution for inference. In stereology the work of Cruz Orive (1980) mentioned earlier is an example of the type of approach described by Royall but it is an isolated case and there would appear to be scope for further work in that field. More common is the adoption of geometric models to describe the specimen; where the geometric model involves a probabilistic structure (as in Wicksell's problem) this can be used to form the basis of inference and design rather than a randomised sampling scheme. The idea of a random set is the basis for this approach and the application of random set theory to stereological problems is an active area of research. In Chapter 3 we look in detail at the application of a random set model to a typical stereological problem and derive new results for the estimation of the variance of the lineal fraction, the natural estimator of volume fraction when performing a lineal analysis on a two phase specimen. The technique is illustrated on some simulated examples and is compared with the known theoretical values for those cases, generally showing good agreement.

Modelling will of course be involved at another level since in any study the interest will ultimately be in a wider population than the observed sample. The objectives of the study will be expressible in terms of inferences or hypotheses for which the data obtained from the specimens will provide the statistics and therefore the way in which a stereological study is carried out should reflect this wider model. Although this has been hinted at (see, for example, Gundersen & Østerby, 1981) there is little reference to this issue in the literature. In §3.7 we conclude the thesis by looking at the way the results of the preceding sections can be fitted into such a wider framework, thus fulfilling the objective of putting stereology (or at least one problem in stereology) into a statistical framework.

Chapter 2 Estimation of Absolute Volume

2.1 Introduction

There are many instances, particularly in biological and materials sciences, where the volume of a certain structure (or *phase*) is of interest, either in absolute value or as a proportion of the *reference space* containing it, and where such measurements cannot be made directly but must be estimated stereologically from lower dimensional information obtained by sectioning and sometimes further application of sampling frames and test sets. The data obtained from the sections and estimators derived from them will have a probabilistic structure determined not only by the geometry of the specimen under scrutiny and the sampling scheme employed but also by the nature of the population of which the specimen is a member. For, whatever the study's aims might be, it is almost certain that the specific specimen being analysed is not of interest purely for its own sake but as part of a sample to make inferences about a wider population.

Ultimately the data are going to be used in a model of some kind in which we wish to fit parameters or test hypotheses, for example as a means of relating structure to function or investigating treatment effect on a diseased organ. Since an exact, complete distributional specification of the model will not be available in any realistic case we must consider which assumptions are necessary, whilst still remaining acceptable, in order to make useful and valid conclusions. This is of course dependent on context and hence we first examine the different situations giving rise to stereological estimation of volume.

2.2 The General Framework

We have noted that in general a specimen is itself not of primary interest but of interest as part of a sample from a population (in the statistical sense of the word). We shall make an important distinction between two types of population, referred to hereafter as Type 1 and Type 2.

Type 1 Each sampling unit is a distinct, isolated entity, being a member of a finite population of such units. For example the population could be all rats, all male rats, all diabetic humans, etc. The structure of interest could be a particular organ or part of an organ within the sampling unit, for instance.

Type 2 Each sampling unit is a small part of a much larger single entity, for example a small block of concrete taken from a batch or from a building. The structure of interest here could be the $Ca(OH)_2$ phase of the material.

There is quite clearly an important difference between these types of population since the former predetermines the sampling units and in general gives rise to variability in a wide range of characteristics whilst the latter allows the experimenter freedom to choose the sampling units himself and control some of the variability.

We see immediately from this distinction that estimation of absolute volume in a sample from a Type 2 population is meaningless, since the volume of the reference space is chosen arbitrarily by the experimenter. Thus any study must be in terms of estimated volume per unit volume of reference phase. However, the situation is not so clear cut for a population of Type 1 and will depend on the context.

Consider a study investigating a relationship between function and structure of, say, an organ of an animal. Here it is quite conceivable that absolute volume is the parameter of interest, although presumably other concomitant variables measured on the animal would need to be taken into account. However, it is equally possible that such a study could be concerned with the *relative* volume of a particular type of structure within the organ, for example as a measure of the efficiency of the organ, in the same sense that the surface area to volume ratio of an animal is the important factor as far as heat loss is concerned, not simply its absolute surface area.

In addition to questions of what exactly we want to estimate we need to consider the arrangement of the phase of interest within the specimen and the scale of magnification at which we need to measure it, thus determining whether we are able to measure throughout the whole phase of interest or just in a small subsampled portion of it. Thus we need to identify the different possible situations that may be encountered in

studies concerned with volume estimation and treat them accordingly.

In this chapter we are interested in the estimation of the absolute volume of a structure (or phase of interest) contained in a sampling unit from a population of Type 1. The phase of interest can be regarded as a compact (not necessarily connected) set in $\mathbb{R}^3$ which we will always denote by Y. It will normally be contained in a reference phase which we also regard as a compact (not necessarily connected) set in $\mathbb{R}^3$; this reference phase may be the sampling unit itself or may simply be one level of a "nested" set of such phases. In a purely theoretical treatment we would not have to be concerned with intermediate phases and magnification but in the real world points have positive area and lines have positive thickness and therefore we are restricted in what we can measure at any given magnification by the resolution of the instruments that we use (including the human eye). This means that we cannot classify problems in terms of rigorous, mathematical criteria but rather in terms of somewhat arbitrary, subjective criteria. When we say that a particular phase cannot be measured with sufficient accuracy at a given magnification we mean that the methods available to us do not allow measurement to be made without an unacceptable experimental error; note that it is experimental error and not sampling error that is under consideration here. It is clear that as the magnification increases the experimental error decreases (although we should be aware of the additional problems that can occur with specimens exhibiting a *fractal* behaviour) but at the same time the size of the field over which we measure decreases. Thus increased accuracy results in a smaller fraction of the object of interest being measured (assuming that finite limits exist on the time and resources available) and so we have to settle for a compromise which will depend very much on specific conditions relating to the particular experiment and experience built up from previous experiments. In this broad classification of problems we are only interested in orders of magnitude of magnification and not in precise numbers; that is, the difference in experimental error between 95× and 100× magnification of an object is likely to be negligible. Our criterion for classifying problems will be whether, given X_j is contained in X_i and M_i is the "optimal" magnification for measuring X_i (that is, the lowest level of

magnification for which experimental error is acceptable) the optimal level of magnification, M_j, for measuring X_j is significantly higher than M_i. By "significantly higher" we understand that under the prevailing experimental constraints the difference between M_i and M_j is such that the total area of the fields in which X_j is measured is a small fraction of the total area of the fields in which X_i is measured, the implication being that small regions of X_i must be sampled from the total amount available for examination at magnification M_i in order to make measurements on X_j. We note in passing that where such sampling is necessary the structure of X_i will have a crucial impact on the way it is performed; this point will recur throughout the discussion.

First we look at the simplest case, where Y can be isolated from the rest of the specimen, thus making certain information available that is useful, such as the caliper diameter; although this case is perhaps unrealistic, and indeed would allow measurement by other methods, we use it to establish some principles. Although we do not wish to restrict Y to be convex it is clear that in order that it can be isolated from the rest of the specimen it must at least be connected and, in some non-technical sense, dense; that is, we would be looking at, for example, an organ rather than a network of capilleries.

The next step is to suppose that the phase of interest, Y, is contained within an opaque reference phase, X, from which it cannot be separated. We take X to be of the type described above in the simplest case whereas Y need not now be so restricted. First we examine the situation where Y can be measured with sufficient accuracy at a low enough magnification that allows measurement over an entire planar section through X. There are many possibilities for the structural form of Y; for example it could be a single compact subset of X, the union of disjoint "particles" or a "tree-like" structure composed of a large number of connected branches. It is possible that there could be an intermediate phase contained in X and containing Y but it would have no relevance in this case since there is no sub-sampling required to measure Y.

Next we consider the case when the magnification required to measure Y

accurately is such that sub-sampling is necessary. Now the possibility of an intermediate phase containing Y is highly significant since the existence of such a phase would enable us to restrict the sub-sampling to be within that phase. Where such a phase exists then its structure and the level of magnification needed to measure it are factors determining further cases of interest.

2.3 Fundamental Principles

We examine first the case where the phase of interest, Y, can be isolated from the sampling unit and we establish some basic principles. We introduce the concept of *Isotropic Uniform Randomness* (*IUR*) as an unambiguous method of defining "random" sections, whose properties make them a natural starting point (analogous to simple random sampling) for the construction of sampling designs. We then consider the implications of practical constraints imposed on sampling schemes and examine the relative efficiency of different designs.

Isotropic Uniform Randomness is the invariant measure, under rigid motions, on the co-ordinate space used to define lines or planes. In the case of planes in $\mathbb{R}^3$ the orientation of the normal corresponds to a point taken uniformly from the surface of a hemisphere and the plane passes through a point taken uniformly along an axis parallel to the normal. More specifically, in our present context we restrict attention to those planes which intersect the specimen and proceed as follows. Let ω represent an orientation in an arbitrary frame of reference in $\mathbb{R}^3$ and let $h_Y(\omega)$ denote the lineal projection of Y onto the direction ω. That is, $h_Y(\omega)$ is the distance between the two tangent planes to Y which have normals with orientation ω. If we take a plane with normal parallel to ω and distance t from the origin of the frame of reference then $\exists$ $a_Y(\omega)$, $b_Y(\omega)$ with $b_Y(\omega)=a_Y(\omega)+h_Y(\omega)$ such that for $t\in (a_Y(\omega),b_Y(\omega))$ the plane intersects Y and not otherwise. We let $T(\omega,t)$ be the plane with normal in the direction of ω and distance t from the origin, and denote by $A(Y\cap T)$ the area of intersection of Y with the plane T. We note that

$$\int_{a_Y(\omega)}^{b_Y(\omega)} A(Y \cap T(\omega,t))dt = V(Y)$$

where $V(.)$ denotes volume and hence for arbitrary, but fixed, ω if we let t be random with the uniform density $(h_Y(\omega))^{-1}$ for $t \in (a_Y(\omega),b_Y(\omega))$ and zero otherwise then the random variable $A(Y \cap T(\omega,t))$ has expectation $V(Y)/h_Y(\omega)$. If ω is also random with probability density $h_Y(\omega)/H(Y)$, where $H(Y)=\int h_Y(\omega)d\omega$ and $d\omega$ is the uniform density on the surface of a unit hemisphere then the random variable $A(Y \cap T(\omega,t))$ has expectation $V(Y)/H(Y)$. The joint density described above may at first appear not to be *IUR* since the marginal distribution of ω is not uniform on the surface of a hemisphere but it is clear that it is exactly the invariant measure described previously restricted to the subset of ω-t space for which the corresponding planes intersect Y and hence the planes so generated are *IUR* planes. The properties of *IUR* planar probes are well known (see, for example, Miles & Davy, 1976) and their generation is straightforward in principle. In practice their obvious drawback is that once cut a specimen cannot be "glued" back together for further independent repetitions. Furthermore $H(Y)$ in general is unknown and would require estimation.

Since the uniform distribution for t described above gives unbiased estimation of $V(Y)/h_Y(\omega)$ by $A(Y \cap T(\omega,t))$ for arbitrary, fixed ω we might consider generating a sequence of independent, identically distributed random variables $\{t_1,...,t_n\}$ with that distribution and forming an estimate

$$\hat{V}_{UR} = \frac{h_Y(\omega)}{n} \sum_{i=1}^{n} A(Y \cap T(\omega,t_i)) \ .$$

In this simplest case there would be no problem in measuring $h_Y(\omega)$ and therefore it can be assumed known. The variance of $\hat{V}_{UR}$ is

$$n^{-1}[h_Y(\omega)]^2 \mathrm{var}(A(Y \cap T(\omega,t)))$$

where $\mathrm{var}(A(Y \cap T(\omega,t)))$ is estimated unbiasedly by

$$\frac{1}{n-1}\left[\sum_{i=1}^{n} A^2(Y\cap T(\omega,t_i)) - \frac{1}{n}\left[\sum_{i=1}^{n} A(Y\cap T(\omega,t_i))\right]^2\right].$$

However, in the same sense that a systematic grid of points is a more efficient estimator of area than a Poisson point process of equal intensity, it seems, intuitively, that if all the sections are parallel it would be more efficient to space them out equally rather than randomly. Let us suppose that the number, n, of parallel sections through Y, is fixed. We may locate the sections by defining

$$t_i = a_Y(\omega) + i\Delta(\omega) \qquad (i=1,\ldots,n)$$

where $\Delta(\omega)=h_Y(\omega)/(n+1)$ or alternatively introduce some randomisation by defining

$$t_i = a_Y(\omega) + (i-1)\Delta^*(\omega) + u \qquad (i=1,\ldots,n)$$

where $\Delta^*(\omega)=h_Y(\omega)/n$ and u is uniformly random on $(0,\Delta^*)$ (see fig. 2.1).

The first scheme lends itself to the techniques of numerical integration, such as Simpson's rule, whose accuracy will depend on the behaviour of $A(Y\cap T(\omega,t))$ as a function of t. Since Simpson's rule is exact for cubic functions we might expect good results for a wide range of shapes. For example, it will be exact for the class of general prismatoids, which includes prismatoids, cylinders, cones, spheres, spheroids and ellipsoids, and therefore could be expected to be good for any shapes which are well approximated by a general prismatoid. Bounds for the error could perhaps be estimated from a measure of the deviation of $A(Y\cap T(\omega,t))$ from cubic form.

The second scheme yields n highly correlated random variables. The obvious estimator,

$$\hat{V}_{SYS} = \Delta^* \sum_{i=1}^{n} A(Y\cap T(\omega,t_i))$$

(which differs from $\hat{V}_{UR}$ in the definition of t_i) is clearly unbiased for $V(Y)$ under the uniform distribution for u.

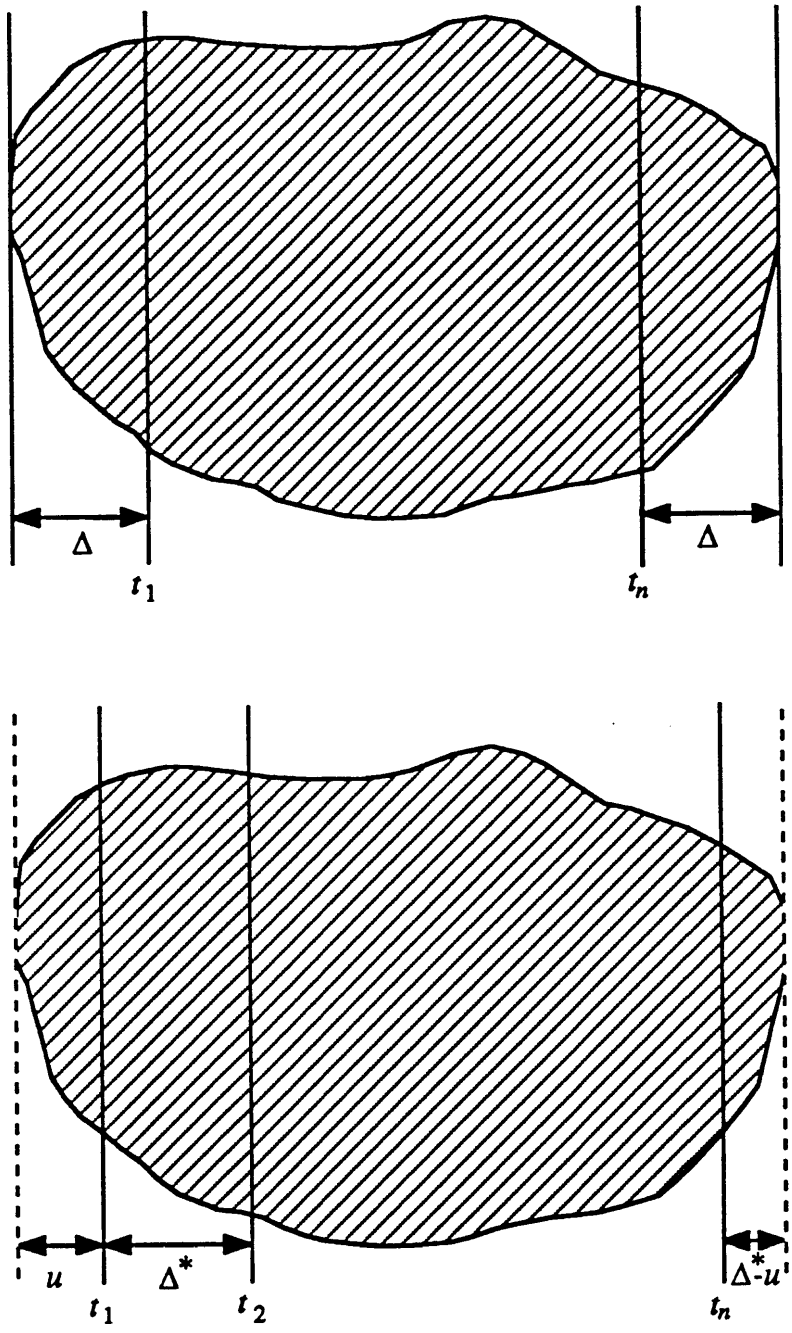


Fig. 2.1 Two Schemes for Systematic Sectioning.

The diagrams illustrate two methods for placing n points systematically in an interval. The first is completely deterministic whereas the second represents a fixed set of points given a random location.

Its variance can be expressed in terms of the covariogram, g, of the function $A(Y \cap T(\omega,t))$ defined by

$$g(y) = \int_{-\infty}^{\infty} A(Y \cap T(\omega,t))A(Y \cap T(\omega,t+y))dt$$

(defining $A(Y \cap T(\omega,t)) = 0$ for $t \notin (a_Y(\omega),b_Y(\omega))$) giving

$$\mathrm{var}(\hat{V}_{SYS}) = \Delta^* \left\{ g(0)+2\sum_{k=1}^{n-1} g(k\Delta^*) \right\} - V^2$$

and

$$V^2 = 2\int_0^{h_Y(\omega)} g(y)dy \ ,$$

where V is the volume of Y. Gundersen & Jensen (1987) give an estimator of this variance based on a quadratic approximation to the covariogram near the origin. The coefficients in the approximation are derived from the sectional areas and have a very high correlation with the volume estimate in the case of a sphere. It is not clear whether this is the case more generally.

An alternative approach, in the spirit of stratified random sampling and which eliminates the correlations in the data, is to define

$$t_i = a_Y(\omega)+(i-1)\Delta^*(\omega)+u_i \qquad (i=1,\ldots,n)$$

with Δ^* as before and $\{u_i : i=1,\ldots,n\}$ independent, identically distributed uniform random variables on $(0,\Delta^*)$. We now have the unbiased estimator

$$\hat{V}_{ST} = \Delta^* \sum_{i=1}^{n} A(Y \cap T(\omega,t_i))$$

(again differing from $\hat{V}_{ST}$ and $\hat{V}_{SYS}$ in the definition of t_i) with variance

$$(\Delta^*)^2 \sum_{i=1}^{n} \mathrm{var}(A(Y \cap T(\omega,t_i))$$

where now the terms in the sum are not identical. The motivation for such an estimator

is that each section is restricted to a slab of the specimen whose thickness diminishes as n increases. Thus each term in the sum will decrease as n increases as opposed to the case when each section is randomly located over the whole specimen, giving a variance which is the sum of n identical terms, independent of n. So we can expect the variance of $\hat{V}_{ST}$ to be at worst $o(n^{-1})$ whereas in general the variance of $\hat{V}_{UR}$ is $O(n^{-1})$. In fact the result will often be much stronger.

To compare the estimators $\hat{V}_{SYS}$ and $\hat{V}_{ST}$ write $f(x)$ for $A(Y \cap T(\omega,x))$, $x_i = a_Y(\omega)+(i-1)\Delta^*$ and f_i, f_i' and f_i'' for $f(x)$ and its first two derivatives evaluated at x_i. Then for both estimators

$$\mathrm{var}(f(t_i)) = \frac{1}{\Delta^*}\int_0^{\Delta^*} [f(x_i+u)]^2 du - \left[\frac{1}{\Delta^*}\int_0^{\Delta^*} f(x_i+u)du\right]^2$$

$$= \int_0^1 \left[f_i + (v\Delta^*)f_i' + \frac{(v\Delta^*)^2}{2}f_i'' + \ldots\right]^2 dv$$

$$- \left[\int_0^1 \{f_i + (v\Delta^*)f_i' + \frac{(v\Delta^*)^2}{2}f_i'' + \ldots\}dv\right]^2$$

$$= \left(f_i^2 + \Delta^* f_i f_i' + \frac{(\Delta^*)^2}{3}((f_i')^2 + f_i f_i'')\right) - \left(f_i + \frac{\Delta^*}{2}f_i' + \frac{(\Delta^*)^2}{6}f_i''\right)^2 + O((\Delta^*)^3)$$

$$= \frac{(\Delta^*)^2}{12}(f_i')^2 + O((\Delta^*)^3) .$$

This gives us

$$\mathrm{var}(\hat{V}_{ST}) = (\Delta^*)^2 \sum_{i=1}^{n} \left\{\frac{(\Delta^*)^2}{12}(f_i')^2 + O((\Delta^*)^3)\right\}$$

whereas the correlations between the terms in $\hat{V}_{SYS}$ give

$$\mathrm{var}(\hat{V}_{SYS}) = (\Delta^*)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \left\{ \frac{(\Delta^*)^2}{12} f_i' f_j' + O\left((\Delta^*)^3\right) \right\}$$

where the extra terms are calculated by a similar expansion to that above. Since $\Delta^*=O(n^{-1})$ an assumption that $f'(x)$ is bounded guarantees $\mathrm{var}(\hat{V}_{ST})=O(n^{-3})$ and $\mathrm{var}(\hat{V}_{SYS})=O(n^{-2})$. However for the particular case $f(x)=\lambda^2-x^2$, corresponding to a triaxial ellipsoid, with $x_i=(-1+2(i-1)/n)\lambda$ $(i=1,...,n)$ and $f'(x)= -2x$ we obtain $\sum\sum f_i' f_j' = 4\lambda^2$ giving $\mathrm{var}(\hat{V}_{SYS})=O(n^{-4})$. In fact in this case it can be shown after lengthy but straightforward calculations that the variance of $\hat{V}_{ST}$ is exactly

$$[V(Y)]^2\{1/n^3 - 4/5n^5\}$$

whereas it has been shown that the variance of $\hat{V}_{SYS}$ is exactly

$$[V(Y)]^2/5n^4$$

(see, for example, Mattfeldt,1987, where this is a special case of a more general result). This appears to be a fortunate consequence of the particular form of $f(x)$ for the triaxial ellipsoid and should not be expected in general. For example, the case where $f(x)=\mu x$ $(x\in[a,b];\ a,b>0)$, also considered by Mattfeldt, would give $\mathrm{var}(\hat{V}_{ST})=O(n^{-3})$ but $\mathrm{var}(\hat{V}_{SYS})=O(n^{-2})$. The important point is that the performance of systematic sampling depends essentially on the stratification of the specimen into strata within which the variance decreases as n increases. However, when sampling is fully systematic the correlations between observations may enhance or reduce the performance of the estimator. Thus systematic sectioning is not *in general* an application of the principle of antithetic variates, as suggested by Mattfeldt, but an application of the principle of stratified random sampling. In the case of the triaxial ellipsoid there is also an effect of antithetic variates, thus enhancing the performance of $\hat{V}_{SYS}$.

2.4 Restricted Case

We now turn our attention to a more realistic case where Y is contained within an opaque reference phase, X, from which it cannot be separated; we assume that Y is still measurable accurately without magnification. This is essentially the restricted case of Miles (1978a) and we use his terminology to emphasise the relationship between our development and the treatment of this class of problems in the literature.

This case is essentially the same as the previous one but with additional variation in the number of sections intersecting Y. (We assume X is "opaque"; if it is not then this case can be treated identically to the previous one). We consider the effect on estimation of $V(Y)$ when applying the techniques of §2.3 to X. We also discuss possible alternatives.

If we take *IUR* sections of X and discard those not intersecting Y the remaining sections are *IUR* sections of Y, irrespective of the arrangement of Y within X. This well known property of *IUR* sectioning is one of the strongest motivations for its prominence in the literature (see, for example, Miles & Davy, 1976, Coleman,1979). Thus any results for the previous case based on *IUR* sections can be carried over to the present situation by applying them to those *IUR* sections of X which intersect Y. Furthermore the estimator $\hat{V}_{UR}$ has an analogue but with the additional problem that $h_Y(\omega)$ is now unknown.

Let $t_1,...,t_N$ be independent and identically distributed random variables with the uniform density $(h_X(\omega))^{-1}$ for $t_i \in (a_X(\omega),b_X(\omega))$ and zero otherwise and let n be the number of sections $T(\omega,t_i)$ of X that intersect Y. Also let $\{s_1,...,s_n\}$ be the subset of $\{t_1,...,t_N\}$ such that

$$Y \cap T(\omega,s_i) \neq \varnothing \qquad i=1,...,n$$

$$Y \cap T(\omega,t) = \varnothing \qquad \forall\ t \in \{t_1,...,t_N\} \backslash \{s_1,...,s_n\}$$

(ie $\{s_1,...,s_n\}$ correspond to the sections that hit Y).

Then $E(A(Y \cap T(\omega,t_i))) = V(Y)/h_X(\omega)$ and the obvious unbiased estimator of $V(Y)$ is

$$\hat{V}_{UR*} = \frac{h_X(\omega)}{N}\sum_{i=1}^{N} A(Y \cap T(\omega,t_i))$$

$$= \frac{h_X(\omega)}{N}\sum_{i=1}^{n} A(Y \cap T(\omega,s_i)) \ .$$

By first conditioning on n and then noting that n has a Binomial$(N,h_Y(\omega)/h_X(\omega))$ distribution it can be shown that the variance of $\hat{V}_{UR*}$ is

$$\frac{1}{N}\left[\{V(Y)\}^2\left(\frac{h_X(\omega)}{h_Y(\omega)}-1\right)+ h_Y(\omega)h_X(\omega)\mathrm{var}(A(Y \cap T(\omega,s_i)))\right] \ .$$

To compare the efficiency of $\hat{V}_{UR}$ with $\hat{V}_{UR*}$ we can look at the ratio of the reciprocals of the variances, giving

$$\frac{\dfrac{h_X(\omega)}{h_Y(\omega)} + \dfrac{[V(Y)]^2\left(\dfrac{h_X(\omega)}{h_Y(\omega)}-1\right)}{}}{[h_Y(\omega)]^2\mathrm{var}(A(Y \cap T(\omega,s)))}$$

as the relative efficiency of $\hat{V}_{UR}$ to $\hat{V}_{UR*}$ where the s in the formula emphasizes that the variance term refers only to those sections which hit Y. The first term in this expression is simply $1/E(n/N)$ and since the second term is always non-negative this expression indicates that the loss in efficiency when Y is contained in an opaque reference phase, X, is greater than the $E(n/N)$, which we might have expected.

No explicit mention has been made of the arrangement and composition of Y; that is, whether Y is a union of disjoint "particles" or whether it is a single, connected domain, and if so whether or not it is convex. These factors are of course included implicitly in the term $\mathrm{var}(A(Y \cap T(\omega,s)))$ but it should be noted that the distribution of n is dependent *only* on $h_Y(\omega)/h_X(\omega)$.

The estimators $\hat{V}_{SYS}$ and $\hat{V}_{ST}$ of §2.3 have their analogues in this case, which we denote by $\hat{V}_{SYS*}$ and $\hat{V}_{ST*}$, but now with $h_Y(\omega)$ unknown and the number of

sections intersecting Y random their analysis becomes much more difficult. In the particular case when Y is a single domain we can show that the variation in the number of sections contributing to $\hat{V}_{SYS*}$ is less than the variation in the number of terms contributing to $\hat{V}_{ST*}$. Following the notation of the previous section let

$$t_i=a_X(\omega)+(i-1)\Delta^*(\omega)+u \qquad (i=1,...,N)$$

$$s_i=a_X(\omega)+(i-1)\Delta^*(\omega)+u_i \qquad (i=1,...,N)$$

with Δ^* now equal to $h_X(\omega)/N$ and $u,u_1,...,u_N$ independent, identically distributed uniform random variables on $(0,\Delta^*)$ and define

$$\hat{V}_{SYS*} = \Delta^*\sum_{i=1}^{N} A(Y\cap T(\omega,t_i))$$

$$\hat{V}_{ST*} = \Delta^*\sum_{i=1}^{N} A(Y\cap T(\omega,s_i)) \quad .$$

Denoting $[a_X(\omega)+(i-1)\Delta^*,a_X(\omega)+i\Delta^*]$ by c_i it is clear (see fig. 2.2) that $\exists l,m\in\{1,...,N\}$ such that

$$T(\omega,t)\cap Y\neq\varnothing \quad \forall\, t\in c_i \qquad i\in\{l+1,...,m\}$$

$$T(\omega,t)\cap Y=\varnothing \quad \forall\, t\in c_i \qquad i\in\{1,...,l\text{-}1,m+2,...,N\}$$

and for $t\in c_l$ and $t\in c_{m+1}$ $T(\omega,t)$ may or may not intersect Y. Denoting by a the Lebesgue measure of the set $\{t:t\in c_l, T(\omega,t)\cap Y\neq\varnothing\}$ and by b the Lebesgue measure of the set $\{t:t\in c_{m+1}, T(\omega,t)\cap Y\neq\varnothing\}$ we can write that

$$N_1=k+I_1$$

$$N_2=k+I_2+I_3$$

where N_1 and N_2 are the numbers of sections making contributions to $\hat{V}_{SYS*}$ and $\hat{V}_{ST*}$ respectively, $k=m-l$ and

$$I_1 = \begin{cases} 0 & \text{with probability } 1-(a+b)/\Delta^* \\ 1 & \text{with probability } (a+b)/\Delta^* \end{cases} \Big\} \; a+b<\Delta^*$$

$$\begin{cases} 1 & \text{with probability } 2-(a+b)/\Delta^* \\ 2 & \text{with probability } (a+b)/\Delta^* -1 \end{cases} \Big\} \; a+b>\Delta^*$$

$$I_2 = \begin{cases} 0 & \text{with probability } 1\text{-}a/\Delta^* \\ 1 & \text{with probability } a/\Delta^* \end{cases}$$

$$I_3 = \begin{cases} 0 & \text{with probability } 1\text{-}b/\Delta^* \\ 1 & \text{with probability } b/\Delta^* \end{cases}$$

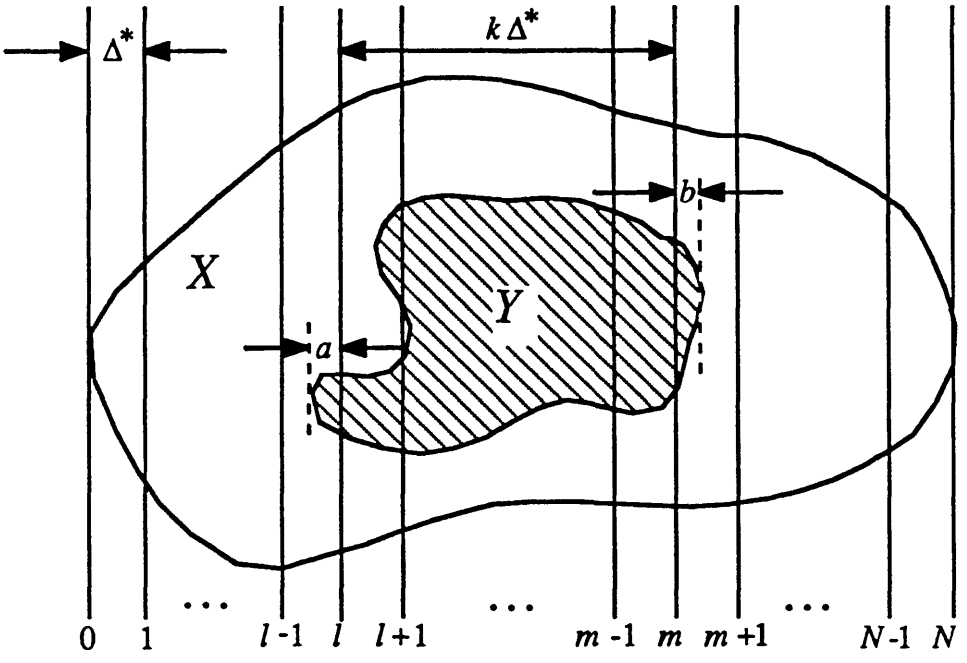


Fig. 2.2 Systematic Sections for an Embedded Specimen.

When systematic sections are taken through X the numbers that hit Y will be random. From the diagram it is clear that the difference between the maximum and minimum number of sections hitting Y is two when Y is connected. If the sections in each stratum are located independently then the variance of the number of sections hitting Y will be greater than if the sections are completely systematic since in the latter case the contributions from the "critical" intervals will be negatively correlated.

It is not difficult to show that

$$\mathrm{var}(N_2) = (a+b)/\Delta^* - (a^2+b^2)/(\Delta^*)^2$$

$$\mathrm{var}(N_1) = \begin{cases} [(a+b)/\Delta^*][1-(a+b)/\Delta^*] & a+b<\Delta^* \\ [(a+b)/\Delta^*-2][1-(a+b)/\Delta^*] & a+b>\Delta^* \end{cases}$$

and hence for $a+b<\Delta^*$ $\mathrm{var}(N_1)$ is clearly less than $\mathrm{var}(N_2)$ whereas for $a+b>\Delta^*$ $\mathrm{var}(N_1)<\mathrm{var}(N_2)$ whenever $\Delta^*-(a+b)<ab/\Delta^*$; but this will always be true since $\Delta^*-(a+b)$ is always negative and ab/Δ^* is always positive.

It seems probable that this behaviour will carry over to the case where the projection of Y onto ω is composed of several disjoint intervals. However, what is far more difficult to analyse in general is the variance of the areas of intersection. We conclude this section with an example of the use of systematic and stratified designs for estimating area in a number of synthetic specimens.

Example Let O be the origin of an arbitrary Cartesian frame in $\mathbb{R}^2$ and let X be a disc of radius R with its centre at O. Let Y be a subset of X composed of the disjoint union of several smaller discs (the different specimens are shown in fig. 2.3). We let $T(\theta,t)$ be the line making an angle θ with the horizontal axis and with distance t from the origin. The set $\{(\theta,t)\colon 0\le\theta<\pi;\ -R\le t\le R\}$ defines all lines hitting X but because of the symmetry of the specimens we need only consider $0\le\theta<\pi/2$. For a given N we let $\Delta^*=2R/N$ and define

$$t_i=-R+(i-1)\Delta^*+u \qquad (i=1,\ldots,N)$$

$$s_i=-R+(i-1)\Delta^*+u_i \qquad (i=1,\ldots,N)$$

where $u,u_1,\ldots,u_N$ are independent, identically distributed uniform random variables on $(0,\Delta^*)$.

Then let

$$\hat{A}_{SYS*} = \Delta^* \sum_{i=1}^{N} L(Y \cap T(\theta, t_i))$$

$$\hat{A}_{ST*} = \Delta^* \sum_{I=1}^{N} L(Y \cap T(\theta, s_i))$$

be two unbiased estimators of the area of Y (analogous to $\hat{V}_{SYS*}$ and $\hat{V}_{ST*}$) where $L(.)$ denotes length. The standard deviations of these estimators for some different values of N and θ are shown in fig. 2.4.

The most striking point to notice from the graphs is that in all cases the systematic estimator is considerably better than the corresponding stratified estimator. The specimens were chosen to cover a wide range of situations, both in terms of heterogeneity and in terms of relative area occupied by the phase of interest, Y, but the relative performance of the estimators seems to be largely independent of these factors. Given also that the systematic estimator will generally be easier to construct in practice (we can use the same grid every time) it does seem that in this type of problem the systematic estimator is preferable to the stratified one. We have, of course, used very regularly shaped specimens in order to make the calculations feasible and we would like to be able to obtain more generally applicable results before drawing too many conclusions. A secondary point of interest is the difference in the performance of the estimators for different orientations of the line probes. One of the concerns when using systematic estimators is the possibility of the results being seriously affected by periodicities in the specimen, and taking all our line probes in a single direction does suggest the possibility of being vulnerable to that sort of error, albeit with a small probability. There are various modifications that we might consider, such as taking two sets of probes at right angles to each other, or two sets of probes with different spacings. Clearly there are many unanswered (and unasked) questions related to systematic sectioning and there is much work still to be done, particulary for arbitrarily shaped specimens.

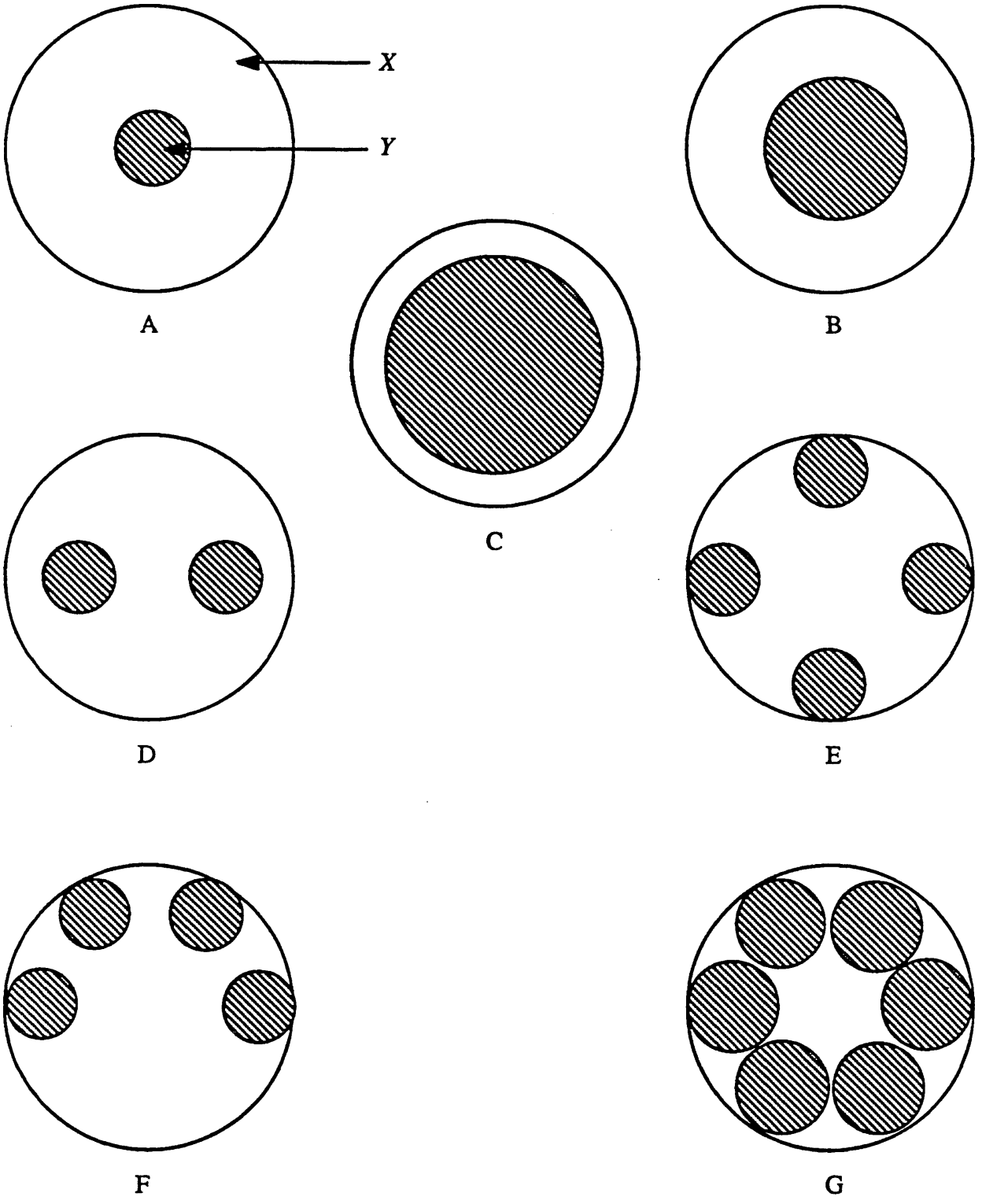
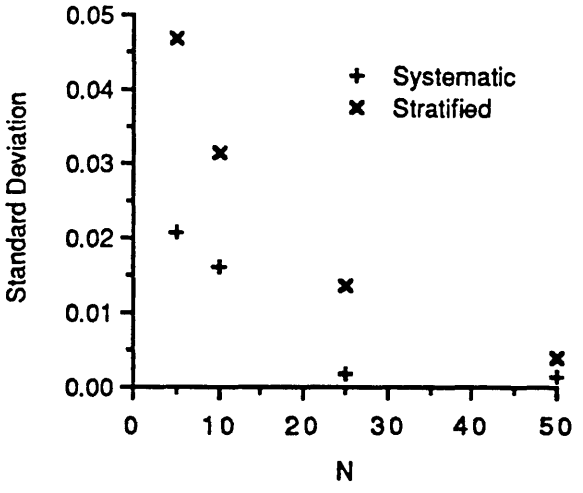
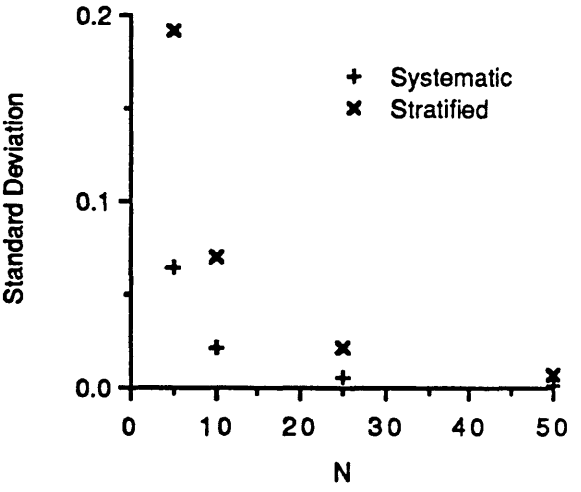


Fig. 2.3 Synthetic specimens used to examine systematic and stratified estimators

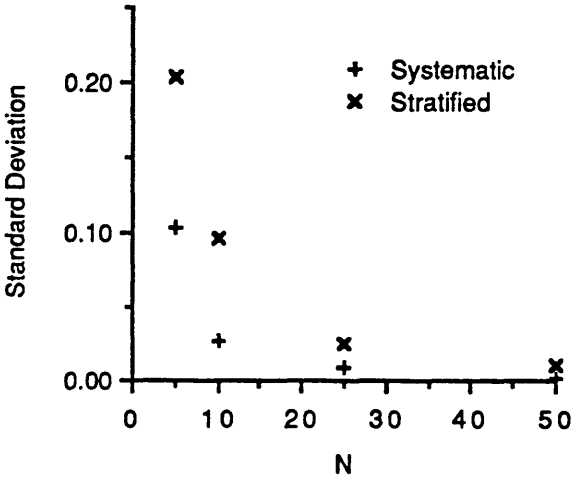
In each case the reference phase, X, is a disc of radius 1 unit. The arrangements of Y were chosen to cover a wide range of situations as far as heterogeneity and area coverage were concerned.



(a) Specimen A

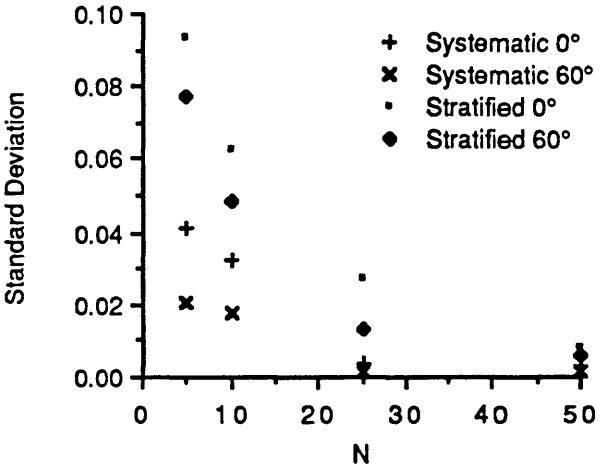


(b) Specimen B

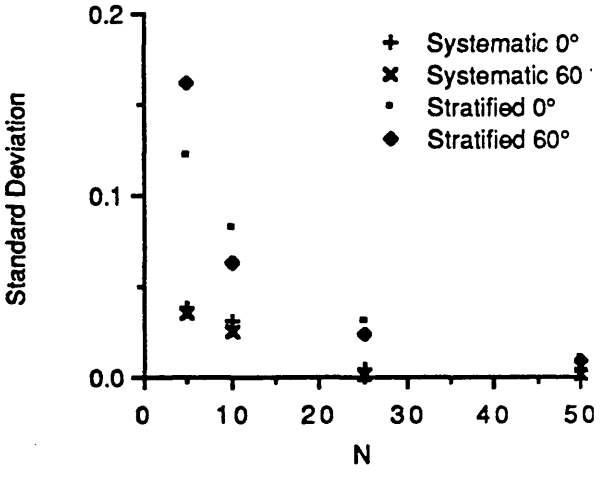


(c) Specimen C

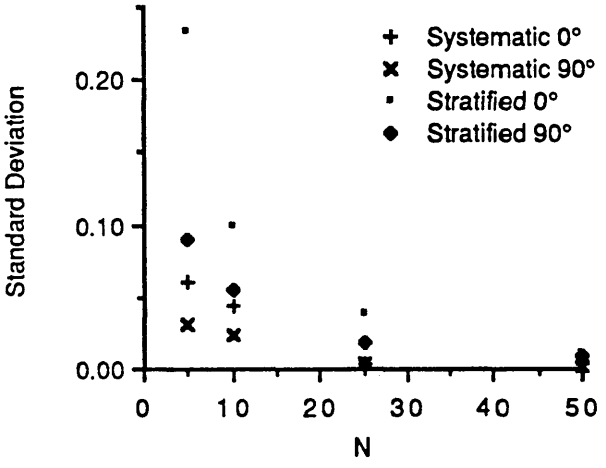
Fig. 2.4 Standard Deviation of Systematic and Stratified Estimators of Area for Synthetic Specimens



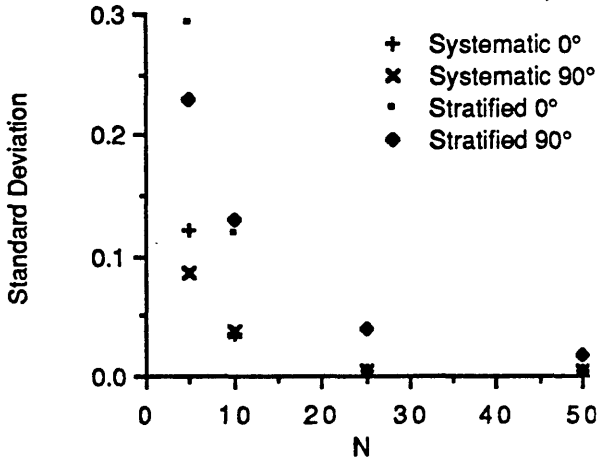
(d) Specimen D



(e) Specimen E



(f) Specimen F



(g) Specimen G

Fig. 2.4 (cont.) Standard Deviation of Systematic and Stratified Estimators of Area for Synthetic Specimens

The graphs show the standard deviations of systematic and stratified estimators of area, based on sets of parallel line probes, for the specimens shown in fig. 2.3. The numbers of probes used are shown on the horizontal axis and their inclination to the horizontal is indicated in the legend. For specimens A, B and C only horizontal line probes were considered as those specimens have rotational symmetry of infinite order. For the other specimens the results shown are for the orientations that gave the most different results of those obtained.

2.5 Unbiased Sampling Methods for Particles

An alternative approach when Y is composed of a large number of disjoint "particles" is to employ recently developed methods such as the *disector* of Sterio (1984) and Cruz-Orive's *selector* (1987), designed to eliminate biases inherent in classical stereological sampling procedures. The achievement of the disector is in being able to estimate number, a zero dimensional quantity, stereologically, without recourse to rigid assumptions such as those of Wicksell's problem. The significance of this is recognised by referring to Fig. 2.5; one of the fundamental relationships in stereology is that the expected value of a measurement of dimension m on a probe of dimension p through a specimen of dimension s is proportional to a characteristic of the specimen of dimension s-(p-m). For example, areas of intersection on planar probes estimate the volume of a three dimensional specimen (s=3, p=2, m=2) and the number of intercepts of a line probe with the boundary of a two dimensional specimen estimates its boundary length (s=2, p=1, m=0). It is quickly seen that to estimate number, a zero dimensional quantity, requires either measurements of dimension -1 or probes of dimension 3; the disector is essentially a three dimensional probe, enabling particles to be sampled with equal probability and hence making unbiased estimation of particle number possible.

The use of the disector and related techniques in the present context amounts to regarding the problem of volume estimation as one of estimation of total number and mean volume in a finite population where the size of the population is unknown, the sample size is random and the measurements are made with errors with unknown distribution. The fundamental principles involved, apart from that of the disector, are the formulae of integral geometry, particularly that due to Crofton (1885), relating powers of lengths of random chords of bodies to global characteristics such as volume and surface area and the theory of finite sample surveys.

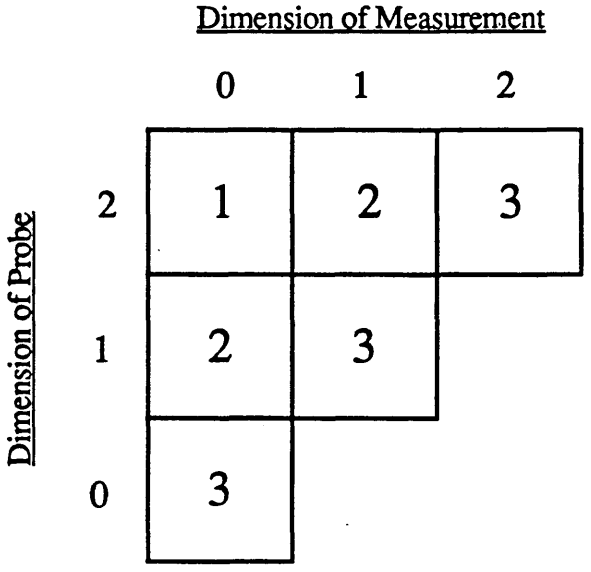


Fig. 2.5 Relationship Between Dimensions of Probes, Measurements and Estimated Characteristics

The figure shows the dimension of the characteristic of a three dimensional specimen which is estimated unbiasedly by measurements on probes of the dimensions stated.

We have a population, Y, of disjoint particles, $Y_1,...,Y_Q$, contained in a reference phase, X, and we wish to make inference about population parameters from measurements made on a sample of particles. In the classical theory of sampling from a finite population (see Cochran, 1977) we would proceed by basing our inference on a *simple random sample* from the population; that is, if our sample were to contain n units then the sampling scheme would ensure that every subset of the population of size n had an equal probability of being chosen. This is a very strong condition in the present stereological context, where in general the population size is unknown and the sample size cannot be fixed *a priori*, and it seems likely that strict simple random sampling is not possible under the most general assumptions. We can, however, use the disector to satisfy the weaker condition that every particle is included in the sample with equal probability.

First we examine the principles underlying the disector in some detail and provide a rigorous proof of its sampling properties. We take a different approach from that in Sterio's original discussion, starting with a finite partition of the reference phase and an association rule between the particles and the elements of the partition. We then regard the elements of the partition as a finite population of disectors from which we take a sample. The treatment is more general in that the disectors can be quite arbitrary in shape, and more useful in practice because the application of finite population sampling theory is immediate. In Sterio's original paper it is not clear how repeated sampling should be achieved in practice; the essential difference here is to work with a finite partition and finite population theory. This then allows an examination of different sampling schemes and estimators and some assessment of their relative efficiencies.

Having examined the question of estimating particle number in some detail we then go on to look at the second aspect of the volume estimation problem, namely estimation of mean particle volume, paying particular attention to the selector of Cruz Orive (1987) based on the principle of *vertical sections* (Baddeley, Gundersen and Cruz Orive, 1986), and we conclude with some comments on ratio estimators.

We start from the following principle:

Let $\mathcal{S}$ be a partition of the reference phase by a collection of sets, $S_1,...,S_M$, satisfying

$$X \subset \bigcup_{s=1}^{M} S_s$$

$$\vdots$$

$$S_s \cap S_{s'} = \varnothing \qquad (s' \neq s).$$

Then if the probability that S_s is sampled is a constant, π, for all s and if for every y there exists a unique s such that

$$\Pr(Y_y \text{ is sampled} \mid S_s \text{ is sampled}) = 1$$

and

$$\Pr(Y_y \text{ is sampled} \mid S_s \text{ is not sampled}) = 0$$

we have that the probability of sampling particle Y_y equals π for all y.

The contribution of Sterio (1984) is to provide a rule which ensures that these requirements are met. In order to apply this rule we construct $\hat{\delta}$ in the following manner:

1 For an arbitrary orientation, ω, in $\mathbb{R}^3$ choose $t_1<t_2<\ldots<t_{\alpha-1}<t_\alpha$ arbitrarily in the interval $(a_X(\omega),b_X(\omega))$ subject to the constraint that $\max\{t_1-a_X(\omega),\ t_2-t_1,\ldots,t_\alpha-t_{\alpha-1},\ b_X(\omega)-t_\alpha\}$ is less than the minimum caliper diameter of any particle. This enables us to define a set of parallel planes intersecting X.

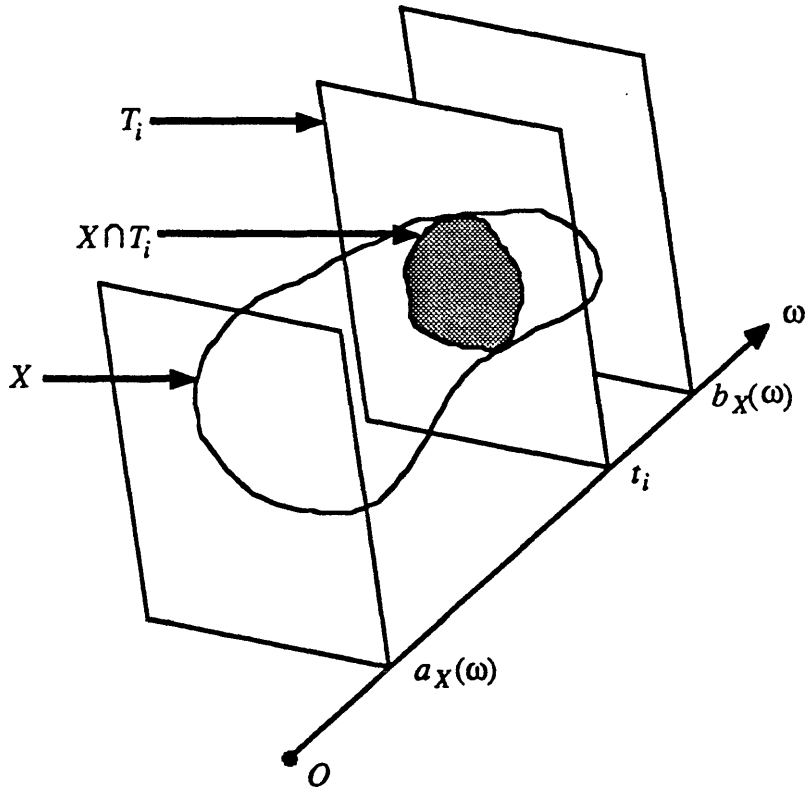


Fig. 2.6 Construction of Partition - First Stage.

A set of α parallel planes is defined by choosing α points arbitrarily between the two tangent planes to X subject to the condition that the distance between adjacent planes is always less than the minimum caliper diameter of any particle. This ensures that every particle must be hit by at least one plane.

2 For each $i \in \{1,...,\alpha\}$ choose $l_{i1}<l_{i2}<...<l_{i\beta_i-2}<l_{i\beta_i-1}$ arbitrarily in the interval $(a_{T_i}(\theta_i),b_{T_i}(\theta_i))$, where $T_i=T(\omega,t_i)$, θ_i is an arbitrary orientation in $\mathbb{R}^2$ and $a_{T_i}(\theta_i)$, $b_{T_i}(\theta_i)$ are the distances of the tangents to $X \cap T_i$ with normal θ_i from an arbitrary origin in T_i. Thus on each plane, T_i, we can define a set of parallel lines, $L(\theta_i,a_{T_i}(\theta_i))$, $L(\theta_i,l_{i1}),...,L(\theta_i,l_{i\beta_i-1})$, $L(\theta_i,b_{T_i}(\theta_i))$, which we denote by $L_{i0},...,L_{i\beta_i}$ respectively, and hence $X \cap T_i$ is divided into β_i parallel strips. On each of these lines we take the direction $\theta_i+\pi/2$ to be positive and $\theta_i-\pi/2$ to be negative.

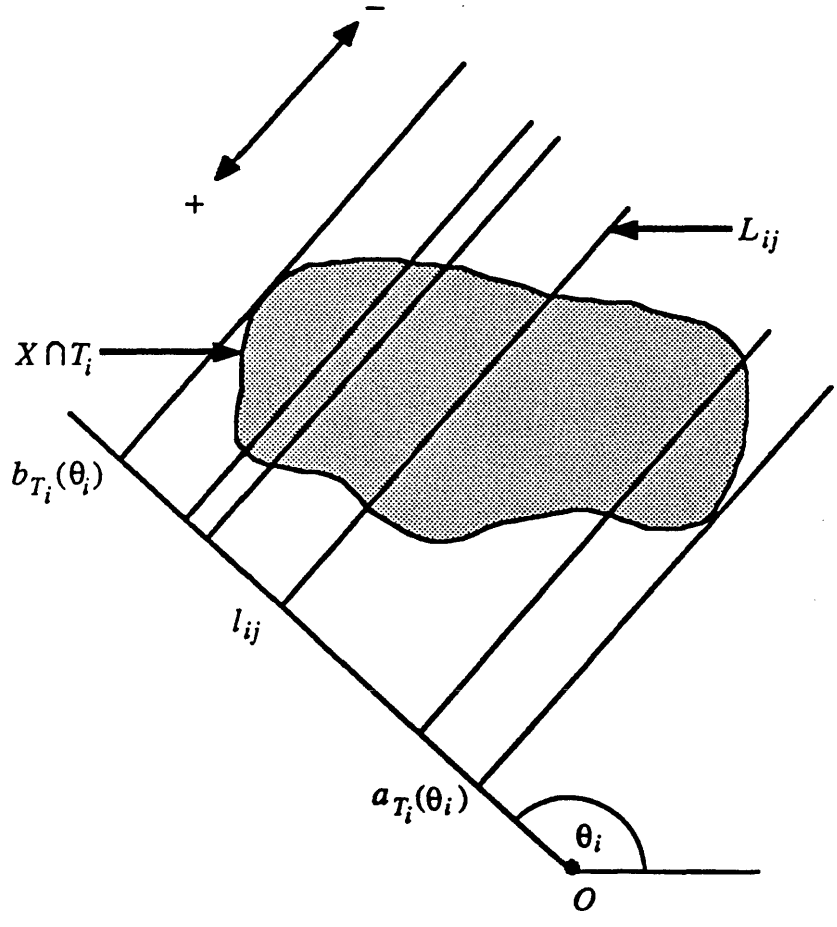


Fig. 2.7 Construction of Partition - Second Stage.

The profile of intersection of X with a planar section is divided into strips in an analogous way to the construction of the planar sections themselves. This time there is no restriction on the distance between the lines because the space inbetween the lines is visible and no particle profiles can be "lost".

3 Each strip on each plane is divided arbitrarily into γ_{ij} rectangles, $D_{ij1},...,D_{ij\gamma_{ij}}$, where L_{ij-1} and L_{ij} form one pair of sides of D_{ijk} and the other pair of sides are denoted B_{ijk-1} and B_{ijk}.

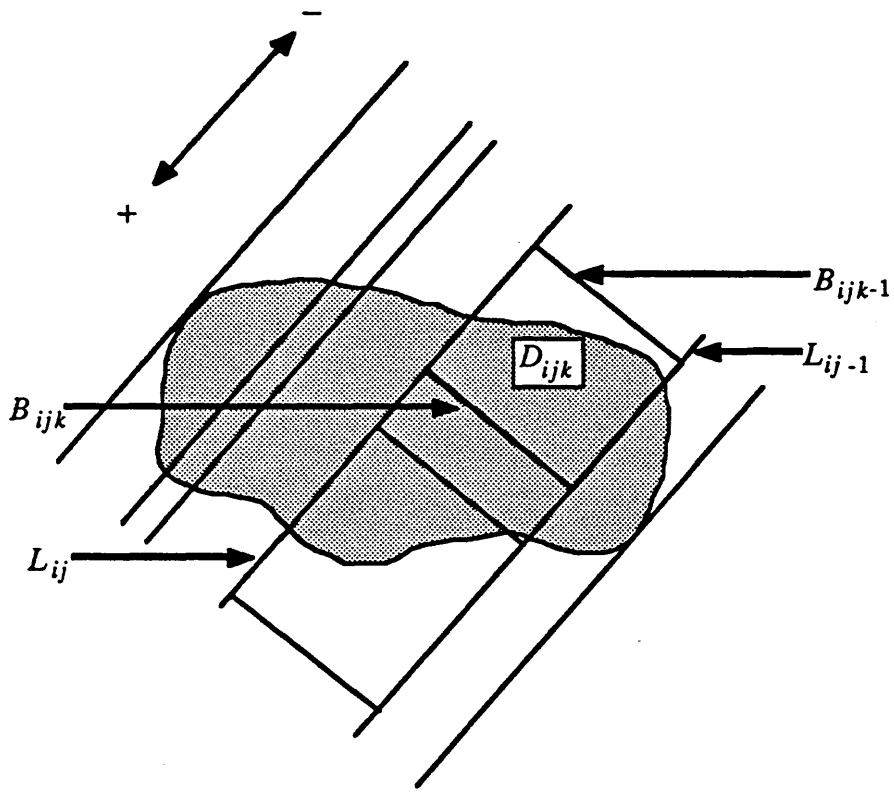


Fig. 2.8 Construction of Partition - Third Stage.

The final stage in the process is to divide each strip on each plane into rectangles. We still need to define a rule such that every particle can be unambiguously associated with a unique rectangle on a unique planar section, preferably without having to construct the whole partition.

We have now established a framework for the sampling of particles with equal probabilities. Rather than a partition by subsets, $S_1,...,S_M$, we have an index set $\Xi=\{(i,j,k)\colon i=1,...,\alpha;\ j=1,...,\beta_i;\ k=1,...,\gamma_{ij}\}$. However the same principle still applies; that is, we require that each $\xi \in \Xi$ is sampled with constant probability, π, and that for each y there exists a unique $\xi \in \Xi$ satisfying

$$\Pr(Y_y \text{ is sampled} \mid \xi \text{ is sampled}) = 1$$

and

$$\Pr(Y_y \text{ is sampled} \mid \xi \text{ is not sampled}) = 0.$$

For a particle,Y_y, contained in X let Λ_y be the subset of $\{1,...,\alpha\}$ such that $i \in \Lambda_y \Leftrightarrow Y_y \cap T_i \neq \varnothing$ and let $i^*=i^*(y)$ be the largest member of Λ_y. (We note that $\Lambda_y \neq \varnothing$ because of the constraints on $t_1,...,t_\alpha$). Also, for $i=i^*(y)$, let $j^*=j^*(y,i^*)$ be the largest j such that $\exists$ a k for which $D_{i^*jk} \cap (Y_y \cap T_{i^*}) \neq \varnothing$ and let $k^*=k^*(y,i^*,j^*)$ be the largest k such that $D_{i^*j^*k} \cap (Y_y \cap T_{i^*}) \neq \varnothing$. Defining

$$I(y, \xi) = \begin{cases} 1 & i=i^*(y), j=j^*(y, i^*), k=k^*(y, i^*, j^*) \\ 0 & \text{otherwise} \end{cases}$$

we include Y_y in the sample if $I(y, \xi) = 1$ and not otherwise. This satisfies the necessary conditions for unbiased sampling.

We could now proceed to construct a sampling design on the population of particles by considering specific partitions of the type described above together with a sampling distribution on the corresponding index set, Ξ. However, first we state two results which enable us to evaluate $I(y,\xi)$ for a particular ξ without having to realise the entire partition in practice.

We introduce the notation $L^+_{ij,k}$ and $L^-_{ij,k}$ for the semi-infinite lines which are co-incident with L_{ij}, have their end points at the intersection of L_{ij} and B_{ijk} and extend infinitely in the positive and negative directions respectively. E_{ijk} is the infinite line formed by the union of $L^-_{ij,k}$, B_{ijk} and $L^+_{ij-1,k}$ (see Fig. 2.9).

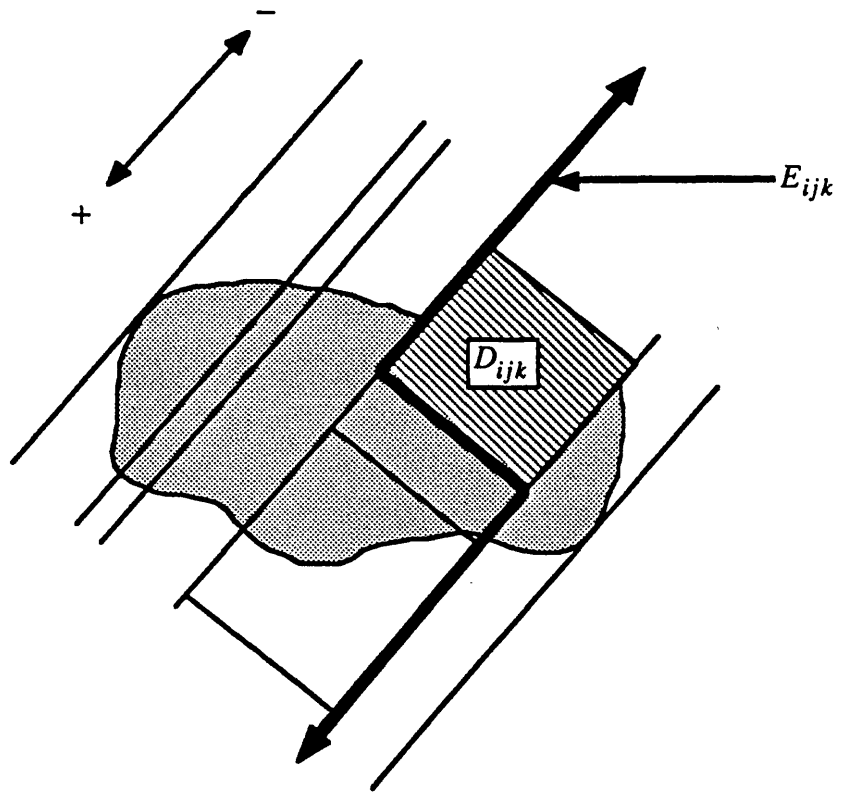


Fig. 2.9 Infinite Exclusion Line for Unbiased 2-D Counting Rule.

The diagram illustrates the unbiased 2-D counting rule of Gundersen(1977). A convex planar subset of $X \cap T_i$ is associated with the rectangle for which it intersects D_{ijk} but not E_{ijk}. See the text for the proof that with this rule every such planar set is associated with exactly one rectangle.

Proposition Let Y_y be a convex particle and, for $i \in \Lambda_y$, let $C = Y_y \cap T_i$. Then $j=j^*$ *and* $k=k^* \Leftrightarrow D_{ijk} \cap C \neq \emptyset$ *and* $E_{ijk} \cap C = \emptyset$.

Proof If $k^* = \gamma_{ij^*}$ then $B_{ij^*k^*} \cap X = \emptyset$ and $L^+_{ij^*-1,k^*} \cap X = \emptyset$. If $k^* \neq \gamma_{ij^*}$ we must have $B_{ij^*k^*} \cap C = \emptyset$ and $L^+_{ij^*-1,k^*} \cap C = \emptyset$ otherwise there would exist a $k' > k^*$ such that $D_{ij^*k'} \cap C \neq \emptyset$. In both cases we have $L^-_{ij^*,k^*} \cap C = \emptyset$ by the definition of j^* and hence $E_{ij^*k^*} \cap C = \emptyset$. Also $D_{ij^*k^*} \cap C \neq \emptyset$ by the definition of j^* and k^*.

Now we show that $D_{ijk} \cap C \neq \emptyset$ and $E_{ijk} \cap C = \emptyset$ *only* holds when $j=j^*$ and $k=k^*$. By the definition of k^*, $D_{ij^*k'} \cap C = \emptyset$ for $k' > k^*$. Suppose $\exists\ k' < k^*$ with $D_{ij^*k'} \cap C \neq \emptyset$. Since C is connected and $D_{ij^*k^*} \cap C \neq \emptyset$ we must have that

either $B_{ij^*k'} \cap C \neq \emptyset$ or $L^+_{ij^*-1,k'} \cap C \neq \emptyset$ and hence $E_{ij^*k'} \cap C \neq \emptyset$. It remains to show that $D_{ij'k} \cap C = \emptyset$ or $E_{ij'k} \cap C \neq \emptyset \ \forall \ j' \neq j^*$. By the definition of j^* we know that $D_{ij'k} \cap C = \emptyset \ \forall \ j' > j^*$. Now consider any j', k with $j' < j^*$. Since $D_{ij'k}$ and $D_{ij^*k^*}$ are separated by $E_{ij'k}$ and C is connected it follows that either $D_{ij'k} \cap C = \emptyset$ or $E_{ij'k} \cap C \neq \emptyset$.

Proposition Let Y_y be a particle contained in X and let $i^*(y)$ be the largest member of Λ_y . Then $\nexists \ i' \neq i^*$ such that $Y_y \cap T_{i'} \neq \emptyset$ and $Y_y \cap T_{i'+1} = \emptyset$.

Proof Suppose $\exists \ i' < i^*$ with $Y_y \cap T_{i'} \neq \emptyset$. Since Y_y is connected we must have that $Y_y \cap T_i \neq \emptyset \ \forall \ i \in \{i', i'+1, \ldots, i^*\}$ and in particular we cannot have $Y_y \cap T_{i'+1} = \emptyset$.

Note that the first proposition requires Y_y to be convex, in order that $Y_y \cap T_i$ is connected. We can relax this condition on Y_y provided that we find a mapping from $Y_y \cap T_i$ onto a connected set in T_i. Miles (1978b) suggests a mapping onto a unique "associated point" and Gundersen (1977) suggests taking the union of the particle profiles together with connecting lines. A further possibility is to take the convex hull, $(Y_y \cap T_i)^C$. We need to be able to identify all the profiles coming from the same particle to be able to implement any of these methods and also to implement the second proposition in practice. This condition is Gundersen's "General Requirement" (1986). Given that it is satisfied we are able to determine from a pair of consecutive planes, T_i and T_{i+1}, (a single plane in the case $i=\alpha$) and a rectangle, D_{ijk}, on T_i the value of $I(y,(i,j,k))$ for a given y. In the case $i=\alpha$ we have $I(y,(i,j,k))=1$ if and only if $Y_y \cap T_i \neq \emptyset$, $D_{ijk} \cap (Y_y \cap T_i)^C \neq \emptyset$ and $E_{ijk} \cap (Y_y \cap T_i)^C = \emptyset$. For $i \neq \alpha$ we also require $Y_y \cap T_{i+1} = \emptyset$. ($T_{i+1}$ is sometimes referred to as the "lookup" plane.) When these conditions hold we say that Y_y is "counted" by D_{ijk}. We note that there is a corresponding result taking T_{i-1} as the lookup plane with i^* defined to be the smallest member of Λ_y.

Let us now put our discussion into the context of Sterio's disector by considering the following construction of $\mathcal{A}$. Let d be a fixed real number less than the minimum caliper diameter of any particle, let $0<\delta\leq d$, $N=[(h_X(\omega)-\delta)/d]+1$ (where [.] denotes "the integer part of") and $t_i=\delta+(i-1)d$ $(i=1,...,N)$. Let τ represent a tessellation of the plane by a rectangle with sides of length r and s and let $\tau(\theta,u,v)$ be τ after rotation through an angle θ and translation parallel to the sides of the rotated rectangles by distances u and v respectively. For a plane $T_i=T(\omega,t_i)$ intersecting X let $\tau_i=\tau_i(\theta_i,u_i,v_i)$ be the specific tessellation covering T_i and denote by τ_i^* the subset of rectangles of τ_i which intersect $T_i\cap X$. Labelling the units of τ_i^* by D_{ij} $(j=1,...,M_i)$ we let $D_{ij}(\pm)$ denote the disector consisting of D_{ij} together with lookup plane $T_{i\pm1}$. Although we have already remarked that a lookup plane is not required for $D_{Nj}(+)$ or $D_{1j}(-)$ we define T_0 and T_{N+1} to be $T(\omega,\delta-d)$ and $T(\omega,\delta+Nd)$ for consistency, so that we may regard $D_{ij}(\pm)$ as a cuboid.

Given this framework how should we interpret the phrase "randomly located disector"? For $\omega,\delta,(\theta_1,u_1,v_1),...,(\theta_N,u_N,v_N)$ fixed we have a finite population of M disectors where

$$M=\sum_{i=1}^{N} M_i \ .$$

(Strictly there are $2M$ disectors but we will restrict attention to $\{D_{ij}(+)\}$ for simplicity.) We can construct a bijective mapping, f, from $\{1,...,M\}$ to $\{(i,j):j=1,...,M_i;\ i=1,...,N\}$ defined by

$$f^{-1}((i,j))=\sum_{k=1}^{i-1} M_k + j$$

so that a sample $\{s_1,...,s_m\}$ from $\{1,...,M\}$ corresponds to a sample of disectors $\{D_{f(s_1)}(+),...,D_{f(s_m)}(+)\}$. Adapting our earlier notation to let $I(y,(i,j))=1$ if Y_y is counted by $D_{ij}(+)$ and zero otherwise, and letting $q_{ij}=\Sigma_y I(y,(i,j))$ then if $\{s_1,...,s_m\}$ is a simple random sample the estimator

$$\hat{Q}=\frac{M}{m}\sum_{i=1}^{m} q_{f(s_i)}$$

is certainly an unbiased estimator of Q. However, although we only have to observe $D_{f(s_i)}$ and the corresponding lookup plane[1] in order to evaluate $q_{f(s_i)}$ one drawback of $\hat{Q}$ is that we need to know M, which means that we need to form all of the tessellations $\tau_1,...,\tau_N$ and enumerate $M_1,...,M_N$. Therefore it might be useful in some cases to be able to replace M by an estimate and tessellate only those planes which contain the disectors in the sample[2]. The problem then is how to generate the sample, since we no longer have a finite population from which to select. We can regard the problem as one of subsampling with units of unequal size (the units being the section planes) where the sizes of the units $(M_1,...,M_N)$ are not known precisely. Indeed in our case they are random variables; however, if we take θ_i, u_i and v_i to be independent, uniform random variables on $(0,2\pi]$, $(0,r]$ and $(0,s]$ respectively then the following result enables us to find unbiased estimates of $\{M_i\}$.

Proposition If a tessellation of congruent rectangles is superimposed uniformly at random on a bounded set B then the expectation of the number of rectangles that intersect B is

$$\frac{\text{Mean area of dilation of } B \text{ by rectangle}}{\text{Area of rectangle}}$$

where the mean is taken over all orientations of the rectangle.

---

[1] In fact we need to observe more than D_{ij} in order to identify all of the constituent parts of any particle which intersects D_{ij} and in order to check that the exclusion line is not intersected but the important point is that only two planes are required for each dissector in the sample and only one needs to be tessellated.

[2] Although two planes are required to form a dissector we naturally associate $D_{ij}(+)$ with T_i and hence say that $D_{ij}(+)$ is "contained in" T_i.

Proof Fix an arbitrary Cartesian frame of reference in the plane with origin O and let $R_{r,s}$ be the rectangle with co-ordinates $(0,0),(0,s),(r,s)$ and $(r,0)$. Let $R(\theta,x,y)$ be $R_{r,s}$ rotated through an angle θ and translated by the vector $(x,y)^T$ (relative to the rotated axes) and let $I(B,R(\theta,x,y))$ equal one if $R(\theta,x,y)$ intersects B and zero otherwise. Then, for any fixed θ, the area of dilation of B by $R(\theta,0,0)$ is given by

$$A(B\oplus R(\theta,0,0)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(B,R(\theta,x,y))\, dxdy$$

$$= \int_0^s \int_0^r \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} I(B,R(\theta, u+ir, v+js))\, dudv$$

$$= rs \times \text{expected number of rectangles of uniformly randomly located tessellation with orientation } \theta \text{ which hit } B.$$

Integrating with respect to a uniform distribution on θ gives the result.

Notice that in the proof we have used the Minkowski sum, $B\oplus R(\theta,0,0,)$, rather than the dilation, $B\oplus {}^{-}R(\theta,0,0,)$ (where ${}^{-}R$ denotes the set $\{-r : r \in R\}$) but since they differ only by a translation their areas will be equal and thus here and elsewhere we use the notation interchangeably when areas are being considered. If B is convex then the mean area of dilation of $B$by a rectangle of sides r and s is given by the formula

$$A(B) + rs + (r+s)B(B)/\pi$$

(where $B(\ .\ )$ denotes boundary length), which can be found in Santaló, 1976, p.94. When B is not convex, or if its perimeter and/or area are unknown the area of dilation can be estimated with an automatic image analyser.

The relevant theory for subsampling with units of unequal size can be found in Cochran (1977, Ch.11). In line with Cochran we use the following notation for the ith unit:

	Population	Sample
Number of elements	M_i	m_i
Mean per element	$\bar{Q}_i$	$\bar{q}_i$
Total	$Q_i=M_i\bar{Q}_i$	$q_i=m_i\bar{q}_i$

There are N units in the population and we let n be the number of units in the sample. Then our previously defined notation is related to the quantities above by

$$M=\sum_{i=1}^{N} M_i \qquad m=\sum_{i=1}^{n} m_i \qquad Q=\sum_{i=1}^{N} Q_i \ .$$

In addition we define $\bar{Q}$ to be Q/N, $\ddot{Q}$ to be Q/M and $\bar{M}$ to be M/N; the sampling fraction in the first stage of sampling is $f_1=n/N$ and the sampling fraction in the ith unit at the second stage is $f_{2i}=m_i/M_i$. We describe below four possible estimators of Q; the first does not require any knowledge of M_i for sections not in the sample whilst the second requires an estimate of M. The third and fourth use estimates of the M_i to construct the sampling scheme.

Let ζ be a simple random sample of size n from $\{1,...,N\}$ and let $\{T_i:i\in\zeta\}$ be the corresponding simple random sample of planar sections. On each plane in the sample we superimpose a tessellation $\tau_i(\theta_i,u_i,v_i)$ where θ_i,u_i,v_i have the uniform distribution described earlier, and from the M_i elements in τ_i^* we draw a simple random sample of size m_i corresponding to a simple random sample η of size m_i from

$\{1,...,M_i\}$. Then

$$\tilde{Q}_1 = \frac{N}{n}\sum_{i\in\zeta} M_i\bar{q}_i$$

is an unbiased estimator of Q with variance

$$\frac{1}{f_1}\sum_{i=1}^{N}\frac{M_i^2 S_{2i}^2}{m_i}(1-f_{2i}) + \frac{N}{(N-1)}\frac{(1-f_1)}{f_1}\sum_{i=1}^{N}(Q_i-\bar{Q})^2 ,$$

where S_{2i}^2 is the population variance in unit i.

With the same sampling scheme as above we can also form the "ratio-to-size" estimator

$$\tilde{Q}_2 = \frac{\mu\sum_{i\in\zeta} M_i\bar{q}_i}{\sum_{i\in\zeta} M_i} .$$

This is a biased estimator because of its random denominator and its mean square error (MSE) is given by $E[(\tilde{Q}_2 - Q)^2]$. We have that

$$\tilde{Q}_2 - Q = \mu\,\frac{\sum_{i\in\zeta} M_i(\bar{q}_i - Q/\mu)}{\sum_{i\in\zeta} M_i}$$

and the usual procedure, in the case of the $\{M_i\}$ being fixed and known, is to replace the $\sum M_i$ in the denominator by $n\bar{M}$. However, since M is a random variable itself in our context, it makes more sense to use $n\bar{\mu}$ in the denominator, giving an approximate MSE for $\tilde{Q}_2$ of

$$\frac{1}{f_1}\sum_{i=1}^{N}\frac{M_i^2 S_{2i}^2}{m_i}(1-f_{2i})+\frac{1}{f_1}\left(1-\frac{n-1}{N-1}\right)\sum_{i=1}^{N}\left(Q_i-\frac{M_i}{\mu}Q\right)^2$$

$$+\left(\frac{1-1/n}{1-1/N}\right)\left(1-\frac{M}{\mu}\right)^2 Q^2 \quad .$$

Our third estimator is based on sampling of units with probability proportional to size. We let μ_i be an estimate of M_i and define

$$v_j=\frac{1}{\mu}\sum_{i=1}^{j}\mu_i \qquad\qquad \mu=\sum_{i=1}^{N}\mu_i \quad .$$

Now we generate n independent and identically distributed uniform random variables $\{w_1,...,w_n\}$ on the interval (0,1) and define ζ by

$$v_{i-1}<w_k\leq v_i \Rightarrow i\in\zeta \qquad\qquad (k=1,...,n).$$

Then $\{T_i: i\in\zeta\}$ represents a sample drawn from $\{T_1,...,T_N\}$ with probability proportional to estimated size. The drawback with this method of selection is that the same unit can be included twice in the sample, but techniques for sampling proportional to size without replacement tend to be rather complicated to implement (except when only two units are being drawn) and probably would not justify the extra effort involved in this context (see Cochran, 1977, § 9A.6). We draw a simple random sample of size m_i without replacement from each unit in the sample. Where a unit has been included λ_i times we have three possible courses of action. We can draw λ_i independent simple random samples, each one being taken from the whole unit, allowing the same element to be included in more than one of the sub-samples. Alternatively we can draw a single simple random sample of size $\lambda_i m_i$ or take a simple random sample of size m_i and weight the final estimate proportional to λ_i. Cochran points out that when the overall sampling fraction is small there is little difference in precision between these methods and therefore we assume that the first option is

adopted. The unbiased estimate of Q in this case is

$$\tilde{Q}_3 = \frac{\mu}{n} \sum_{i \in \zeta} \frac{M_i}{\mu_i} \bar{q}_i$$

with variance

$$\frac{\mu}{n} \sum_{i=1}^{N} \frac{M_i}{\mu_i} \frac{M_i S_{2i}^2}{m_i} (1-f_{2i}) + \frac{\mu}{n} \sum_{i=1}^{N} \mu_i \left(\frac{Q_i}{\mu_i} - \frac{Q}{\mu} \right)^2 .$$

All of the methods described so far have restricted the sample to come from n units of the population. Another possibility, defining μ_i and ν_i as before, is to take m independent and identically distributed uniform random variables $\{w_1,...,w_m\}$ on the interval (0,1) and define $J(i,k)=1$ if $\nu_{i-1}<w_k \leq \nu_i$ and zero otherwise. Then for each i such that $\sum_k J(i,k)>0$ we include T_i in the sample and select a simple random sample of m_i elements from T_i, where $m_i=\sum_k J(i,k)$. Then

$$\tilde{Q}_4 = \frac{\mu}{m} \sum_{i \in \zeta} \frac{M_i}{\mu_i} q_i$$

is an unbiased estimate of Q. Since the numbers of elements of the sample from each unit have a multinomial distribution we have introduced dependencies between the contributions to the estimator from each unit. The expression for the variance of this estimator is

$$\frac{\mu}{m} \sum_{i=1}^{N} \left(\frac{(M_i-1)}{\mu_i} - \frac{(m-1)}{\mu} \right) M_i S_{2i}^2 + \frac{\mu}{m} \sum_{i=1}^{N} \left(\frac{Q_i^2}{\mu_i} - \frac{\bar{Q}^2}{\bar{\mu}} \right) .$$

Thus we have, for the "within-units" variance

$$\frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2 S_{2i}^2}{m_i}(1-f_{2i})$$

$$\frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2 S_{2i}^2}{m_i}(1-f_{2i})$$

$$\frac{N}{n}\sum_{i=1}^{N}\frac{\bar{\mu}}{\mu_i}\frac{M_i^2 S_{2i}^2}{m_i}(1-f_{2i})$$

$$\frac{N}{n}\sum_{i=1}^{N}\left\{\frac{\bar{\mu}}{\mu_i}\frac{M_i(M_i-1)S_{2i}^2}{\bar{m}} - \frac{n}{N}\left(1-\frac{1}{m}\right)M_i S_{2i}^2\right\}$$

and for the "between-units" variance

$$\frac{N}{n}\left(1-\frac{n-1}{N-1}\right)\sum_{i=1}^{N}\left(Q_i-\bar{M}\frac{Q}{M}\right)^2$$

$$\frac{N}{n}\left(1-\frac{n-1}{N-1}\right)\sum_{i=1}^{N}\left(Q_i-M_i\frac{Q}{\mu}\right)^2 + \left(\frac{n-1}{N-1}\right)\left(1-\frac{M}{\mu}\right)^2 Q^2$$

$$\frac{N}{n}\bar{\mu}\sum_{i=1}^{N}\left(\frac{Q_i^2}{\mu_i}-\frac{\bar{Q}^2}{\bar{\mu}}\right)$$

$$\frac{N}{n}\frac{\bar{\mu}}{\bar{m}}\sum_{i=1}^{N}\left(\frac{Q_i^2}{\mu_i}-\frac{\bar{Q}^2}{\bar{\mu}}\right).$$

Clearly the within-units variance of $\tilde{Q}_1$ is equal to that of $\tilde{Q}_2$ whereas the equivalent expression for $\tilde{Q}_3$ is weighted inversely to the expected size of unit i. The effect of this will depend on the sampling fractions, $\{f_{2i}\}$, used. If we suppose S_{2i}^2 to be a constant, σ^2, for all i then, noting that $m_i = f_{2i} M_i$, we can rewrite the within-units variances as:

$$\frac{N}{n}\sigma^2 \sum_{i=1}^{N} M_i \left(\frac{1}{f_{2i}} - 1\right) \qquad (\tilde{Q}_1 \text{ and } \tilde{Q}_2)$$

$$\frac{N}{n}\sigma^2 \sum_{i=1}^{N} \frac{\bar{\mu}}{\mu_i} M_i \left(\frac{1}{f_{2i}} - 1\right) \qquad (\tilde{Q}_3)$$

$$\frac{N}{n}\sigma^2 \sum_{i=1}^{N} \left\{ \frac{\bar{\mu}}{\mu_i} \frac{M_i (M_i - 1)}{\bar{m}} - \frac{n}{N}\left(1 - \frac{1}{m}\right) M_i \right\} \qquad (\tilde{Q}_4).$$

$\tilde{Q}_4$ does not involve any "second-stage" sampling but for $\tilde{Q}_1$, $\tilde{Q}_2$ and $\tilde{Q}_3$ we would like to find the optimal values for $\{f_{2i}\}$ to minimise these expressions. In fact, since we are considering the second stage of sampling, the choice of $\{f_{2i}\}$ should be sample dependent, being a minimisation of the conditional expectation of the within-units variance given the particular set of units selected at the first stage. Thus for $\tilde{Q}_1$ and $\tilde{Q}_2$ we want to minimise

$$\sum_{i \in \zeta} M_i \left(\frac{1}{f_{2i}} - 1\right)$$

subject to the constraint

$$\sum_{i \in \zeta} f_{2i} M_i = m .$$

This reduces to the solution of the equations

$$\frac{-M_i}{f_{2i}^2} + \lambda M_i = 0 \qquad i \in \zeta$$

$$\sum_{i \in \zeta} f_{2i} M_i = m$$

giving

$$f_{2i} = \frac{1}{\sqrt{\lambda}} \qquad i \in \zeta$$

$$\sqrt{\lambda} = \frac{\sum_{i \in \zeta} M_i}{m}$$

that is

$$f_{2i} = \frac{m}{\sum_{i \in \zeta} M_i} .$$

Thus the optimal choice is f_{2i} = constant for all i. In theory this constant is sample dependent, being a function of the M_i's in the sample, but in practice, since the $\{m_i\}$ must be integer anyway, we would take $f_{2i} = f_2 = \bar{m}/\bar{M}$ as being close enough, since the mean of the M_i's in the sample will be close to $\bar{M}$ in most circumstances. In our particular context we would replace the unknown $\bar{M}$ by its expected value, $\bar{\mu}$, thus giving $f_{2i} = \bar{m}/\bar{\mu}$ for all i and for all first stage samples.

For $\tilde{Q}_3$ a similar argument leads to the value $f_{2i} = \bar{m}/M_i$, again being independent of the first stage sampling. Using the fact that

$$\sum_{i=1}^{N} M_i = M$$

and substituting the values for $\{f_{2i}\}$ derived above, the expressions for within-units variance are now

$$\sigma^2 M\left(\frac{\mu}{m} - \frac{N}{n}\right) \qquad (\tilde{Q}_1 \text{ and } \tilde{Q}_2)$$

$$\sigma^2 \frac{\mu}{m} \sum_{i=1}^{N} \frac{M_i}{\mu_i}\left(M_i - \frac{m}{n}\right) \qquad (\tilde{Q}_3)$$

$$\sigma^2 \left\{ \frac{\mu}{m} \sum_{i=1}^{N} \frac{M_i^2 - M_i}{\mu_i} - M\left(1 - \frac{1}{m}\right) \right\} \qquad (\tilde{Q}_4).$$

Taking expectations over the distribution of $\{M_i\}$ gives

$$\sigma^2 \mu \left\{ \frac{\mu}{m} - \frac{N}{n} \right\} \qquad (\tilde{Q}_1 \text{ and } \tilde{Q}_2)$$

$$\sigma^2 \mu \left\{ \frac{\mu}{m} - \frac{N}{n} + \frac{1}{m} \sum_{i=1}^{N} \frac{\text{var}(M_i)}{\mu_i} \right\} \qquad (\tilde{Q}_3)$$

$$\sigma^2 \mu \left\{ \frac{\mu}{m} - \frac{(N-1)}{m} - 1 + \frac{1}{m} \sum_{i=1}^{N} \frac{\text{var}(M_i)}{\mu_i} \right\} \qquad (\tilde{Q}_4).$$

Clearly $\tilde{Q}_3$ is always inferior to $\tilde{Q}_1$ and $\tilde{Q}_2$ in terms of within-units variance. We do not have an explicit expression for var(M_i) but there are approximations available. We recall that M_i is the number of points of a regular grid falling within a region, which, up to a translation, is the dilation of the ith planar section through the specimen by a rectangle of the grid. For a uniformly randomly located grid an approximate formula for the variance of M_i due to Matheron can be found in Matérn (1985), namely

$$\text{var}(M_i) = (0.0531b^{-1} + 0.0194ba^{-2})B_i$$

where the rectangles of the grid have sides a and b ($a<b$) and B_i is the boundary of the dilated planar section; it only remains to find B_i. For a convex planar set, K, with perimeter $U(K)$, the perimeter of $K \oplus b(o, r)$, where $b(o, r)$ is a disc of radius r, is given by $U(K) + 2\pi r$ (see Stoyan, Mecke and Kendall, 1987, p.27). This allows us to find upper and lower bounds for B_i whenever the section through the specimen

is convex. Furthermore, it is easily seen that the perimeter of the dilation of a convex polygon by a rectangular set is equal to the sum of the perimeters of the polygon and the rectangle. Thus for convex specimens we can find a reasonably good approximation to var(M_i). Table 2.1 shows the ratios of the within-units variances of $\tilde{Q}_3$ to $\tilde{Q}_1$ and $\tilde{Q}_4$ to $\tilde{Q}_1$ for a variety of different situations, under the assumption of constant variance within sections, using the expressions derived above. The specimen is taken to be spherical, so that all sections are circular. In this case the length of the perimeter of the dilation of a section by a rectangle involves elliptic integrals so we take the mean of the lower and upper bounds (that is, the perimeters of dilations by discs of radius a and $\sqrt{(a^2 + b^2)}$ respectively, where the rectangle has dimensions a, b with $a<b$). As the rectangular unit of the grid is in general much smaller than the specimen section the error is likely to be negligible.

The difference between $\tilde{Q}_3$ and $\tilde{Q}_1$ is seen to be very slight in all the cases considered, whereas the performance of $\tilde{Q}_4$ relative to $\tilde{Q}_1$ is considerably lower in some instances. It is particularly bad when a large number of elements is chosen from a small number of units, but when the first stage sampling fraction is high and the second stage sampling fraction is low there is not much to choose between $\tilde{Q}_4$ and $\tilde{Q}_1$ in terms of within-units variance, under the assumption of constant population variance within units.

N	n	m	a	b	μ	W1	W3/W1	W4/W1	N/n	M/m
10	2	25	0.20	0.40	392	4188	1.0025	1.3432	5	16
10	2	50	0.20	0.40	392	1114	1.0046	2.3491	5	8
10	5	50	0.20	0.40	392	2290	1.0022	1.1426	2	8
10	5	100	0.20	0.40	392	753	1.0034	1.4772	2	4
10	2	25	0.25	0.25	474	6606	1.0014	1.2624	5	19
10	2	50	0.25	0.25	474	2119	1.0021	1.8561	5	9
10	5	50	0.25	0.25	474	3540	1.0013	1.1110	2	9
10	5	100	0.25	0.25	474	1296	1.0017	1.3343	2	5
10	2	6	0.40	0.80	135	2373	1.0114	1.1539	5	23
10	2	12	0.40	0.80	135	848	1.0159	1.5341	5	11
10	5	12	0.40	0.80	135	1254	1.0107	1.0377	2	11
10	5	25	0.40	0.80	135	461	1.0140	1.2017	2	5
10	2	6	0.50	0.50	158	3353	1.0070	1.1245	5	26
10	2	12	0.50	0.50	158	1283	1.0091	1.4086	5	13
10	5	12	0.50	0.50	158	1755	1.0067	1.0291	2	13
10	5	25	0.50	0.50	158	679	1.0083	1.1569	2	6
40	8	100	0.20	0.40	1559	16519	1.0026	1.3434	5	16
40	8	200	0.20	0.40	1559	4361	1.0049	2.3654	5	8
40	20	200	0.20	0.40	1559	9039	1.0024	1.1412	2	8
40	20	400	0.20	0.40	1559	2960	1.0036	1.4790	2	4
40	8	100	0.25	0.25	1884	26078	1.0014	1.2623	5	19
40	8	200	0.25	0.25	1884	8329	1.0023	1.8630	5	9
40	20	200	0.25	0.25	1884	13981	1.0014	1.1098	2	9
40	20	400	0.25	0.25	1884	5106	1.0018	1.3348	2	5
40	8	25	0.40	0.80	538	8880	1.0120	1.1598	5	22
40	8	50	0.40	0.80	538	3095	1.0172	1.5766	5	11
40	20	50	0.40	0.80	538	4709	1.0113	1.0364	2	11
40	20	100	0.40	0.80	538	1817	1.0146	1.1952	2	5
40	8	25	0.50	0.50	627	12583	1.0074	1.1289	5	25
40	8	50	0.50	0.50	627	4724	1.0098	1.4370	5	13
40	20	50	0.50	0.50	627	6605	1.0070	1.0279	2	13
40	20	100	0.50	0.50	627	2676	1.0087	1.1516	2	6

Table 2.1 Within-units Components of Variance - S_{2i}^2 Constant.

The sixth column, headed μ, is the expected number of rectangles of the grid intersecting all sections. The column W1 is $1/\sigma^2$ times the within-units component of the variance of $\tilde{Q}_1$. The columns W3/W1 and W4/W1 are the ratios of the within-units components of the variances of $\tilde{Q}_3$ and $\tilde{Q}_4$ relative to $\tilde{Q}_1$.

The assumption of constant variance from one section to another seems reasonable. For a sparse population of particles, where each sampling element contains either zero or one particle, the population variance within a unit, given the total is Q_i, is given by $Q_i/M_i(1 - Q_i/M_i)$. In such a situation, even if Q_i/M_i is not constant, S_{2i}^2 is bounded between 0 and 0.25 and is likely to be dominated by the between-units variance, so the assumption of constant variance should not be disastrous. If, on the other hand, the number of particles is large relative to the number of sampling elements but is fairly homogeneous we might expect the numbers of particles in the elements of the ith unit to have a multinomial distribution with total Q_i and cell probabilities $1/M_i$ for each element. This would give $S_{2i}^2 = Q_i/M_i$, and, with Q being large relative to μ and the distribution of particles being homogeneous, an assumption that Q_i is approximately proportional to M_i seems fair, giving $S_{2i}^2 =$ constant. Let us now consider the effect of the assumption of constant variance being violated. We consider two cases, namely $S_{2i}^2 = \sigma^2 M_i$ and $S_{2i}^2 = \sigma^2/M_i$. When $S_{2i}^2 = \sigma^2 M_i$ and we use the same values of $\{f_{2i}\}$ as before we obtain for the within-units variance:

$$\left(\frac{\mu}{m} - \frac{N}{n}\right)\sum_{i=1}^{N} M_i^2 \qquad (\tilde{Q}_1 \text{ and } \tilde{Q}_2)$$

$$\frac{\mu}{m}\sum_{i=1}^{N}\frac{1}{\mu_i}\left(M_i^3 - \frac{m}{n}M_i^2\right) \qquad (\tilde{Q}_3)$$

$$\frac{\mu}{m}\sum_{i=1}^{N}\left\{\frac{1}{\mu_i}(M_i^3 - M_i^2) - \frac{(m-1)}{\mu}M_i^2\right\} \qquad (\tilde{Q}_4).$$

These expressions are not very helpful as we do not know anything about $E(M_i^3)$. As a very rough guide we can replace M_i by μ_i and we find that the expression for $\tilde{Q}_3$ is always greater than that for $\tilde{Q}_1$, although the expression for $\tilde{Q}_4$ is not so revealing. Table 2.2 shows the same ratios as Table 2.1 for the same values of N, n, m, etc. but under the altered assumption about the within-units variance.

N	n	m	a	b	μ	W1	W3/W1	W4/W1	N/n	M/m
10	2	25	0.20	0.40	392	186481	1.0559	1.3452	5	16
10	2	50	0.20	0.40	392	49598	1.2103	2.3529	5	8
10	5	50	0.20	0.40	392	101969	1.0409	1.1445	2	8
10	5	100	0.20	0.40	392	33528	1.1244	1.4800	2	4
10	2	25	0.25	0.25	474	357278	1.0445	1.2646	5	19
10	2	50	0.25	0.25	474	114594	1.1389	1.8595	5	9
10	5	50	0.25	0.25	474	191448	1.0333	1.1131	2	9
10	5	100	0.25	0.25	474	70106	1.0908	1.3371	2	5
10	2	6	0.40	0.80	135	35439	1.0269	1.1515	5	23
10	2	12	0.40	0.80	135	12669	1.0753	1.5308	5	11
10	5	12	0.40	0.80	135	18729	1.0204	1.0355	2	11
10	5	25	0.40	0.80	135	6889	1.0554	1.1988	2	5
10	2	6	0.50	0.50	158	58765	1.0236	1.1254	5	26
10	2	12	0.50	0.50	158	22476	1.0618	1.4097	5	13
10	5	12	0.50	0.50	158	30764	1.0180	1.0300	2	13
10	5	25	0.50	0.50	158	11894	1.0467	1.1580	2	6
40	8	100	0.20	0.40	1559	738770	1.0606	1.3456	5	16
40	8	200	0.20	0.40	1559	195038	1.2294	2.3697	5	8
40	20	200	0.20	0.40	1559	404254	1.0443	1.1433	2	8
40	20	400	0.20	0.40	1559	132388	1.1352	1.4822	2	4
40	8	100	0.25	0.25	1884	1416754	1.0480	1.2647	5	19
40	8	200	0.25	0.25	1884	452479	1.1504	1.8668	5	9
40	20	200	0.25	0.25	1884	759557	1.0358	1.1121	2	9
40	20	400	0.25	0.25	1884	277419	1.0981	1.3379	2	5
40	8	25	0.40	0.80	538	133119	1.0312	1.1578	5	22
40	8	50	0.40	0.80	538	46404	1.0896	1.5738	5	11
40	20	50	0.40	0.80	538	70590	1.0236	1.0346	2	11
40	20	100	0.40	0.80	538	27233	1.0611	1.1928	2	5
40	8	25	0.50	0.50	627	221359	1.0272	1.1303	5	25
40	8	50	0.50	0.50	627	83111	1.0724	1.4388	5	13
40	20	50	0.50	0.50	627	116193	1.0207	1.0292	2	13
40	20	100	0.50	0.50	627	47069	1.0512	1.1531	2	6

Table 2.2 Within-units Components of Variance - S_{2i}^2 Proportional to M_i.

The estimators follow a similar pattern to that of the previous case, although the difference between $\tilde{Q}_1$ and $\tilde{Q}_3$ is slightly more marked. However, we would not feel worried about using $\tilde{Q}_3$ in preference to $\tilde{Q}_1$ if there were other grounds on which we suspected that $\tilde{Q}_3$ could be better, such as between-units variance. It seems that in this particular case the violation of the assumption of constant population variance within-units will not invalidate conclusions made on the grounds of within-units variance about the relative merits of the various estimators, although the second stage sampling fractions will almost certainly not now be optimal.

Finally, when $S_{2i}^2 = \sigma^2/M_i$ the within-units variances, with the same values of f_{2i} as before, and after taking expectations over the distribution of $\{M_i\}$, are

$$\left(\frac{\mu}{m} - \frac{N}{n}\right) N \qquad (\tilde{Q}_1 \text{ and } \tilde{Q}_2)$$

$$N\frac{\mu}{m} - \frac{\mu}{n}\sum_{i=1}^{N}\frac{1}{\mu_i} \qquad (\tilde{Q}_3)$$

$$N\frac{\mu}{m} - \frac{\mu}{n}\sum_{i=1}^{N}\frac{1}{\mu_i} - \left(1 - \frac{1}{m}\right) N \qquad (\tilde{Q}_4).$$

The ratios of the expressions for $\tilde{Q}_1$ and $\tilde{Q}_3$, and the expressions for $\tilde{Q}_1$ and $\tilde{Q}_4$ are shown in Table 2.3.

In this case, although the results are very similar for $\tilde{Q}_4$ relative to $\tilde{Q}_1$ we see a reversal with $\tilde{Q}_3$ and $\tilde{Q}_1$, with $\tilde{Q}_3$ now exhibiting a smaller within-units variance than $\tilde{Q}_1$ in all instances, sometimes markedly so. In all cases we have to weigh the within-units variance against the between-units variance, and that is what we consider next.

N	n	m	a	b	μ	W1	W3/W1	W4/W1	N/n	M/m
10	2	25	0.20	0.40	392	107	0.8844	1.3315	5	16
10	2	50	0.20	0.40	392	28	0.5653	2.3271	5	8
10	5	50	0.20	0.40	392	58	0.9154	1.1319	2	8
10	5	100	0.20	0.40	392	19	0.7428	1.4610	2	4
10	2	25	0.25	0.25	474	139	0.9052	1.2534	5	19
10	2	50	0.25	0.25	474	45	0.7043	1.8422	5	9
10	5	50	0.25	0.25	474	75	0.9292	1.1026	2	9
10	5	100	0.25	0.25	474	27	0.8067	1.3229	2	5
10	2	6	0.40	0.80	135	175	0.9521	1.1265	5	23
10	2	12	0.40	0.80	135	63	0.8659	1.4959	5	11
10	5	12	0.40	0.80	135	93	0.9637	1.0119	2	11
10	5	25	0.40	0.80	135	34	0.9014	1.1680	2	5
10	2	6	0.50	0.50	158	213	0.9565	1.1030	5	26
10	2	12	0.50	0.50	158	81	0.8863	1.3805	5	13
10	5	12	0.50	0.50	158	111	0.9668	1.0086	2	13
10	5	25	0.50	0.50	158	43	0.9141	1.1315	2	6
40	8	100	0.20	0.40	1559	424	0.8129	1.3258	5	16
40	8	200	0.20	0.40	1559	112	0.2914	2.3322	5	8
40	20	200	0.20	0.40	1559	232	0.8632	1.1252	2	8
40	20	400	0.20	0.40	1559	76	0.5824	1.4545	2	4
40	8	100	0.25	0.25	1884	554	0.8439	1.2483	5	19
40	8	200	0.25	0.25	1884	177	0.5114	1.8412	5	9
40	20	200	0.25	0.25	1884	297	0.8836	1.0968	2	9
40	20	400	0.25	0.25	1884	108	0.6812	1.3171	2	5
40	8	25	0.40	0.80	538	660	0.9252	1.1238	5	22
40	8	50	0.40	0.80	538	230	0.7854	1.5251	5	11
40	20	50	0.40	0.80	538	350	0.9436	1.0026	2	11
40	20	100	0.40	0.80	538	135	0.8537	1.1513	2	5
40	8	25	0.50	0.50	627	803	0.9314	1.0996	5	25
40	8	50	0.50	0.50	627	301	0.8173	1.3980	5	13
40	20	50	0.50	0.50	627	421	0.9477	1.0000	2	13
40	20	100	0.50	0.50	627	171	0.8710	1.1171	2	6

Table 2.3 Within-units Components of Variance - S_{2i}^2 Inversely Proportional to M_i.

Turning to the between-units variance we first consider the most extreme case, that is

$$Q_k = Q$$
$$Q_i = 0 \qquad\qquad i \neq k$$

for some $k \in \{1, \ldots, N\}$. The expected values of the corresponding components of the variance for $\tilde{Q}_1$, $\tilde{Q}_2$, $\tilde{Q}_3$ and $\tilde{Q}_4$ are

$$\frac{N}{n} Q^2 \left(1 - \frac{n-1}{N-1}\right)\left(1 - \frac{1}{N}\right)$$

$$\frac{N}{n} Q^2 \frac{1}{\mu^2}\left[\sum_{i=1}^{N} \mathrm{var}(M_i) + \left(1 - \frac{n-1}{N-1}\right)\left\{\sum_{i=1}^{N} \mu_i^2 + \mu^2 - 2\mu\mu_k\right\}\right]$$

$$\frac{N}{n} Q^2 \frac{1}{N}\left(\frac{\mu}{\mu_k} - 1\right)$$

$$\frac{N}{n} Q^2 \frac{1}{N\bar{m}}\left(\frac{\mu}{\mu_k} - 1\right)$$

The expression for $\tilde{Q}_1$ is less than or equal to that for $\tilde{Q}_2$ whenever $\mu_k \leq \bar{\mu}$ whereas the situation is not clear for the case $\mu_k > \bar{\mu}$. If we take $\mu_k = \lambda\mu$ then, approximating the expression for $\tilde{Q}_1$ by

$$\frac{N}{n} Q^2 \left(1 - \frac{n}{N}\right)\left(1 - \frac{1}{N}\right)$$

we find that it is smaller than that for $\tilde{Q}_3$ whenever $\lambda < 1/[(1 - f_1)N]$, that is when $\mu_k < \bar{\mu}/(1 - f_1)$. On the other hand the expression for $\tilde{Q}_4$ is smaller than that for $\tilde{Q}_1$ whenever

$$\mu_k > \frac{\bar{\mu}}{\frac{m}{N}\left(\frac{N}{n} - 1\right)}$$

which we would expect to be true most of the time. Table 2.4 shows the ratios of the expressions for $\tilde{Q}_2$ and $\tilde{Q}_3$ relative to that for $\tilde{Q}_1$, for the same situations as were considered in the study of the within-units variance. We do not consider different values of m since it only appears in the expression for $\tilde{Q}_4$ and this is easily derived from that for $\tilde{Q}_3$, simply dividing by $\bar{m}$.

N	n	μ_k	a	b	μ	B1	B2/B1	B3/B1
10	2	0.38	0.20	0.40	392	4	1.1525	3.1422
10	2	0.81	0.20	0.40	392	4	1.0576	1.4190
10	2	1.40	0.20	0.40	392	4	0.9269	0.7693
10	5	0.38	0.20	0.40	392	1	1.1526	5.0276
10	5	0.81	0.20	0.40	392	1	1.0577	2.2705
10	5	1.40	0.20	0.40	392	1	0.9270	1.2308
10	2	0.37	0.25	0.25	474	4	1.1559	3.2535
10	2	0.80	0.25	0.25	474	4	1.0595	1.4306
10	2	1.41	0.25	0.25	474	4	0.9252	0.7629
10	5	0.37	0.25	0.25	474	1	1.1559	5.2056
10	5	0.80	0.25	0.25	474	1	1.0596	2.2889
10	5	1.41	0.25	0.25	474	1	0.9253	1.2206
10	2	0.45	0.40	0.80	135	4	1.1340	2.6310
10	2	0.84	0.40	0.80	135	4	1.0481	1.3629
10	2	1.34	0.40	0.80	135	4	0.9361	0.8049
10	5	0.45	0.40	0.80	135	1	1.1346	4.2096
10	5	0.84	0.40	0.80	135	1	1.0487	2.1806
10	5	1.34	0.40	0.80	135	1	0.9367	1.2878
10	2	0.44	0.50	0.50	158	4	1.1384	2.7412
10	2	0.83	0.50	0.50	158	4	1.0503	1.3766
10	2	1.36	0.50	0.50	158	4	0.9336	0.7957
10	5	0.44	0.50	0.50	158	1	1.1387	4.3859
10	5	0.83	0.50	0.50	158	1	1.0507	2.2026
10	5	1.36	0.50	0.50	158	1	0.9340	1.2732

Table 2.4 Between-units Variance - Extreme Case.

N	n	μ_k	a	b	μ	B1	B2/B1	B3/B1
40	8	0.16	0.20	0.40	1559	4	1.0469	7.7381
40	8	1.08	0.20	0.40	1559	4	0.9996	1.1231
40	8	1.42	0.20	0.40	1559	4	0.9824	0.8509
40	20	0.16	0.20	0.40	1559	1	1.0469	12.3809
40	20	1.08	0.20	0.40	1559	1	0.9996	1.7970
40	20	1.42	0.20	0.40	1559	1	0.9825	1.3614
40	8	0.15	0.25	0.25	1884	4	1.0476	8.3355
40	8	1.08	0.25	0.25	1884	4	0.9997	1.1226
40	8	1.43	0.25	0.25	1884	4	0.9820	0.8445
40	20	0.15	0.25	0.25	1884	1	1.0476	13.3368
40	20	1.08	0.25	0.25	1884	1	0.9997	1.7961
40	20	1.43	0.25	0.25	1884	1	0.9821	1.3513
40	8	0.23	0.40	0.80	538	4	1.0425	5.3018
40	8	1.08	0.40	0.80	538	4	0.9992	1.1282
40	8	1.36	0.40	0.80	538	4	0.9847	0.8867
40	20	0.23	0.40	0.80	538	1	1.0426	8.4829
40	20	1.08	0.40	0.80	538	1	0.9994	1.8051
40	20	1.36	0.40	0.80	538	1	0.9848	1.4186
40	8	0.22	0.50	0.50	627	4	1.0435	5.7334
40	8	1.08	0.50	0.50	627	4	0.9992	1.1271
40	8	1.38	0.50	0.50	627	4	0.9840	0.8775
40	20	0.22	0.50	0.50	627	1	1.0436	9.1734
40	20	1.08	0.50	0.50	627	1	0.9993	1.8033
40	20	1.38	0.50	0.50	627	1	0.9841	1.4040

Table 2.4 (cont.) Between-units Variance - Extreme Case.

Between-units variances for the case when all of the particles are in a single section. The column B1 is $1/Q^2$ times the between-units component of the variance of $\tilde{Q}_1$. The columns B2/B1 and B3/B1 are the ratios of the between-units components of the variances of $\tilde{Q}_2$ and $\tilde{Q}_3$ relative to $\tilde{Q}_1$. The column headed μ is the expected number of rectangles of the grid intersecting all sections and the column headed μ_k is $1/\bar{\mu}$ times the expected number of rectangles of the grid intersecting the section containing the particles. In each case the three values of μ_k used correspond to the smallest, the largest and the median section sizes.

The values of μ_k used are the smallest, the largest and the median. The general impression is that there is little to choose between $\tilde{Q}_1$ and $\tilde{Q}_2$ whereas overall $\tilde{Q}_3$ is inferior to $\tilde{Q}_1$. The gain in using $\tilde{Q}_3$ when μ_k is large is far outweighed by the loss when μ_k is small since, by the nature of the sampling scheme for $\tilde{Q}_3$, the probability of selecting a small unit is small. $\tilde{Q}_4$, on the other hand, performs much better, only having a greater between-units variance than $\tilde{Q}_1$ for a few of the cases where all of the particles are in the smallest section (considering the same values of m as were used before). In all other cases it performs considerably better in terms of between-units variance.

The case considered above is extreme and we now consider two alternative possibilities, namely when the number of particles per unit is approximately constant and when the number of particles per unit is approximately proportional to the size of the unit.

The first possibility amounts to

$$Q_i = \bar{Q} + e_i \qquad |e_i| < 1 \quad \forall\, i$$

$$\Sigma\, e_i = 0.$$

In this case the expected values of the between-units variances are

$$\frac{N}{n}\left(1 - \frac{n-1}{N-1}\right)\sum_{i=1}^{N} e_i^2 \qquad (\tilde{Q}_1)$$

$$\frac{N}{n}\frac{Q^2}{\mu^2}\left[\sum_{i=1}^{N} \mathrm{var}(M_i) + \left(1 - \frac{n-1}{N-1}\right)\sum_{i=1}^{N} (\mu_i - \bar{\mu})^2\right]$$

$$+ \frac{N}{n}\left(1 - \frac{n-1}{N-1}\right)\left[\sum_{i=1}^{N} e_i^2 - \frac{2Q}{\mu}\sum_{i=1}^{N} e_i \mu_i\right] \qquad (\tilde{Q}_2)$$

$$\frac{N}{n}\left[Q^2\left(\frac{1}{N^2}\sum_{i=1}^{N}\frac{\bar{\mu}}{\mu_i} - \frac{1}{N}\right) + 2\bar{Q}\bar{\mu}\sum_{i=1}^{N}\frac{e_i}{\mu_i} + \bar{\mu}\sum_{i=1}^{N}\frac{e_i^2}{\mu_i}\right] \qquad (\tilde{Q}_3)$$

The second possibility can be expressed as

$$Q_i = \frac{\mu_i}{\mu} Q + e_i \qquad \forall i, \quad \Sigma e_i = 0, \quad |e_i| < 1$$

and gives expected between-units variances of

$$\frac{N}{n}\left(1 - \frac{n-1}{N-1}\right)\left[\frac{Q^2}{\mu^2}\sum_{i=1}^{N}(\mu_i - \bar{\mu})^2 + \frac{2Q}{\mu}\sum_{i=1}^{N} e_i \mu_i + \sum_{i=1}^{N} e_i^2\right] \qquad (\tilde{Q}_1)$$

$$\frac{N}{n}\left[\frac{Q^2}{\mu^2}\sum_{i=1}^{N}\mathrm{var}(M_i) + \left(1 - \frac{n-1}{N-1}\right)\sum_{i=1}^{N} e_i^2\right] \qquad (\tilde{Q}_2)$$

$$\frac{N}{n}\bar{\mu}\sum_{i=1}^{N}\frac{e_i^2}{\mu_i} \qquad (\tilde{Q}_3)$$

(The expressions for $\tilde{Q}_4$ have not been given since they are simply $1/\bar{m}$ times those for $\tilde{Q}_3$.)

Clearly $\tilde{Q}_1$ in the first case and $\tilde{Q}_4$ in the second case are going to be optimal, except in very extreme cases, as they have no term in Q. Tables 2.5 and 2.6 show performances of the estimators in terms of between-units variance for the same situations as before but with specific values of Q introduced, since now there is no common factor of Q^2 in all of the expressions. The values of Q used are intended to represent contrasting degrees of sparseness in the population of particles; for N=10 the values of Q are 15 and 95, giving values of 1.5 and 9.5 for $\bar{Q}$ and approximate expected values of $\ddot{Q}$ for the different sizes of disector of 0.04, 0.03, 0.11, 0.09 (Q=15) and 0.24, 0.20, 0.70, 0.60 (Q=95), whereas for N=40 the values of Q are 60 and 380, giving the same values of $\bar{Q}$ and $\ddot{Q}$. Although no mention is made of particle size there is an implicit assumption, at least about the minimum size, since the unbiasedness of the disector is dependent upon the distance between sections being at least as small as the minimum caliper diameter of any particle.

N	n	Q	a	b	μ	B1	B2/B1	B3/B1
10	2	15	0.20	0.40	392	11.11	1.2274	1.9849
10	2	95	0.20	0.40	392	11.11	44.2076	89.5598
10	5	15	0.20	0.40	392	2.78	1.2365	3.1758
10	5	95	0.20	0.40	392	2.78	44.5711	143.2957
10	2	15	0.25	0.25	474	11.11	1.2534	2.0917
10	2	95	0.25	0.25	474	11.11	46.0368	96.1350
10	5	15	0.25	0.25	474	2.78	1.2588	3.3467
10	5	95	0.25	0.25	474	2.78	46.2534	153.8160
10	2	15	0.40	0.80	135	11.11	1.1392	1.5223
10	2	95	0.40	0.80	135	11.11	36.3713	60.1372
10	5	15	0.40	0.80	135	2.78	1.1883	2.4357
10	5	95	0.40	0.80	135	2.78	38.3406	96.2195
10	2	15	0.50	0.50	158	11.11	1.1439	1.6180
10	2	95	0.50	0.50	158	11.11	37.6272	66.3927
10	5	15	0.50	0.50	158	2.78	1.1751	2.5887
10	5	95	0.50	0.50	158	2.78	38.8749	106.2283
40	8	60	0.20	0.40	1559	41.03	1.1260	2.5613
40	8	380	0.20	0.40	1559	41.03	47.1137	153.9964
40	20	60	0.20	0.40	1559	10.26	1.1359	4.0981
40	20	380	0.20	0.40	1559	10.26	47.5095	246.3942
40	8	60	0.25	0.25	1884	41.03	1.1483	2.7588
40	8	380	0.25	0.25	1884	41.03	48.9031	168.2218
40	20	60	0.25	0.25	1884	10.26	1.1542	4.4141
40	20	380	0.25	0.25	1884	10.26	49.1390	269.1549
40	8	60	0.40	0.80	538	41.03	1.0561	1.7569
40	8	380	0.40	0.80	538	41.03	39.3416	94.5674
40	20	60	0.40	0.80	538	10.26	1.1096	2.8110
40	20	380	0.40	0.80	538	10.26	41.4907	151.3078
40	8	60	0.50	0.50	627	41.03	1.0549	1.9052
40	8	380	0.50	0.50	627	41.03	40.5196	105.7977
40	20	60	0.50	0.50	627	10.26	1.0888	3.0482
40	20	380	0.50	0.50	627	10.26	41.8807	169.2764

Table 2.5 Between Units Variance - $Q_i \approx \bar{Q}$.

N	n	Q	a	b	μ	B3	B1/B3	B2/B3
10	2	15	0.20	0.40	392	20.59	1.9787	0.7866
10	2	95	0.20	0.40	392	21.52	31.2208	1.0214
10	5	15	0.20	0.40	392	8.24	1.2367	0.4947
10	5	95	0.20	0.40	392	8.61	19.5130	0.7557
10	2	15	0.25	0.25	474	20.35	2.0019	0.7803
10	2	95	0.25	0.25	474	17.78	37.7959	0.9505
10	5	15	0.25	0.25	474	8.14	1.2512	0.4895
10	5	95	0.25	0.25	474	7.11	23.6224	0.6787
10	2	15	0.40	0.80	135	22.23	1.8325	0.8337
10	2	95	0.40	0.80	135	19.77	25.2883	2.6811
10	5	15	0.40	0.80	135	8.89	1.1453	0.5364
10	5	95	0.40	0.80	135	7.91	15.8052	2.3674
10	2	15	0.50	0.50	158	21.78	1.8707	0.8158
10	2	95	0.50	0.50	158	17.50	28.5684	2.1518
10	5	15	0.50	0.50	158	8.71	1.1692	0.5198
10	5	95	0.50	0.50	158	7.00	17.8553	1.8400
40	8	60	0.20	0.40	1559	83.76	1.9294	0.7440
40	8	380	0.20	0.40	1559	71.59	6.2225	0.7045
40	20	60	0.20	0.40	1559	33.50	1.2059	0.4680
40	20	380	0.20	0.40	1559	28.64	3.8890	0.4534
40	8	60	0.25	0.25	1884	93.27	1.9852	0.7293
40	8	380	0.25	0.25	1884	84.44	6.0670	0.5987
40	20	60	0.25	0.25	1884	37.31	1.2407	0.4574
40	20	380	0.25	0.25	1884	33.78	3.7919	0.3808
40	8	60	0.40	0.80	538	83.14	1.7608	0.7614
40	8	380	0.40	0.80	538	79.60	6.5178	0.7952
40	20	60	0.40	0.80	538	33.26	1.1005	0.4924
40	20	380	0.40	0.80	538	31.84	4.0736	0.5609
40	8	60	0.50	0.50	627	87.12	1.7985	0.7643
40	8	380	0.50	0.50	627	79.11	6.4760	0.8424
40	20	60	0.50	0.50	627	34.85	1.1241	0.4877
40	20	380	0.50	0.50	627	31.64	4.0475	0.5672

Table 2.6 Between Units Variance - $Q_i \propto M_i$.

The results in the previous two tables were obtained by averaging over several different sets of values of the $\{e_i\}$ to avoid the possibility of misleading results from "cancelling out" effects. Clearly no estimator is optimal for all situations on the basis of between-units variance so we now look at several plausible situations where we can calculate the within-units and between-units variances and compare the overall performance of the estimators. As has already been pointed out, the first case considered above, that is that all the particles are contained in a single section, is an extreme case and we do not consider it further (presumably if conditions were such that its occurrence were possible the experimenter would have some idea and be able to act accordingly). The three models we consider are described below and will be referred to hereafter as Models 1, 2 and 3.

Model 1 We assume that the particles are distributed homogeneously throughout the specimen. This implies that the number in each section is approximately proportional to the section size (Table 2.6 applies for between-units variance) and that within each section the joint distribution of the numbers of particles in the individual rectangles of the grid will be multinomial.

Model 2 Models 2 and 3 both assume that the number of particles in each section is approximately constant (Table 2.5 applies for between-units variance). One plausible way in which this might occur is if there is an "inhibition zone" in which there are no particles, for example the "core" of the specimen, thus leaving all the particles distributed around the edge. Then, although sections through the centre will generally contain many more disectors many of them will not be able to contain particles whereas sections near to the edge of the specimen will contain fewer disectors but all will be close to the edge and therefore all may contain particles. Model 2 attempts to describe this situation by assuming that the particles within each section are homogeneously distributed amongst a fixed number, k, of disectors, irrespective of the size of the section.

Model 3 Model 3 is similar to Model 2 in that the number of particles in each section is assumed to be approximately constant but now the particles within each section are assumed to be homogeneously distributed amongst all of the disectors. It is more difficult to imagine a situation where this could occur for all orientations of the section planes but nevertheless in a particular direction it might occur, for example in a spherical specimen where the included particles are attracted to two opposite "poles" as happens with the chromosomes in a cell undergoing division.

With these three models the within-units and between-units variances can be calculated for a selection of the situations considered above (again averaging over different sets of values of the $\{e_i\}$). The between-units variances have already been calculated in Tables 2.5 and 2.6. For the within-units variances we first note that S_{2i}^2 is given by

$$\frac{Q_i}{M_i} \qquad \text{Models 1 \& 3}$$

$$\frac{Q_i}{M_i - 1}\left\{1 - \frac{1}{k} + Q_i\left(\frac{1}{k} - \frac{1}{M_i}\right)\right\} \qquad \text{Model 2}$$

and we substitute these expressions into the original expressions for the within-units variance (before taking expectations over the $\{M_i\}$) and then take expectations where possible (substituting μ_i for M_i where not). The values obtained are shown in Table 2.7 below.

						Within units variance			Between units variance				Total standard deviation			
N	Q	a	b	n	m	$\tilde{Q}_1, \tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$	$\tilde{Q}_1$	$\tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$	$\tilde{Q}_1$	$\tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$
10	15	0.25	0.25	2	50	67.10	74.62	124.70	40.74	15.88	20.35	0.81	10.38	9.11	9.75	11.20
				5	100	41.05	44.06	54.85	10.19	3.99	8.14	0.41	7.16	6.71	7.23	7.43
		0.5	0.5	2	12	122.04	129.62	172.06	40.74	17.77	21.78	3.63	12.76	11.82	12.30	13.26
				5	25	64.58	67.61	74.79	10.18	4.53	8.71	1.74	8.65	8.31	8.74	8.75
40	60	0.25	0.25	8	200	265.23	308.24	495.25	185.16	68.02	93.27	3.73	21.22	18.26	20.04	22.34
				20	400	162.62	179.82	217.63	46.29	17.07	37.31	1.87	14.45	13.41	14.74	14.82
		0.5	0.5	8	50	452.21	486.11	650.84	156.69	66.59	87.12	13.94	24.68	22.78	23.94	25.78
				20	100	256.11	269.67	295.42	39.18	17.00	34.85	6.97	17.18	16.53	17.45	17.39
10	95	0.25	0.25	2	50	424.95	433.45	788.19	672.01	16.90	17.78	0.71	33.12	21.02	21.24	28.09
				5	100	259.98	263.37	346.60	167.96	4.83	7.11	0.36	20.69	16.27	16.45	18.63
		0.5	0.5	2	12	772.94	774.89	1082.01	499.95	37.66	17.50	2.92	35.68	28.47	28.15	32.94
				5	25	409.01	409.79	469.97	124.99	12.88	7.00	1.40	23.11	20.54	20.42	21.71
40	380	0.25	0.25	8	200	1679.79	1724.83	3127.49	512.30	50.55	84.44	3.38	46.82	41.60	42.54	55.95
				20	400	1029.90	1047.91	1373.75	128.09	12.86	33.78	1.69	34.03	32.29	32.89	37.09
		0.5	0.5	8	50	2864.00	2891.04	4091.93	512.32	66.64	79.11	12.66	58.11	54.14	54.50	64.07
				20	100	1622.00	1632.82	1855.96	128.06	17.95	31.64	6.33	41.83	40.50	40.80	43.15

Table 2.7a Variance Under Model 1

						Within units variance			Between units variance				Total standard deviation			
N	Q	a	b	n	m	$\tilde{Q}_1, \tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$	$\tilde{Q}_1$	$\tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$	$\tilde{Q}_1$	$\tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$
10	15	0.25	0.25	2	50	67.52	77.37	128.19	11.11	13.93	23.24	0.93	8.87	9.02	10.03	11.36
				5	100	41.30	47.33	56.55	2.78	3.50	9.30	0.47	6.64	6.69	7.53	7.55
		0.5	0.5	2	12	124.85	133.38	187.50	11.11	12.71	17.98	3.00	11.66	11.73	12.30	13.80
				5	25	66.06	70.58	82.02	2.78	3.27	7.20	1.44	8.30	8.33	8.82	9.14
40	60	0.25	0.25	8	200	284.13	340.83	541.55	41.03	47.12	113.19	4.53	18.03	18.20	21.31	23.37
				20	400	174.20	208.97	238.64	10.26	11.84	45.29	2.26	13.58	13.64	15.95	15.52
		0.5	0.5	8	50	501.90	537.72	769.60	41.03	43.28	78.17	12.51	23.30	23.35	24.82	27.97
				20	100	284.25	304.54	351.50	10.26	11.17	31.28	6.26	17.16	17.19	18.33	18.92
10	95	0.25	0.25	2	50	561.10	653.06	1065.35	11.11	511.47	1069.17	42.77	23.92	32.75	41.50	33.29
				5	100	343.27	399.53	469.96	2.78	128.58	427.61	21.38	18.60	21.72	28.76	22.17
		0.5	0.5	2	12	1549.03	1655.83	2326.45	11.11	418.04	737.62	122.94	39.50	44.35	48.92	49.49
				5	25	819.69	876.21	1017.69	2.78	108.07	295.32	59.06	28.68	30.46	34.23	32.81
40	380	0.25	0.25	8	200	3598.53	4413.06	6858.82	41.03	2006.49	6902.14	276.09	60.33	74.87	106.37	84.47
				20	400	2206.29	2705.69	3022.38	10.26	504.17	2761.53	138.08	47.08	52.06	73.94	56.22
		0.5	0.5	8	50	8850.11	9646.47	13570.57	41.03	1662.52	4340.88	694.54	94.29	102.53	118.27	119.44
				20	100	5012.18	5463.19	6198.16	10.26	429.70	1736.78	347.36	70.87	73.77	84.85	80.90

Table 2.7b Variance Under Model 2

						Within units variance			Between units variance				Total standard deviation			
N	Q	a	b	n	m	$\tilde{Q}_1, \tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$	$\tilde{Q}_1$	$\tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$	$\tilde{Q}_1$	$\tilde{Q}_2$	$\tilde{Q}_3$	$\tilde{Q}_4$
10	15	0.25	0.25	2	50	67.10	54.70	123.90	11.11	13.93	23.24	0.93	8.84	9.00	8.83	11.17
				5	100	41.05	36.09	54.45	2.78	3.50	9.30	0.47	6.62	6.67	6.74	7.41
		0.5	0.5	2	12	122.04	114.05	169.46	11.11	12.71	17.98	3.00	11.54	11.61	11.49	13.13
				5	25	64.58	61.38	73.54	2.78	3.27	7.20	1.44	8.21	8.24	8.28	8.66
40	60	0.25	0.25	8	200	265.23	186.57	490.38	41.03	47.12	113.19	4.53	17.50	17.67	17.31	22.25
				20	400	162.62	131.15	215.19	10.26	11.84	45.29	2.26	13.15	13.21	13.28	14.75
		0.5	0.5	8	50	452.21	406.11	638.03	41.03	43.28	78.17	12.51	22.21	22.26	22.01	25.51
				20	100	256.11	237.66	289.02	10.26	11.17	31.28	6.26	16.32	16.35	16.40	17.18
10	95	0.25	0.25	2	50	424.95	306.73	783.12	11.11	511.47	1069.17	42.77	20.88	30.60	37.09	28.74
				5	100	259.98	212.69	344.06	2.78	128.58	427.61	21.38	16.21	19.71	25.30	19.12
		0.5	0.5	2	12	772.94	690.97	1068.03	11.11	418.04	737.62	122.94	28.00	34.51	37.80	34.51
				5	25	409.01	376.22	463.25	2.78	108.07	295.32	59.06	20.29	22.74	25.91	22.85
40	380	0.25	0.25	8	200	1679.79	909.92	3094.90	41.03	2006.49	6902.14	276.09	41.48	60.72	88.39	58.06
				20	400	1029.90	721.95	1357.45	10.26	504.17	2761.53	138.08	32.25	39.17	59.02	38.67
		0.5	0.5	8	50	2864.00	2377.32	4009.73	41.03	1662.52	4340.88	694.54	53.90	67.28	81.97	68.59
				20	100	1622.00	1427.33	1814.87	10.26	429.70	1736.78	347.36	40.40	45.30	56.25	46.50

Table 2.7c Variance Under Model 3

The results of the simulations give support to using the simplest estimator, $\tilde{Q}_1$. In only a few cases is it bettered by any of the others in terms of overall variance and then not by very much. Although $\tilde{Q}_4$ has a very small between-units variance in many cases, the within-units variance tends to dominate and there would be little justification for using $\tilde{Q}_4$ on the basis of this evidence, especially as it is a more complicated estimator to realise in practice. The ratio estimator, $\tilde{Q}_2$, performs well overall and is based on the same sampling design as $\tilde{Q}_1$, thus making it simple to implement. However, it is in general biased and, although the comparison has been made using MSE, the desirability of unbiasedness expressed in the stereological literature suggests that few practicioners would advocate its use over $\tilde{Q}_1$.

Many questions still need to be asked. First, approximations have been used for the variance of the number, M_i, of disectors hitting a given section and although they are probably quite good they only apply to convex sections (and hence convex specimens). There appears to be no simple extension to non-convex specimens and even for convex specimens we have no expressions for the moments of the $\{M_i\}$ other than the first two. Secondly, the calculations presented are for a spherical specimen and only for certain specific values of the various parameters involved, such as the number of particles, number of sections, etc., although these values were chosen to represent a wide range of circumstances. How well these results reflect the general case of an arbitrary specimen is difficult to judge, although the specific shape of the sections only enters through the variance of the $\{M_i\}$. Thirdly, the assessment of the estimators has been from a "design-based" point of view, that is, how well they perform under the probabilistic structure induced by a randomised sampling scheme. In practice there may be extra knowledge available or restrictions imposed which render this assessment inappropriate, for example in the presence of strong heterogeneity, anisotropy or periodicity of which the experimenter has some knowledge he may be able to construct a sampling scheme and/or estimator which performs better in those particular circumstances.

What has been achieved is a proper theoretical framework for sampling schemes involving the disector which are realisable in practice whilst still upholding the essential properties of the disector and which allow the application of standard theory to examine their second order properties. Some of the standard deviations in Table 2.7 are alarmingly high and highlight the fact that unbiasedness alone is not sufficient to ensure accurate estimation; indeed the indications are that fairly large sampling fractions are needed to obtain reasonable accuracy when estimating particle number using the disector.

Having obtained an estimate of the number of particles in the population the next stage is to estimate the mean volume of the particles. The most economical way to do this would be to use the same particles that are counted by the disectors to estimate the mean volume, as with Cruz-Orive's selector (1987), since they are sampled from the population with uniform probability. Continuing with the same partition as described above we now let $w \in \mathbb{Z}$ be the smallest integer such that wd is greater than the maximum caliper diameter of any particle and whenever $D_{ij}(\pm)$ is included in the sample we also observe $T_{i\mp1},...,T_{i\mp w}$[1]. Thus for any particle counted by a disector we observe a "complete" set of parallel planar profiles with separation d. We know that the total area of the profiles multiplied by d is an unbiased estimator of the volume of the particle and there are many ways of estimating the profile areas. For example, we could incorporate into the rectangular grid with which we tessellate the planes $T_1,...,T_N$ a lattice of points; the number of those points falling inside a particle profile, suitably normalized, would provide an estimate of its area.

Cruz-Orive (1987) is concerned with the case when d is not known and the separation between planes is not constant. This means that we need a direct estimator of volume, which he derives from the results of geometric probability relating expectations of powers of lengths of random chords to global characteristics such as

[1] We need to modify this definition slightly when $i-w<1$ or $i+w>N$ but the principle remains the same.

surface area and volume. There are some non-trivial problems here in defining a precise probabilistic structure for the sampling mechanism which can be realised in practice; Cruz-Orive leaves several questions unanswered in his account and we examine briefly below some points for consideration.

The foundation on which the approach is based is the three dimensional version of Crofton's Theorem, namely that for an *IUR* line probe T through a convex domain Y the length of intersection, $L(Y \cap T)$ satisfies

$$E([L(Y \cap T)]^4) = 12[V(Y)]^2/\pi S(Y)$$
$$E(L(Y \cap T)) = 4V(Y)/S(Y),$$

where V and S denote volume and surface area respectively, giving

$$\frac{E([L(Y \cap T)]^4)}{E(L(Y \cap T))} = \frac{3V(Y)}{\pi}.$$

The first result can be found in Santaló (1976, p.237) although Cruz-Orive gives an alternative proof and generalises to the non-convex case. The second result is Cauchy's formula, which can be proved very simply under mild regularity conditions on the surface of Y (basically that it is piecewise smooth). The next important step is to notice that if T^* is a length-weighted line probe, that is the density of T^* is proportional to its length, then the densities of $L(Y \cap T)$ and $L(Y \cap T^*)$ are related by

$$f_{T^*}(l) = l f_T(l)/E_T(L)$$

and hence

$$E_{T^*}(L^3) = E_T(L^4)/E_T(L)$$
$$= 3V(Y)/\pi .$$

Thus if we can generate length-weighted probes through a particle Y then we can form a direct estimate of $V(Y)$.

We note the following:

1 A length-weighted probe through Y can either be generated as a length-weighted probe in the profile of intersection of an area-weighted planar probe through Y or as a line with uniformly random orientation (in $\mathbb{R}^3$) through a point chosen uniformly at random within Y (for details see Davy & Miles, 1977, Proposition 2 and Coleman, 1979, Theorem 3.3).

2 Describing an orientation in $\mathbb{R}^3$ relative to a fixed frame $Oxyz$ by standard spherical polar co-ordinates (φ,θ) we have that a line making an angle Θ with Oz contained in a plane whose normal has orientation $(\Phi,\pi/2)$ where Θ has the density $\sin\theta$ $(0\leq\theta\leq\pi/2)$ and Φ is uniformly random on $(0,2\pi]$ has uniformly random orientation in $\mathbb{R}^3$ (see Fig. 2.10). This observation of Baddeley (Baddeley, Gundersen & Cruz Orive, 1986) is the basis of the so-called *vertical sections*.

3 Given a plane whose normal has a fixed, arbitrary orientation ω_0 and whose distance from a fixed point is uniformly random and given a uniformly random point P on that plane then P is conditionally uniformly random in Y given that it lies inside Y.

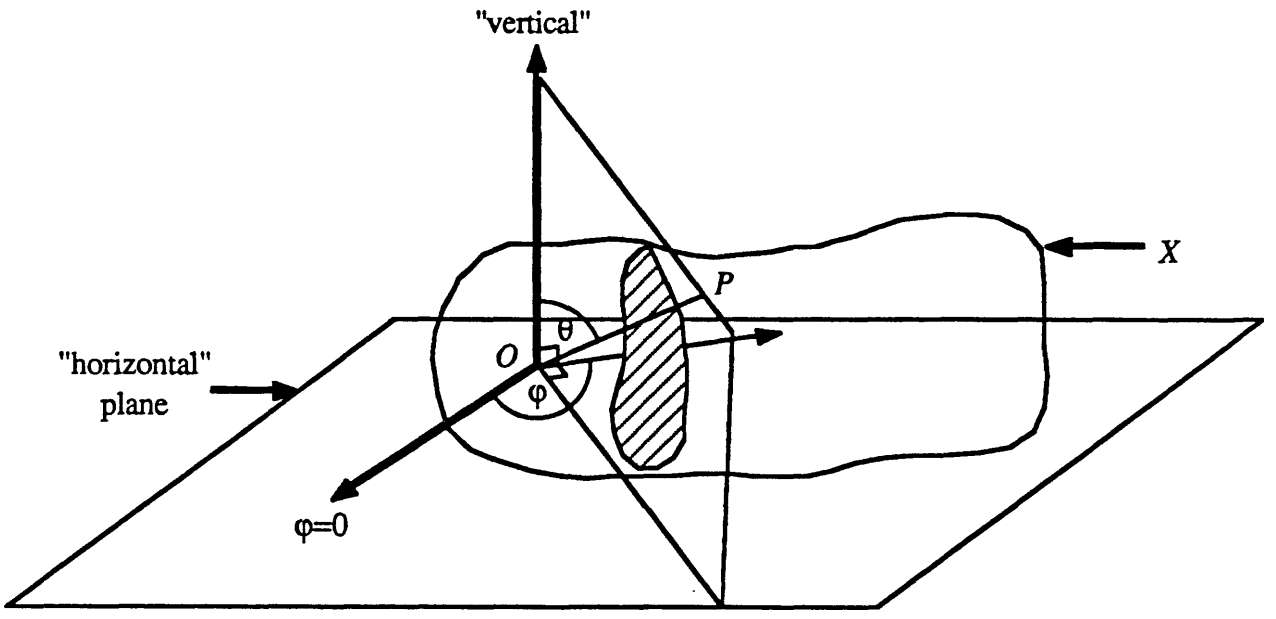


Fig. 2.10 Construction of Lines with Uniformly Random Orientation in $\mathbb{R}^3$ Using "Vertical" Sections.

The underlying principle for this method of generating lines with uniformly random orientation is the fact that every line in $\mathbb{R}^3$ is contained in a vertical plane. Therefore if vertical planes are generated first and lines are generated within them with the correct density the lines will have the desired properties. In practice the drawback is that if more than one plane is to be generated then they must be parallel and the line probes will not be independent.

We can now see how to proceed in the context of the selector.

1 Take an arbitrary plane through an arbitrary origin of co-ordinates, O, as the "horizontal" and fix an arbitrary direction in the horizontal plane as $\varphi=0$.

2 Generate a uniformly random Φ from $(0,2\pi]$ and take a planar section of X uniformly at random from all of those with normal parallel to $(\Phi,\pi/2)$ which intersect X.

3 Generate points uniformly at random within the planar profile of X formed by this section.

4 For each point falling inside the profile of the particle Y generate a Θ from the density $\sin\theta$ $(0\leq\theta\leq\pi/2)$ and form the line making an angle Θ to the vertical (see fig. 2.11).

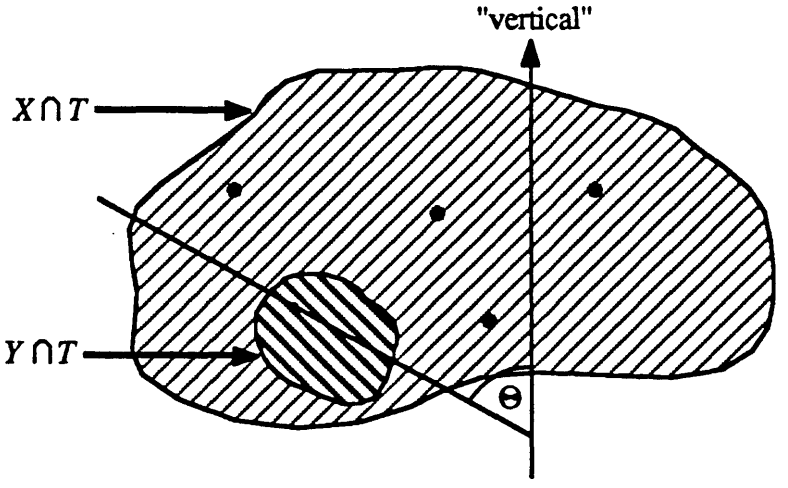


Fig. 2.11 Construction of Length-Weighted Lines Through a Particle.

If a set of uniformly random points is generated within X then those that hit Y are uniformly random in Y. Thus if we take a line with uniformly random orientation through each of the points that hit Y then the lines will be length-weighted probes through Y.

There are obvious drawbacks to this procedure. First we notice that although an individual point falling inside Y is marginally uniform within Y the set of points that fall inside Y are clearly not independent since they all lie in the same plane; this event occurs with probability zero for a set of independent, uniformly random points in Y and clearly a set of coplanar points is unsatisfactory. We can improve on this by ensuring that any particle to be analysed is intersected by several planes but unfortunately, as specimens cannot be reconstructed after sectioning, we have to restrict the planes to be parallel. Although a set of points generated on a set of parallel planes still represents an event of probability zero for independent, uniformly random points in Y it is intuitively far more acceptable than the original suggestion. Our problem now is that the line probes, although marginally length-weighted, are again highly correlated as they all lie perpendicular to a single orientation, again representing an event of probability zero for independent, length-weighted probes. The potentially enormous variance that could result for markedly anisotropic particles is worrying, particularly if all the particles are grouped around a common preferred direction. Cruz-Orive's

answer to this problem is to regard each selector as a small block sampled from the specimen with one pair of faces parallel to the horizontal plane and another pair of faces having normal with orientation $(\Phi_i,0)$ where each Φ_i is an independent uniform random variable on $(0,2\pi]$. The set of parallel planes making up the selector is then taken parallel to this second pair of sides and over a set of selectors the potentially disastrous effects of aligned, anisotropic particles are greatly reduced. However, in practice it is difficult to see how each block should be located with the correct probabilistic structure without destroying the specimen to the extent that subsequent blocks cannot be obtained. A second possibility is to divide the specimen into several blocks first and then obtain selectors from each block after they have been given independent, random orientations. This raises "edge-effect" problems; how do we identify disjoint parts of particles that are divided between two blocks and estimate the volume of those particles? Also we greatly increase the probability of having a selector which does not lie wholly within the specimen since we are increasing the surface area to volume ratio of the specimen and therefore we are losing information, since selectors which lie partly outide the specimen are necessarily sampling less of it. Apart from these reservations it is also evident that the procedure outlined so far would be extremely time consuming to put into operation with so many independent points and lines to be generated and located on the specimen. Cruz-Orive's answer here is to use a grid of equally spaced, parallel lines, each containing a set of equally spaced points. On each plane of a selector the grid is dropped at random but with the lines having orientation Θ, where Θ has the same distribution as before and is independent from one plane to the next. Whilst the points of the grid which lie inside Y are still marginally uniformly random in Y they are now highly correlated; furthermore the line probes on a given plane all have exactly the same orientation, although they too still have the correct marginal distribution. However, the most serious flaw is that we are no longer generating length-weighted probes. The set of lines that we end up with are marginally *IUR* lines (although still correlated) to which we apply random weights whose expected values are the lengths of the lines. Since we must then divide by the sum of

the weights it is clear that what Cruz-Orive has suggested is in fact no more than a (biased) ratio estimator derived under *IUR* sampling.

We can write his estimator as

$$\hat{V}_{CO} = \frac{\sum_{i=1}^{m} n_i L_i^3}{\sum_{i=1}^{m} n_i}$$

where n_i is the number of points on the ith line probe intersecting Y and L_i is the length of intersection of that line with Y. The $\{n_i\}$ are conditionally independent, given $\{L_i\}$, with

$$E(n_i \mid \{L_i\}) = L_i$$

and

$$\text{var}(n_i \mid \{L_i\}) = e_i(1-e_i)$$

where $e_i = L_i/d_p - [L_i/d_p]$, d_p is the distance between points on the same line and [.] denotes "the integer part of".

The sole reason for Cruz-Orive's elaborate sampling scheme and estimator is the supposition that distances between sections are unknown and not necessarily constant. The price we have to pay is a complicated sampling mechanism, which is not clearly defined in theory and poses several problems in terms of practical realisation, and a biased estimator based on highly dependent variables. Furthermore, the denominator is itself the basis for a far simpler, more robust (in terms of anisotropy) estimator for which we lack only the knowledge of distances between sections. An alternative approach therefore would be to form an estimate of volume based on point counts taken on parallel planes and to incorporate an estimate of the mean distance between sections. Unless there is a very large variation in between-section distances this would surely be a preferable and less sensitive method since methods based on areas of intersection with parallel planes are conditionally unbiased given a fixed orientation for the planes whereas methods based on Crofton's Theorem (or its higher dimensional analogues) applied to a set of parallel line probes are only unbiased when

integrated over the distribution of the orientation of the lines. In terms of variance if we consider the general formula

$$\mathrm{var}(X) = \mathrm{var}_Y[E(X|Y)] + E_Y[\mathrm{var}(X|Y)]$$

then the first term on the right hand side will be zero in the former case but not in the latter.

We conclude this section with some remarks concerning the use of ratio estimators. Ratio estimators are commonly advocated in the stereological literature as a means of variance reduction; that is, even when we know $V(X)$ it may be advantageous to estimate the ratio $V(Y)/V(X)$ by $A(Y \cap T)/A(X \cap T)$ and then multiply by the known $V(X)$ to estimate $V(Y)$. This will be particularly beneficial when $A(Y \cap T)$ varies greatly between sections and $A(Y \cap T)$ and $A(X \cap T)$ are highly positively correlated but in practice it may not be possible. Miles (1978a) gives an excellent account of the underlying theory; here we state the main results. If we take a number of planar sections $T_1,...,T_n$ then

$$\sum_{i=1}^{n} \frac{A(Y \cap T_i)}{A(X \cap T_i)}$$

is only unbiased for $V(Y)/V(X)$ with area-weighted sections, which are difficult, if not impossible to generate. On the other hand

$$\frac{\Sigma A(Y \cap T_i)}{\Sigma A(X \cap T_i)}$$

is biased for *IUR* sections but the bias may be negligible when the denominator is almost constant (for example when n is large); in that case the estimation will be practically the same as direct estimation of $V(Y)$ so we will have gained very little. If

one section is area-weighted and the rest are *IUR* then the second estimator described above is unbiased for $V(Y)/V(X)$ but again we have the problem of generating an area-weighted section. In all of the above cases any implementation in practice would necessitate restricting the sections to be parallel; the estimators are still valid but it seems unlikely that in general they will perform better than those based on systematic sections.

2.6 Extended Case

Again we adopt the terminology of Miles (1978a) and use "extended case" to mean any situation where it is necessary to sub-sample from planar sections through a deterministic specimen. However, his theoretical treatment allows independent *IUR* planar probes to be generated within an opaque reference phase whereas here we take a more practical stance and assume a two stage sampling scheme where the first stage involves the generation of planar sections through the reference phase and the second requires the sub-sampling by quadrats of the profiles of intersection. Thus in many ways this case is similar to the previous one, the main difference being in the way that we measure $A(Y \cap T)$. To a large extent the estimation of $A(Y \cap T)$ in this context is similar to the estimation of $V(Y)$ itself but with the considerable bonus that we can see the interior of the profile of intersection of the reference phase with a planar probe without having to destroy it, thus enabling us to generate genuinely independent repetitions of sampling procedures and achieve sampling with probability proportional to included area. Therefore techniques that would theoretically be superior but which are not feasible practically in the case of volume estimation now become a reality.

First we define some terms to describe this situation. We assume that our starting point is a specimen X_0 which can be isolated from the rest of the sampling unit and which is of a similar structure to the Y of §2.3. We may then have a number of "nested" phases, $X_1,\ldots,X_n$, satisfying $X_0 \supset X_1 \supset \ldots \supset X_n \supset Y$ where the structure of the intermediate phases may be quite general and the levels of magnification required to measure them may vary, although naturally the magnification required for any given phase must be at least as great as that for all the phases which contain it.

Our method is to take planar sections through X_0 and then to sub-sample small regions from the profiles of intersection for magnification and measurement of Y. Clearly, the most efficient scheme will sub-sample only from X_n since we know that there is no phase of interest outside that phase and therefore we will regard X_n as a "local" reference phase and for brevity and consistency with previous sections refer to it simply as X.

We start our discussion with the case when X is measurable accurately without magnification. Here we need to consider the structure of X and the implications in terms of the probability of a random quadrat hitting the boundary of X given that it hits X. For convex sets there are a few results in integral geometry that can be applied. The measure of the set of positions of a convex set K_1 (integrating over all rotations and translations) such that it is contained within a fixed convex set K_0 is given by

$$2\pi(A_0 + A_1) - B_0B_1$$

when the boundaries of both sets have continuous radii of curvature and when the greatest radius of curvature of the boundary of K_1 is less than the least radius of curvature of the boundary of K_0. (A_0 is the area of K_0 and B_0 is the length of the boundary of K_0 , etc.) Also the measure of the set of positions of K_1 such that it intersects K_0 is given by

$$2\pi(A_0 + A_1) + B_0B_1.$$

(Both results can be found in Santaló, 1976, pp.94-95). Thus the probability of the boundary of K_1 intersecting the boundary of K_0 given that the two sets intersect is

$$1 - \frac{2\pi(A_0 + A_1) - B_0B_1}{2\pi(A_0 + A_1) + B_0B_1}$$

$$= \frac{2B_0B_1}{2\pi(A_0 + A_1) + B_0B_1}.$$

This result is not applicable to rectangular quadrats as their boundaries have infinite radii of curvature but it could be applied to circular "quadrats". However, an alternative formula of some use is Poincaré's formula (Santaló, 1976, p.111) which states that the measure of the set of locations of a set K_1 such that its boundary intersects the boundary of a second set K_0 , when weighted with respect to the *number* of intersections in each such location, is equal to 4 times the product of the boundary lengths. If both sets are convex it is clear that the number of intersections in every location is two and therefore the measure of the set of locations in which the boundaries

of the two sets intersect is twice the product of the boundary lengths. As a specific example of the implication of these formulae a circular quadrat of radius r dropped at random on a circular planar set of radius R has a probability $4rR/(r+R)^2$ of intersecting the boundary given that it intersects the set. Even when $R = 10r$ the probability of intersecting the boundary is only just under 1/3, even though the ratio of the areas is 1/100. Since the circle is the shape with the shortest boundary length for a given area this suggests that in general not only must the area of the quadrat be very small relative to the area of the specimen but also that the area-to-perimeter ratio of the specimen must be high in order that we can regard the probability of the quadrat hitting the boundary as being negligible.

With this warning in mind we identify two cases of interest; case A, where the probability of the quadrat intersecting the boundary of the specimen is "small", that is, with high probability $A((X \cap T) \cap Q) = A(Q)$ (where $X \cap T$ is a planar section through X and Q is the quadrat) and case B, where there is a significant probability that the quadrat intersects the boundary of the specimen and hence that $A((X \cap T) \cap Q) \neq A(Q)$. These two cases correspond largely to Models I and II of Cruz-Orive (1980).

Finally we briefly discuss the generalisation to the situation where X is not measurable accurately without magnification but is measuable accurately at a significantly lower magnification than Y.

The problem of area estimation using quadrats has been dealt with extensively in the literature; see particularly Miles & Davy (1977) and Cruz-Orive & Weibel (1981). Miles & Davy define an *FUR* quadrat, Q, to be one whose orientation is fixed and whose location is uniformly random over all positions such that $(X \cap T) \cap Q \neq \varnothing$. (A more rigorous definition is available in their paper). With this probabilistic structure we find that

$$E[A((Y \cap T) \cap Q)] = \frac{A(Y \cap T)\, A(Q)}{A((X \cap T) \oplus Q)}$$

and hence that

$$A((Y\cap T)\cap Q)\left(\frac{A((X\cap T)\oplus Q)}{A(Q)}\right)$$

is an unbiased estimator of $A(Y\cap T)$ (where $\oplus$ denotes dilation of $X\cap T$ by the structuring element Q). However, if there is a significant probability of Q having a location such that $A((X\cap T)\cap Q)\neq A(Q)$ then, as with volume estimation, it is well known that a superior approach is to estimate $A(Y\cap T)$ via the ratio $A((Y\cap T)\cap Q)/A((X\cap T)\cap Q)$ since denominator and numerator will tend to be positively correlated, thus reducing the variance. The quadrats must now be located with probability proportional to $A((X\cap T)\cap Q)$ to achieve unbiasedness (further improving the estimator as quadrats containing little of X will be less likely) which is a realistic proposition in two dimensions. To estimate $A(Y\cap T)$ we must then have an independent estimate of $A(X\cap T)$, which we can obtain, for example by point counting, from the whole section, since $X\cap T$ can be measured without magnification. We note that there is little difference between these two approaches in case A when Q is small compared to $X\cap T$ since $A((X\cap T)\oplus Q)$ will be close to $A(X\cap T)$ and $A((X\cap T)\cap Q)$ will equal $A(Q)$ with high probability. In contrast the second approach should generally be superior in case B. With both methods we may generate independent repetitions, simply taking the arithmetic mean of the estimates obtained from each position of the quadrat.

As with volume estimation an alternative to independent repetitions is to take a systematic set of quadrats. If the set is fixed and bounded then we must regard it as a single quadrat of complex shape and giving it a location taken uniformly at random from those for which it "hits" $X\cap T$ will result in a loss of efficiency since the expected value of $\sum A((X\cap T)\cap Q)$ will be less than for the same number of independently, uniformly located quadrats. (This follows from the fact that the dilation of $X\cap T$ by the set of quadrats will be greater than the dilation by a single quadrat, a fact that is obvious when one considers that there are positions possible where all but one of the quadrats making up the set lie outside X.) However, the set of quadrats can

be located with probability proportional to total included area (of X) giving

$$\frac{\Sigma A((Y \cap T) \cap Q)}{\Sigma A((X \cap T) \cap Q)}$$

as an unbiased estimator of

$$\frac{A(Y \cap T)}{A(X \cap T)} .$$

(See Miles & Davy, 1977 for details of the construction.)

This estimator (or its analogue) also arises for both volume estimation and area estimation as a weighted version of

$$\sum \frac{A((Y \cap T) \cap Q)}{A((X \cap T) \cap Q)} ,$$

that is with weight $A((X \cap T) \cap Q)/\Sigma A((X \cap T) \cap Q)$ attatched to the contribution $A((Y \cap T) \cap Q)/A((X \cap T) \cap Q)$ under the intuitive hypothesis that more information about Y is contained in quadrats containing more of X. Unfortunately, to be unbiased this estimator can no longer be based on independent area-weighted quadrats (the only practical method for independent quadrats is to take one area-weighted and the rest *FUR*; see Miles, 1978b for the details of this, based on a result of Midzuno) and therefore it is not clear whether there is any overall gain over the unweighted

$$\sum \frac{A((Y \cap T) \cap Q)}{A((X \cap T) \cap Q)}$$

based on area-weighted quadrats. One further possibility is to take an "unbounded" systematic set of quadrats. Two approaches, which will not differ much in practice when the individual quadrat areas are small relative to $A(X \cap T)$, are either to define a regular grid of quadrats which is located uniformly at random on $X \cap T$ or to define a complete partition of the plane consisting of k sets of quadrats (where $1/k$ is the fraction to be sampled) and use the set in which a uniformly random point of $X \cap T$ lies (see Cruz-Orive & Weibel, 1981 or Miles, 1978b); we consider only the former case as there will be so little difference in our context. We summarise the estimators described above in Table 2.8.

Sampling scheme		Estimator	
n independent FUR quadrats	$\hat{A}_1$	$\left[\frac{1}{n}\sum_{i=1}^{n} A(Y_T \cap Q_i)\right] \frac{A(X_T \oplus Q)}{A(Q)}$ [1]	Unbiased
n independent FUR quadrats	$\hat{A}_2$	$\frac{\sum_{i=1}^{n} A(Y_T \cap Q_i)}{\sum_{i=1}^{n} A(X_T \cap Q_i)} A(X_T)$	Biased
1 area-weighted quadrat and n-1 independent FUR	$\hat{A}_3$	$\frac{\sum_{i=1}^{n} A(Y_T \cap Q_i)}{\sum_{i=1}^{n} A(X_T \cap Q_i)} A(X_T)$	Unbiased
n independent area-weighted quadrats	$\hat{A}_4$	$\left[\sum_{i=1}^{n} \frac{A(Y_T \cap Q_i)}{A(X_T \cap Q_i)}\right] A(X_T)$	Unbiased

Table 2.8 Area Estimators for Two-Phase Specimens Based on Quadrat Sampling.

[1] X_T and Y_T are shorthand for $X \cap T$ and $Y \cap T$ respectively.

Sampling scheme		Estimator	
bounded, systematic set of n quadrats (FUR)	$\hat{A}_5$	$\left[\frac{1}{n}\sum_{i=1}^{n} A(Y_T \cap Q_i)\right] \frac{A(X_T \cap Q_{TOT})}{A(Q_{TOT})}$ [1]	Unbiased
bounded, systematic set of n quadrats (FUR)	$\hat{A}_6$	$\frac{\sum_{i=1}^{n} A(Y_T \cap Q_i)}{\sum_{i=1}^{n} A(X_T \cap Q_i)} A(X_T)$	Biased
bounded, systematic set of n quadrats (area-weighted)	$\hat{A}_7$	$\frac{\sum_{i=1}^{n} A(Y_T \cap Q_i)}{\sum_{i=1}^{n} A(X_T \cap Q_i)} A(X_T)$	Unbiased
unbounded, systematic lattice of quadrats	$\hat{A}_8$	$\left[\frac{1}{n}\sum_{i=1}^{n} A(Y_T \cap Q_i)\right] \frac{A(X_T \oplus Q)}{A(Q)}$	Unbiased
unbounded, systematic lattice of quadrats	$\hat{A}_9$	$\frac{\sum_{i=1}^{n} A(Y_T \cap Q_i)}{\sum_{i=1}^{n} A(X_T \cap Q_i)} A(X_T)$	Biased

Table 2.8 (cont.) Area Estimators for Two-Phase Specimens Based on Quadrat Sampling.

[1] Q_{TOT} is the union of the quadrats regarded as a single non-connected planar set.

There are no general quantitative results in the literature concerning the second order properties of these estimators; as is so often the case in stereology these properties depend on the geometric structure of the specimen. However, various qualitative observations can be made, some of which have been noted already. In the framework we have developed we are supposing that subsampled quadrats are being used to estimate $A(Y \cap T)$ on a given plane T because Y cannot be measured accurately without magnification. This implies that the quadrats are necessarily small compared with the sectional area and therefore that remarks pertaining to the case of $A(Q)$ being small relative to $A(X \cap T)$ apply. Thus we expect little difference between $\hat{A}_1, \hat{A}_2, \hat{A}_3$ and $\hat{A}_4$ in case A although in case B we should expect $\hat{A}_2$ to be preferable to $\hat{A}_1$ and both $\hat{A}_3$ and $\hat{A}_4$ to be preferable to $\hat{A}_2$.

On the other hand $\hat{A}_5$ and $\hat{A}_6$ should still be avoided whereas the behaviour of $\hat{A}_7, \hat{A}_8$ and $\hat{A}_9$ may be highly dependent on the arrangement of Y within $X \cap T$. For instance, if Y is very "clustered" then $\hat{A}_7$ may exhibit a large variance, resulting from the possibility of hitting a cluster or missing Y entirely, whereas, unless the clusters are periodic with the same period as the lattice of the unbounded set of quadrats we should expect $\hat{A}_8$ and $\hat{A}_9$ to be more efficient. Whatever the arrangement of Y we would expect little difference between $\hat{A}_8$ and $\hat{A}_9$ in case A when $A(Q)$ is small relative to $A(X \cap T)$ but this would not necessarily still be true in case B.

To conclude this discussion of area estimation by quadrats we examine a simulation performed on three synthetic planar specimens. In order to make calculations feasible without requiring unrealistic amounts of computer time the specimens were constructed with circles. The reference phase, X[1], was taken to be a circle of radius 20 in all cases and the phase of interest, Y, was taken to be the realisation of a stationary, hard core process of circles whose radii were uniform on (0,1). The hard core process was produced by generating a Poisson point process in the plane (representing the centres of the circles) and attaching a "mark" to each point

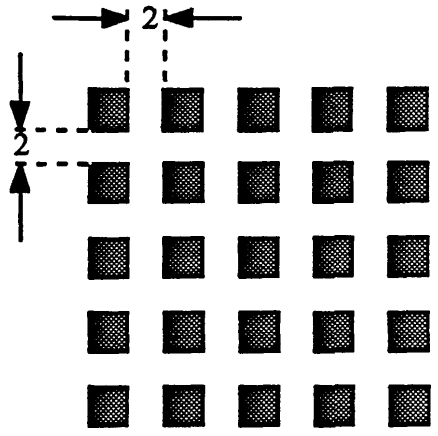
[1] We drop the notation $X \cap T$ since here we are interested only in the estimation of area in the plane.

(representing the radius), where the mark was an independent observation from a uniform distribution on (0,1). Each point was also given an independent, uniformly random "birth-time" and where circles overlapped the younger one was deleted; the only difference between the three specimens was in the intensity of the Poisson point process. (Detailed information relating to the specimens is shown in Table 2.9.) Table 2.10 shows the performance of $\hat{A}_1, \hat{A}_2, \hat{A}_4, \hat{A}_7$ and $\hat{A}_9$ in 100 independent repetitions with n=25[1] using a square quadrat of side 2. The lattices for $\hat{A}_7$ and $\hat{A}_9$ are shown in Fig. 2.12.

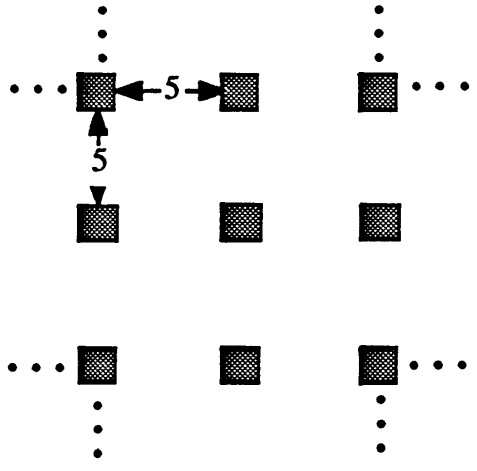
As expected, $\hat{A}_7$ performed significantly worse than the rest of the estimators, even though the set of quadrats was area weighted. A simple geometrical argument shows that the probability that the set of quadrats lies wholly inside X is 0.08, and therefore nearly all of the time at least some part of the set of quadrats lies outside the specimen, often with whole quadrats having no intersection with the specimen at all (when the set of quadrats is not area weighted the corresponding probability drops to approximately 0.02). The performances of $\hat{A}_1$ and $\hat{A}_2$ are interesting in that they highlight the fact that unbiasedness alone is not a good criterion for judging an estimator. In this case the bias of $\hat{A}_2$ is undetectable in comparison with the variance and in terms of mean squared error $\hat{A}_2$ comes out slightly better than $\hat{A}_1$. The best estimators in this simulation appear to be $\hat{A}_4$ and $\hat{A}_9$ although for the second specimen $\hat{A}_9$ breaks the pattern established by the rest of the results. This simulation was repeated to check that the result was not erroneous and the most likely conclusion must be that the specimen exhibits some form of periodicity. Certainly overall $\hat{A}_4$ performs best of all the estimators but since $\hat{A}_9$ is far easier to realise in practice it might be preferred, provided that the experimenter is confident that there is no periodicity or regularity in the specimen that might inflate the variance.

---

[1] For $\hat{A}_9$ n is a random variable but the lattice was constructed so that the expected value of n was close to 25.



(a) Bounded lattice of 25 quadrats.



(b) Part of unbounded, systematic lattice of quadrats.

Fig. 2.12 Systematic Sets of Quadradts for Area Estimation.

The two patterns of systematic quadrats shown here differ in that the top one is a bounded lattice whose location must be taken from all those positions for which it has a non-void intersection with the reference phase whereas for the unbounded lattice only translations up to the distance between adjacent quadrats need be considered.

Specimen	Intensity of Poisson process	No. of circles	Area of Y
1	0.1	98	85
2	0.2	180	173
3	0.5	319	268

Table 2.9 Details of Specimens Used in Simulation.

	Specimen 1		Specimen 2		Specimen 3	
Estimator	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
$\hat{A}_1$	82	31	171	44	268	52
$\hat{A}_2$	83	30	173	40	266	48
$\hat{A}_3$	83	27	168	38	267	41
$\hat{A}_7$	86	47	175	58	268	68
$\hat{A}_9$	86	25	173	51	276	43

Table 2.10 Variance of Quadrat Estimators.

The generalisation of the preceeding discussion to the situation where there is an intermediate phase between X_0 and Y which cannot be measured without magnification but which requires a lower magnification than Y itself is fairly straightforward, at least in theory. We started this section with a general model of nested phases, $X_1,...,X_n$, satisfying $X_0 \supset X_1 \supset ... \supset X_n \supset Y$, where X_0 is the specimen itself and is assumed to be measurable accurately without magnification and Y is the phase of interest. We will be interested in the subset of phases $X_0 \supset X_1^* \supset ... \supset X_m^* \supset Y$ with the property that the optimal levels of magnification, $M_1^*,...,M_m^*$, for measuring the phases in the subset are all significantly different (and different from those for X_0 and Y). Where several phases in the original set have the same optimal level of magnification we retain the "innermost" phase in the reduced set to avoid unnecessary measurement of material that we know not to contain any of the phase of interest. Now if we consider a multi-stage sampling scheme where at each stage we estimate the ratio of the volumes of two "adjacent" phases we can form an estimator of the volume of Y as

$$\hat{V}(X_0)\,\hat{R}_{10}\,\hat{R}_{21} \ldots \hat{R}_{m(m\text{-}1)}\,\hat{R}$$

where $\hat{R}_{ij}$ is an estimate of the ratio $V(X_i)/V(X_j)$ and $\hat{R}$ is an estimate of $V(Y)/V(X_m)$. The first term in the product requires the methods of §2.3 whereas the remaining terms are each examples of the type of estimation problem discussed in this section, with the difference that at each stage the local reference phase, X, will be the phase of interest of the previous stage and therefore only the material sampled in the previous stage will be available for analysis. The obvious advantage of this type of design over a single stage procedure which ignores the intermediate phases is that at each stage we hope to maximise, in some sense, the proportion of phase of interest and consequently the amount of useful information. The disadvantage might be that we introduce multiplicative errors but at least with ratio estimators we can expect to reduce the variance at each stage; indeed this is a situation in which ratio estimators can be used to great effect.

Chapter 3 Relative Volume Estimation

3.1 Introduction

We have already made some remarks in the preceding chapter concerning the use of ratio estimators as a means of reducing variance when estimating absolute volume. These remarks are clearly very pertinent in the context of estimation of relative volume in a sampling unit from a population of Type 1 and indeed much of the discussion of the previous chapter can be seen to be applicable in this context. However, we will be concerned in this chapter with the estimation of relative volume (or volume fraction) in a sampling unit from a population of Type 2, where we can construct stochastic models for the geometric structure of the specimen and we will look in some detail at approaches based on lineal analysis.

3.2 Model Based Inference

The methods of the previous chapter were founded on the assumption that the specimen under examination was a finite, deterministic structure and the randomisation underpinning the discussion was that derived from the sampling scheme. Whereas such an approach is entirely appropriate in the context of a population of Type 1 it is much less so for a population of Type 2, where the sampling unit is somewhat arbitrary. An alternative is to regard the specimen as a realisation of a stochastic process about which we wish to make inference. The sampling scheme then becomes less central to the discussion and instead the probability structure is determined by the stochastic model that is assumed. The introduction of modelling assumptions always begs justification and results must be judged not only on the basis of their value given the model but also on the suitability of the model itself.

We are presented with at least two levels at which we can form models. First, and perhaps most obviously, is the geometric level, that is, the modelling of the physical structure of the specimen, and in this context we can call on random set theory. Secondly, there is the statistical modelling of the data available for analysis and

here there are some interesting problems arising from the nature of the study. For example, there are often unusual dependencies in the data resulting from the spatial relationship of the objects represented by the data. Clearly a geometric model together with a sampling scheme implies a model for the data but this may not be tractable and hence the possibility exists of a need for further modelling. Furthermore the geometric model may be only a partial description, perhaps assuming certain conditions such as convexity of particles, which is then complemented by distributional assumptions regarding the data. Certainly there is no obvious way to approach the issue of modelling and the various viewpoints are all inter-related. Indeed, fundamental to any theory of random sets is the notion of characterisation of the random set in terms of the probability of it intersecting a "test set". Thus it is quite natural to define a model in terms of the intercepts it makes with lower dimensional probes in a stereological study.

Where possible we will attempt to assess the scope and appropriateness of particular modelling assumptions and highlight their deficiencies; in section 3.4 we give a simple example where model based and design based approaches give different estimators. One area which has not received much attention in the stereological literature is goodness-of-fit tests although Ripley (1988, Ch.6) discusses the use of summary plots, based on the morphological transformations of erosion, dilation, opening and closure developed by Serra and Matheron (Serra, 1982), for assessing the suitablity of a model fitted to a given image.

3.3 General Framework

Our aim is to describe a random partition of the reference space (usually a subset of $\mathbb{R}^2$ or $\mathbb{R}^3$) into two phases. That is, if the reference phase is X and the phase of interest is Y, with $Y \subset X$, we want to assign each point $x \in X$ to either Y or $X \setminus Y$ by some probabilistic law, thus defining a binary-valued stochastic process, $\{Z(x) ; x \in X\}$. We can then base our inference on a sample taken from a realisation of the process. The sample will consist of the values that Z takes on the intersection of X with a test set (usually a set of points, lines or regions) and the natural estimator of

volume fraction will be a linear combination of the $\{Z(\mathbf{x})\}$; we can write this estimator formally as

$$\frac{1}{\mu(T)} \int_T Z \, d\mu$$

for a general test set T, where μ is Lebesgue measure in the case of T being a set of lines or regions and $\mu(A)$=number of points in A, for all $A \subset T$, when T is a set of points. (The existence of the integral relies on the assumption that all realisations of Z are measurable functions but this does not worry us in our context since it is unlikely that we should be interested in a model that could generate an unmeasurable set!) Of course we may wish to consider a weighted sum (integral) in some circumstances but here we restrict our attention to the usual unweighted version. Since the estimator is linear its variance will be expressible in terms of the quantities var$[Z(\mathbf{x})]$ and cov$[Z(\mathbf{x}), Z(\mathbf{y})]$ and thus a knowledge of the second order properties of the process $\{Z(\mathbf{x})\}$ will be sufficient to estimate the variance of the estimator of volume fraction. There are essentially three approaches that we can take, corresponding to increasingly strong assumptions about the process $\{Z(\mathbf{x})\}$: use the sample estimate of the covariance (the covariogram) directly, use a model for the covariance (fitted using the covariogram) or model the process $\{Z(\mathbf{x})\}$ and deduce the corresponding covariance structure from that model.

In general the covariance is represented by a function $c(\mathbf{x}, \mathbf{y}) = E[(Z(\mathbf{x}) - m(\mathbf{x}))(Z(\mathbf{y}) - m(\mathbf{y}))]$, where $m(\mathbf{x})=E[Z(\mathbf{x})]$, but usually some form of stationarity is assumed, most often that $m(\mathbf{x})$ is constant for all $\mathbf{x}$ and that $c(\mathbf{x}, \mathbf{y})=c(\mathbf{h})$ for all $\mathbf{x}$, $\mathbf{y}$, where $\mathbf{h}=\mathbf{x}-\mathbf{y}$. The reason for assuming stationarity is that since only one realisation of the process is normally available we have to use measurements made at different locations as if they were repeated realisations from the same distribution.

The ease of implementing the first approach mentioned above will depend on the type of sampling scheme employed. Certainly if the specimen is sampled by a regular grid of points then we will only require the value of $c(\,.\,)$ for a relatively small

number of values of **h** and there will be many pairs of points with **x**-**y** = **h** for each of those values, thus making estimation of the required values of $c(\,.\,)$ straightforward. On the other hand, if the sampling scheme is a set of irregularly spaced points then we will find that there are far more values of **h** for which we require the value of $c(\,.\,)$ and only a few pairs of points available for the estimation each time. In this situation it makes more sense to fit a model to the covariogram and much attention has been given to this topic in the geostatistical literature, including the design of optimum sampling schemes for the estimation of the covariogram (see, for example, Christakos, 1984, Cressie, 1985, Russo, 1984, Warrick & Myers, 1987. In fact the variogram, $\gamma(\mathbf{h})=\mathrm{var}[(Z(\mathbf{h})-Z(\mathbf{0})]$, is used rather than the covariogram because of the weaker conditions required for its existence but when both exist, as they will do in the context we are interested in, there is a simple linear relationship between them and so a knowledge of one implies a knowledge of the other.) An interesting feature of covariogram estimation is the paucity of models available as valid covariance functions, due to the requirement that a covariance function be positive definite. Interestingly, some of the covariance models commonly employed are derived from the random set models considered here in later sections; thus we could end up with the same form for the covariance function either by estimating it directly from the sample covariogram or by fitting a random set model and then deducing the corresponding covariance function, although it is not clear that we would necessarily obtain the same parameter values. One reason for fitting a model to the covariogram (or variogram) in geostatistical applications is that values of the covariance function are required for values of **h** that do not correspond to any pairs of points in the sample. In the same way, if our test set is a set of lines or regions then we will need to evaluate integrals involving the covariance function which will require knowledge of the its value over whole intervals in $\mathbb{R}^1$ or $\mathbb{R}^2$. Thus we would like to have a functional form for the covariance, although of course there exists the possibility of evaluating the integrals numerically, in which case a knowledge of the covariance for a finite set of values of **h** will suffice. Furthermore, since sample estimates of the covariance require us to sum over sets of point pairs, we

will need to use a separate test set to estimate the covariance from that used to estimate the volume fraction when T is a set of lines or regions. In §3.6 we will see how the covariance function can be estimated directly from the data obtained from a lineal analysis by assuming a specific Boolean model, thus avoiding the need for a second test set.

In the work that follows we will use random set models, such as the Boolean model, built up from geometric "building blocks" to model the process $\{Z(\mathbf{x})\}$ but here we mention briefly some models that might be applicable when the data are available on a regular grid, as will be the case with a digitised image represented by an array of pixels. In this case the data will be a finite collection of binary random variables and therefore it is theoretically possible to assign a discrete probability distribution to the set of all possible images. One model that does this is the Gibbs distribution (see, for example, Isham, 1981, Geman & Geman, 1984 and Ripley, 1988, Ch. 5), which assigns an *energy* to each configuration of 0's and 1's given by a *potential* function (the terminology comes from Physics, where these models were originally developed). It turns out that a particularly useful form of the Gibbs distribution is that which has a *nearest neighbour* potential. By neighbours we usually refer to sites which are adjoining vertically or horizontally; thus the sets of sites such that every member of a set is a neighbour of every other member of the same set (known as *cliques*) are simply pairs of adjoining horizontal or vertical sites. Of course it is possible to define any neighbourhood structure (see Besag, 1974 for a fuller discussion) and the particular type of neighbourhood structure employed in a model will reflect the nature of the spatial dependencies. The nearest neighbour potential, then, is one that is a sum of terms, each one of which is made up of contributions from sites which are all in the same clique. The importance of this structure is that it can be shown (see, for example, Isham, 1981, Geman & Geman, 1984, Besag, 1974) that the class of Gibbs distributions with nearest neighbour potentials is equivalent to the class of Markov random fields (MRF's) (essentially the Hammersley-Clifford theorem) and thus there is an equivalent formulation in terms of conditional probabilities, which

allows a description of the model in terms of local characteristics. This is particularly useful in image processing applications where images are restored using iterative methods which "update" pixel values by an algorithm depending on the values at neighbouring sites (see Geman & Geman, 1984 and Besag, 1986 for two examples from the extensive literature; see Ripley, 1988, Ch. 5 for a general overview of the methodology seen from a Bayesian point of view).

Although the Gibbs/MRF formulation described above is a purely spatial model it is easy to see how to extend the idea to incorporate a time dimension. Specifically, instead of a model of the form

$$E(Z_{i,j} \mid \{Z_{u,v} : (u, v) \neq (i, j)\}) = f_{i,j}(Z_{i-1,j}, Z_{i+1,j}, Z_{i,j-1}, Z_{i,j+1})$$

we would have something like

$$E(Z_{i,j,t} \mid \{Z_{u,v,s} : (u, v, s) \neq (i, j, t)\}) = f_{i,j,t}(Z_{i,j,t-1}, Z_{i-1,j,t-1}, Z_{i+1,j,t-1}, Z_{i,j-1,t-1}, Z_{i,j+1,t-1})$$

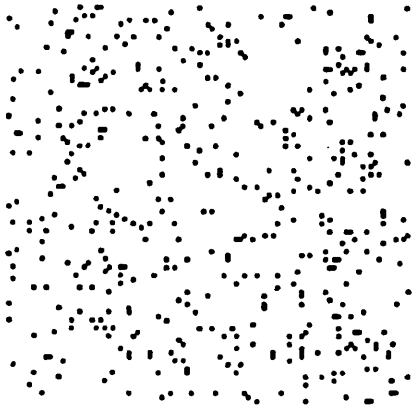
where $f_{i,j,t}$ is some suitable function which could be defined so as to allow the existence of an equilibrium distribution at $t = \infty$, perhaps depending on the initial state of the process. The initial state could be defined for example by

$$Z_{i,j,t_0} = \begin{cases} 0 & \text{with probability } 1-p \\ 1 & \text{with probability } p \end{cases}$$

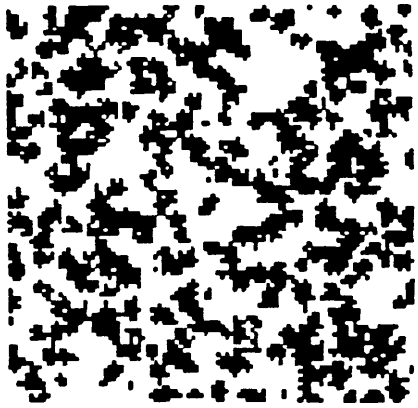
with $\{Z_{i,j,t_0}\}$ being mutually independent. A simulation of such a model is shown in fig. 3.1 for

$$f_{i,j,t} = Z_{i,j,t-1}(1 - \exp(-\lambda t)) + (Z_{i-1,j,t-1}+Z_{i+1,j,t-1}+Z_{i,j-1,t-1}+Z_{i,j+1,t-1})\exp(-\lambda t)$$

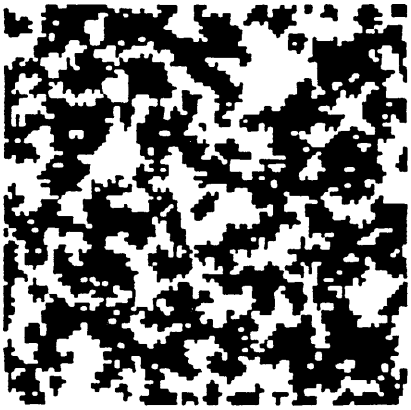
with $\lambda = 0.4$ and $p = 0.05$. (The edge effects have been dealt with by retaining a border around the edge which is not shown in the picture).



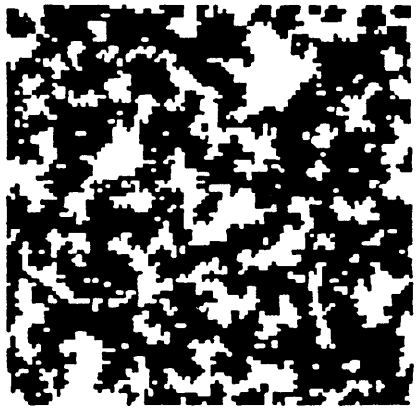
(a) $t = 0$



(b) $t = 4$



(c) $t = 7$



(d) $t = 16$

Fig. 3.1 Simulation of Time Ordered Spatial Process

The figure shows the evolution through time of a spatial process on a 100×100 lattice in which the probabilities at each site are dependent on the values at that site and the four adjacent sites at the previous time point. The last picture shows the equilibrium state for the realisation, whose existence is guaranteed by the choice of dependence function. The initial state is given by generating independent Bernoulli random variables at each site and the whole model is specified by two parameters (see text).

For a discussion of spatial-temporal models of this type and their equilibrium distributions see Isham (1981). Besag (1974) points out that with observations at a single time point analysis is not usually possible for this type of spatial-temporal model without very restrictive assumptions of independence and stationarity. However, it might be fruitful simply to look at the class of equilibrium models as a class of spatial models in its own right. Furthermore, we may in fact be able to have several observations on a process at different time points. In the study of the microstructure of concrete and cement there is interest in the evolution of the various constituent phases over time as the material dries out and hardens. Although each observation in time is from a different spatial location it would still be possible to design an experiment to analyse the full spatial-temporal model, particularly if the spatial component is stationary.

Despite the appealing nature of the models discussed above and their success in image processing applications it is not clear how we should use them in the context of volume fraction estimation. Clearly the neighbourhood structure and the form of the potential function describe the spatial dependence of the model but it is not immediately obvious how this relates to the covariance function that we require in order to estimate the variance of the volume fraction estimator. We turn now to random set models, and in particular the Boolean model, as a means of describing the process $\{Z(\mathbf{x})\}$ in geometric terms.

3.4 Boolean Models

In order to construct stochastic models for geometric structures we need to develop a theory of random sets. Two very general theories, developed in the last twenty years, are those of Kendall (1974) and Matheron (1975), both founded on earlier work of Choquet . Kendall's theory, drawing inspiration from the work of Davidson (reproduced in Kendall, 1974) is based on the idea of observing only whether or not a random set is intersected by a "test set" rather than observing the size of the intersection. This enables the treatment of a very large class of problems, indeed

much larger than we require in the present context. Matheron's approach is linked closely to the French school of mathematical morphology and image analysis and is inspired by practical experience. His treatment is perhaps more easily recognisable to stereologists than that of Kendall, with reference to familiar topics such as size distribution, random intercepts, Minkowski functionals, Steiner's formula, Crofton's formula, covariance measures and so on. We will not be concerned with such an abstract approach as those of Kendall and Matheron but rather with concrete examples of random sets relating to specific stereological situations. Principally we shall be interested in Boolean models and possible extensions and generalisations of them.

In the previous section we have suggested a form of growth model developing in time from a set of isolated sites scattered throughout the specimen; the Boolean model is in many ways an analogous model in a continuous reference space, viewed at $t = \infty$. Because points occupy zero volume in the continuous spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ we have to move from countable collections of points to countable unions of compact sets. The locations, orientations, shapes and sizes of the compact sets can be regarded as random and given a probability distribution; the part of the specimen observed in the sampling frame is then viewed as a realisation of the random set, from which the parameters of the model can be estimated.

One justification for this approach is the following. Because we are concerned with a sampling unit from a population of Type 2 we can regard it as being arbitrary and the population as being very large relative to the sampling unit. Therefore, provided that the sampling unit is chosen independently of its contents the observations contained in it form a simple random sample from the population distribution of the corresponding quantities. Replacing the population distribution by a continuous distribution is justified on the grounds of the large size of the population. Indeed in some contexts we may be interested in a genuine "process", for example an industrial process, rather than in any existing, physical entity and then the justification of such an approach is almost automatic.

The problems that arise are in describing a sufficiently realistic model which still allows the development of theoretical results and in modelling highly irregular geometric structures. In these respects the Boolean model is appealing, since it allows the derivation of some useful and important characteristics whilst being quite flexible. The fundamental building blocks of the Boolean model are compact sets, such as discs, lines, spheres, ellipsoids, etc. whose locations are determined by a point process. Although the formulation could be very general, the Boolean model represents a somewhat restricted class (with certain independence conditions imposed) which allows the derivation of some useful results; the Boolean model with convex grains, in particular, gives some elegant results.

Formal definitions of the Boolean model can be found in many places, for example Stoyan, Kendall & Mecke (1987); loosely, it is constructed from a stationary Poisson point process, giving a set of "germs" at each of which is located a "primary grain", being a realisation of a compact random set, which is independent of the location of the germ. When we say that a primary grain "is located at" a germ we mean that the realisation of the random set is translated by the vector $\mathbf{x}$, where $\mathbf{x}$ is the position of the germ. The compact random set forming the primary grain is usually taken to be of a relatively simple form, although we shall see later that assuming only that it is convex is sufficient to obtain useful results. We look briefly at two very simple examples of Boolean models to illustrate the basic ideas.

Example 1 Let $f_R(r)$ be a probability density function corresponding to a random variable R which takes only positive values and let $Y = \cup B(\mathbf{x}_i, R_i)$ where $B(\mathbf{x}, r)$ is a sphere of radius r and centre $\mathbf{x}$, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ is the realisation of a stationary Poisson point process in $\mathbb{R}^3$ and $R_1, R_2, \ldots$ are independent and identically distributed observations from the density $f_R(\,.\,)$. Then Y is a Boolean model with spherical primary grains. We note that Y is stationary (by the stationarity of the point process and the independence of germs and grains) and in this case the model is isotropic, although this second property does not always hold for Boolean models; it will only hold in general when the primary grains have a uniformly random

orientation. This model could be represented as a marked point process with the radii regarded as marks attatched to the points but it is not clear that anything valuable can be gained from this approach. Generally, although such a treatment provides a very elegant theory, it is only possible to derive second order properties when factorisations of the second moment measure exist, these corresponding to the independence of the Boolean model, for which results can be derived more straightforwardly (although perhaps less rigorously). It is tempting to regard the grains of the model as physical particles, but we should be aware that there is nothing in the model to preclude "overlapping" and the probability of overlapping becomes quite large quite quickly as the intensity of the point process increases.

Example 2 Suppose that $f_{\Theta, L}(\theta, l)$ is a joint density on $[0, \pi) \times (0, \infty)$ and that (θ_1, l_1), (θ_2, l_2), ... are independent and identically distributed observations from $f_{\Theta, L}(\,.\,,\,.\,)$. Then if we construct a process of line segments (in $\mathbb{R}^2$) such that the ith line has length l_i and orientation θ_i, and where the midpoints of the lines are the realisation of a stationary Poisson point process with intensity λ, this process will be a Boolean model. An interesting estimation problem associated with this model is to estimate the mean length of lines per unit area of the plane from information collected on a linear probe. This problem is in fact a two dimensional analogy of a problem in minerology in which the areal density per unit volume of planar "cracks" in a geological region is to be estimated from measurements obtained from a cylindrical core (thin enough to be regarded as a line probe). Returning to the two dimensional problem, it is assumed that the angle of intersection between the line probe and a line of the process hit by it can be measured. First we note that the intensity of the point process on the line probe formed by intersections with lines whose length is in the interval $[l, l+\mathrm{d}l)$ and whose orientation is in the interval $[\theta, \theta+\mathrm{d}\theta)$ is given by

$$\lambda l \sin\theta f_{\Theta, L}(\theta, l)\, \mathrm{d}\theta\, \mathrm{d}l$$

(see fig. 3.2).

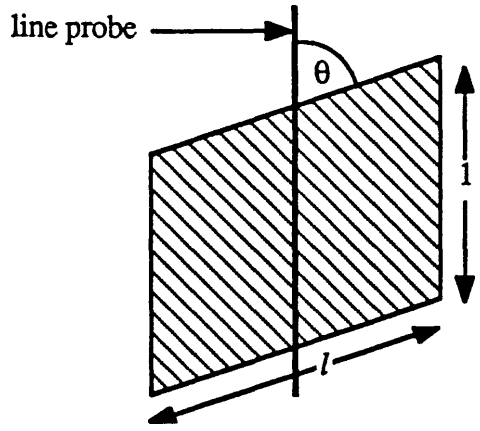


Fig. 3.2 Intersection of Line Process with Linear Test Probe

Lines with orientation θ and length l whose midpoints fall in the shaded area will intersect the unit length of line probe within the shaded area.

Therefore the intensity of the point process on the line probe formed by intersections with all lines of the process is

$$\lambda \iint l \sin\theta f_{\Theta,L}(\theta,l)\mathrm{d}\theta\mathrm{d}l \ ,$$

which we will denote λ^*. Also, the joint density of Θ and L for lines intersecting the line probe must be proportional to $l \sin\theta f_{\Theta,L}(\theta, l)$ and hence is given by

$$g_{\Theta,L}(\theta, l) = \frac{l\sin\theta f_{\Theta,L}(\theta, l)}{\iint l\sin\theta f_{\Theta,L}(\theta, l)\mathrm{d}\theta\mathrm{d}l} \ .$$

It follows from these formulae that

$$E\left[\frac{1}{\sin\Theta}\right] = E(L) \Big/ \iint l\sin\theta f_{\Theta,L}(\theta,l)\mathrm{d}\theta\mathrm{d}l$$

where the left hand side expectation is over lines intersecting the line probe and $E(L)$ is the mean length of the lines of the original process. Hence for a line probe of length T intersected by N lines with orientations $\theta_1, \ldots, \theta_N$ we have that

$$\frac{1}{T}\sum_{i=1}^{N}\frac{1}{\sin\theta_i}$$

is an unbiased estimator of $\lambda E(L)$ since $E(N) = T\lambda^*$. Analagous arguments can be applied to the three dimensional version of the problem. However, the classical design based approach to this problem yields a different estimator. Briefly the formulation of the problem would be that a fixed number, n, of lines are contained in a bounded region and that the ith line has midpoint $\mathbf{x}_i$, length l_i and orientation θ_i. An *IUR* line probe of length T is generated in such a way that all positions for which it hits at least one of the lines are possible. Then for any given orientation, ψ, of the probe the probability of it hitting the ith line is proportional to $l_i|\sin(\psi-\theta_i)|$ but because the probe is isotropic the expected probability of it hitting the ith line is simply proportional to l_i. Hence the usual estimator in this framework would simply be proportional to the number of "hits" and would therefore not be robust in the presence of anisotropy. By contrast the estimator derived in the model based framework takes account of anisotropy automatically. Of course this estimator is still perfectly valid in the design based framework but the irony is that the design is of no value in this situation and, indeed, it may lead us to consider sub-optimal estimators.

There is clearly scope for quite complex models within the class of Boolean models. The only restrictions are:

1) The germs form a stationary Poisson point process.
2) The primary grains are independent and identically distributed realisations of a compact random set.

In practice the most convenient way to describe a compact random set is to define a particular geometric structure (such as a circle or ellipse) which can be specified by a set of parameters and then to specify a joint distribution for those parameters. This, of course, is not as general as the definitions to be found in the work of Kendall, Matheron and others but it is more intuitively accessible and quite flexible. For example, it is perfectly admissible to define a mixture of several distributions, each one associated with a different geometric shape; this is still a Boolean model. Even with circles quite irregular specimens can be modelled, although sparsely distributed, irregular structures cannot be so readily represented since the irregularity stems from

overlapping circles. This situation might well be modelled by having a cluster of circles at each germ but we have to take care not to violate the conditions associated with the Boolean model. Indeed, extensions to the Boolean model may often arise naturally, but usually the price we pay is in the loss of analytic results. One alternative approach is to examine not the Boolean model itself but the intercepts it makes with a stereological test probe. We do this next but first we introduce the covariance function for random set models.

3.5 The Covariance

We define first some notation. Let Y_0 be a compact random set in the (unbounded) reference space X (usually X will be $\mathbb{R}^3$) and let $Y_1, Y_2, \ldots$ be independent and identically distributed (i.i.d.) replicates of Y_0. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be a stationary Poisson point process in X and denote by

$$Y = \bigcup_{i=1}^{\infty} (Y_i + \mathbf{x}_i)$$

the Boolean model with primary grains $Y_1, Y_2, \ldots$ and germs $\mathbf{x}_1, \mathbf{x}_2, \ldots$.

The covariance function is defined as

$$C(\mathbf{r}) = \Pr(\mathbf{x} \in Y \cap \mathbf{x}+\mathbf{r} \in Y) \qquad \mathbf{x}, \mathbf{x}+\mathbf{r} \in X$$

which is equal to $\Pr(\mathbf{0} \in Y \cap \mathbf{r} \in Y)$ (where $\mathbf{0}$ is an arbitrary origin in X) by the stationarity of the Boolean model. We will also use the notation $K(\mathbf{r})$ for the conditional form of the covariance, namely $\Pr(\mathbf{r} \in Y \mid \mathbf{0} \in Y) = C(\mathbf{r})/\pi_0$, where $\pi_0 = \Pr(\mathbf{0} \in Y)$ is the volume fraction occupied by Y. Notice that this covariance function is simply an uncentred version of the standard covariance function $c(\mathbf{h})$ discussed in §3.3, since for a binary-valued process $\{Z(\mathbf{x})\}$ we have that $E[Z(\mathbf{x})Z(\mathbf{y})] = \Pr(Z(\mathbf{x})=1 \cap Z(\mathbf{y})=1)$ and hence $C(\mathbf{h}) = c(\mathbf{h}) + \pi_0^2$.

For a Boolean model we have

$$C(\mathbf{r}) = 2\pi_0 - 1 + (1-\pi_0)^2 \exp\{\lambda D_{Y_0}(\mathbf{r})\}$$

where λ is the intensity of the Poisson point process,

$$D_{Y_0}(\mathbf{r}) = E\,[\mu(Y_0 \cap (Y_0 - \mathbf{r}))]$$

and μ is Lebesgue measure on X.

For a compact set, A_0, $D_{A_0}(\mathbf{r})$ is the "global covariance" of Serra (1982), which is equal to the erosion of A_0 by the two point set $\{\mathbf{0}, \mathbf{r}\}$. This quantity has also been termed the "geometric reduction factor" (Kellerer, 1986) and is closely related to the "point-pair-distance" distribution, that is the distribution of the distance between pairs of points taken independently and uniformly from A_0, and hence it occurs in the study of area and volume estimators based on point counting methods (see, for example, Matérn, 1985).

Both Kellerer and Serra use the same terminology whether A_0 is random or deterministic, simply adding the expectation in the definition when it is random; the expectation is of course with respect to the randomness of A_0. When A_0 is an isotropic random set $D_{A_0}(r)$ (which is a function of distance only) is referred to by physicists as the "distance probability function" (Stoyan, Kendall & Mecke, 1987) and has applications in X-ray analysis. (Alternatively this could refer to the mean function, $\bar{D}_{A_0}(r)$, averaged over all orientations, for an anisotropic random set.) The function D has some interesting properties, such as that the mean (over all orientations) of the derivative at the origin is proportional to the surface area (or boundary in $\mathbb{R}^2$) of the compact set to which it is applied. That is (Serra, 1982)

$$-\frac{1}{4\pi}\int D_{A_0}'((0, \omega))\, d\omega = \frac{E(S(A_0))}{4} \qquad \text{in } \mathbb{R}^3$$

$$-\frac{1}{2\pi}\int D_{A_0}'((0, \alpha))\, d\alpha = \frac{E(B(A_0))}{\pi} \qquad \text{in } \mathbb{R}^2.$$

Here expectations on the right hand side are with respect to the randomness of A_0, $S(A_0)$ is the surface area of A_0 and $B(A_0)$ is the boundary of A_0. $D_{A_0}'((0, \alpha))$ is the limit as $h \to 0$ of $[D_{A_0}((h, \alpha)) - D_{A_0}((0, \alpha))]/|h|$ where (h, α) is a vector of length h and orientation α (in $\mathbb{R}^3$ the definitions are equivalent but with α replaced by the solid angle ω) and integration is over the surface of a unit ball in the appropriate number of dimensions. (We note that $(0, \alpha)$ is a vector of zero length but nevertheless $D_{A_0}((0, \alpha))$ is unambiguous, given the definition of D, and is simply the Lebesgue measure of A_0.)

Clearly D is a fundamental quantity in many areas of geometric probability and related fields. The relationship between this global covariance and the "local covariance", which is given by $C(\mathbf{r})$, is an interesting one. We can see it better, perhaps, if we think not of $D_{A_0}(\mathbf{r})$ but $D_{A_0}(\mathbf{r})/\mu(A_0)$, which is precisely the conditional probability that $\mathbf{x}+\mathbf{r} \in A_0$ given that $\mathbf{x} \in A_0$. We then note that for some $B_0 \supset A_0$ we have

$$D_{A_0}(\mathbf{r}) = \frac{D_{A_0}(\mathbf{r})}{\mu(A_0)} \frac{\mu(A_0)}{\mu(B_0)} \mu(B_0)$$

$$= \Pr(\mathbf{x}+\mathbf{r} \in A_0 \mid \mathbf{x} \in A_0) \Pr(\mathbf{x} \in A_0 \mid \mathbf{x} \in B_0) \mu(B_0)$$

$$= \Pr(\mathbf{x} \in A_0 \cap \mathbf{x}+\mathbf{r} \in A_0 \mid \mathbf{x} \in B_0) \mu(B_0)$$

and this now closely resembles the local covariance, $C(\mathbf{r}) = \Pr(\mathbf{x} \in Y \cap \mathbf{x}+\mathbf{r} \in Y)$. The difference, of course, is that Y is a stationary random set defined on the whole space ($\mathbb{R}^2$ or $\mathbb{R}^3$ usually) such that $E[\mu(Y \cap B_0)/\mu(B_0)]$ is constant for all B_0 whereas A_0 is a compact set (and therefore bounded) for which the probability $\Pr(\mathbf{x} \in A_0 \mid \mathbf{x} \in B_0)$ (and hence $\Pr(\mathbf{x} \in A_0 \cap \mathbf{x}+\mathbf{r} \in A_0 \mid \mathbf{x} \in B_0)$) tends to zero as $\mu(B_0) \to \infty$.

In the Boolean model the local covariance $C(\mathbf{r})$ depends on the global covariance of the primary grains, which are compact sets, with the stationarity

introduced through the Poisson point process of germ centres. In practice it will usually not be possible to obtain an exact expression for $C(\mathbf{r})$ because $D_{Y_0}(\mathbf{r})$ is only tractable for a small number of special cases (see Kellerer, 1986). In the case where the grains are spheres in $\mathbb{R}^3$ with random radii, having density $f_R(\,.\,)$, we note that because of the isotropy of the primary grains both the local covariance of the Boolean model and the global covariance of the grains are functions of distance only. Writing $B(\mathbf{0}, r)$ for a sphere with centre $\mathbf{0}$ and radius r we note that

$$D_{B(0,\,r)}(y) = \begin{cases} 0 & y \geq 2r \\ \frac{4\pi}{3}\left\{r^3 - \frac{3}{4}yr^2 + \frac{1}{16}y^3\right\} & y < 2r \end{cases}$$

(see fig. 3.3).

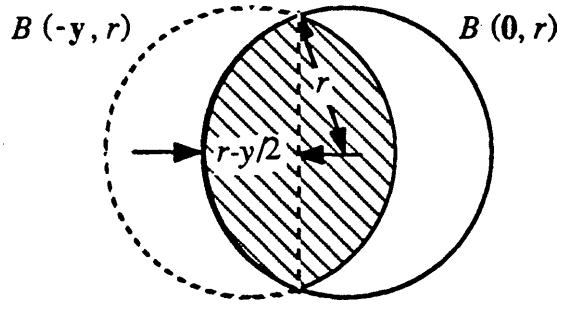


Fig. 3.3 Global Covariance for a Sphere.

Our primary grain Y_0 is a sphere whose radius has density $f_R(\,.\,)$ and therefore

$$D_{Y_0}(y) = E_R[D_{B(0,\,R)}(y)]$$

$$= \frac{4}{3}\pi \int_{\frac{y}{2}}^{\infty} \left\{t^3 - \frac{3yt^2}{4} + \frac{y^3}{16}\right\} f_R(t)\,dt.$$

3.6 Lineal Analysis

Stereological analyses in practice generally require the measurement of various quantities on planar sections. There are essentially three choices in the context of volume estimation; we can measure areas of intersection with planar sampling frames, lengths of intersection with linear probes or numbers of included points from a test set. Whereas measurements of area will usually be the most efficient means of estimating volume they are probably the least easy measurements to obtain; conversely point counts are quick and easy to obtain but are likely to be the least efficient. Lineal analysis, in which intersection of a line probe with the phase of interest is the measurement on which estimates are based, falls inbetween the extremes with respect to efficiency and ease of measurement. However, it has one distinct advantage, and that is that we can obtain a simplified form for the covariance function, albeit with an extra assumption about the model.

Let us first define the situation more precisely and introduce some notation. As before we use X to denote the reference phase and $Y \subset X$ to denote the phase of interest, assumed to be a Boolean model with convex grains. The sampling unit will be a block of material sampled from the specimen, and within this block we will take planar sections on which we will perform the analysis. We need not be concerned greatly with the sampling scheme since it plays no rôle in a model based approach such as we are considering. However, in cases where strong anisotropy is suspected we may wish to be more specific about the sampling scheme. On each planar section we place a set of n equally spaced, parallel linear test probes of length l and separation d, which we will denote $T_1, \ldots, T_n$. On each linear probe there will be an alternating sequence of intercepts with Y and $X \setminus Y$. The lengths of the intercepts made by T_i with Y will be denoted by $\{y_{ij} : j = 1, \ldots, m_i\}$ and similarly the intercepts made with $X \setminus Y$ will be $\{x_{ij} : j = 1, \ldots, n_i\}$ as shown in fig. 3.4. To start with we will consider just a single planar section but later we shall put the results into a wider context.

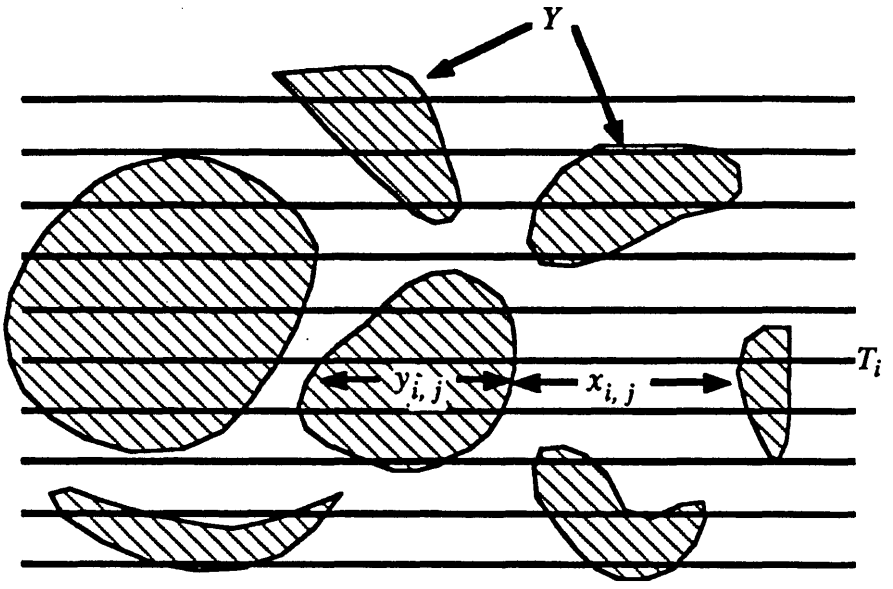


Fig. 3.4 Lineal Analysis of a Boolean Model.

Note that in general the Boolean model does not necessarily have convex grains. However the development in the text is for a model with convex grains.

Our data, then, consists of the $\{x_{ij}\}$ and $\{y_{ij}\}$ together with the random quantities $\{n_i\}$ and $\{m_i\}$. The structure of the data is quite complex with many dependencies but the fundamental result on which we build is that the $\{x_{ij}\}$ have an exponential distribution. This result can be found in the literature (Serra, 1982, Stoyan, Kendall & Mecke, 1987) but here we provide an alternative derivation.

We suppose that the primary grain, Y_0, of the Boolean model is convex and can be parameterised by the vector of parameters θ which has distribution function $F_\Theta(\theta)$ in the parameter space $\mathcal{B}_\Theta$; that is, the germs of the grains with parameters in the subset S of $\mathcal{B}_\Theta$ are a stationary Poisson point process with intensity $\lambda\mu_F(S)$, where λ is the intensity of the process of germs for the whole model and μ_F is the probability measure induced on $\mathcal{B}_\Theta$ by F. Now we restrict attention to an arbitrary line in X containing an arbitrary origin and denote by $\mathbf{s}=s\mathbf{u}$ (where $\mathbf{u}$ is a unit vector parallel to the line) the point on the line at a distance s from the origin. We then define

$$Z_s = \begin{cases} 0 & s \in X \setminus Y \\ 1 & s \in Y \end{cases}$$

and

$$\pi_{01} = \Pr\,(Z_{s+ds} = 1 \mid Z_s = 0)$$
$$\pi_{10} = \Pr\,(Z_{s+ds} = 0 \mid Z_s = 1)$$

and let $Y_0(\Theta_1)$, $Y_0(\Theta_2)$, ... represent realisations of the compact random set Y_0. Representing the primary grain $Y_0(\theta)$ with germ **x** more concisely by the pair (x, θ) we note first that $s \in (x, \theta)$ if and only if $x \in (s, \theta^-) = {}^-Y_0(\theta)+s = \{s-r: r \in Y_0(\theta)\}$. Then it is clear (see fig. 3.5) that there will be a transition from $X \setminus Y$ to Y at s if and only if there exists a θ_0 for which at least one germ of the process with primary grain $Y_0(\theta_0)$ falls inside $(s + ds, \theta_0^-)$ given that for every θ no germ with primary grain $Y_0(\theta)$ falls inside (s, θ^-).

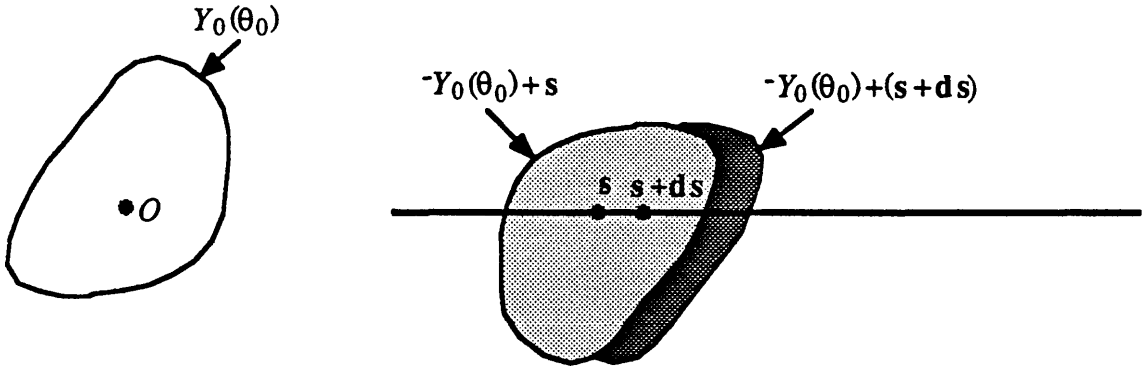


Fig. 3.5 Probability of Transition from $X \setminus Y$ to Y.

The probability that s+ds is in Y given that s is in $X \setminus Y$ is equal to the probability that for some θ_0 at least one germ of the Boolean model with primary grain $Y_0(\theta_0)$ lies in the darker region. For small ds this region has an area approximately equal to ds times the projection of $Y_0(\theta_0)$ onto a plane whose normal is the line probe. Thus in the limit the transition probability is proportional to ds and hence the intervals of intersection with $X \setminus Y$ have exponential length distribution.

Therefore, if Θ is a discrete random variable taking the values $\theta_1, \theta_2, \ldots$ we have

$$\pi_{01} = 1 - \prod \Pr(\text{No germs with primary grains } Y_0(\theta_i) \text{ fall in } (\mathbf{s} + \mathbf{ds}, \theta_i^-) \cap (\mathbf{s}, \theta_i^-)^c)$$
$$= 1 - \exp\left(-\lambda \sum V\{(\mathbf{s} + \mathbf{ds}, \theta_i^-) \cap (\mathbf{s}, \theta_i^-)^c\} \Pr(\Theta = \theta_i)\right).$$

Rewriting the sum as a Lebesgue-Stieltjes integral with respect to the measure μ_F introduced above we see that the result is easily generalised to the case when Θ is continuous by expressing the distribution function, F, as the limit of a monotone increasing sequence of step functions, which correspond to discrete distributions. (This will always be possible by the continuity of F.)

Now for small ds we have that $V\{(\mathbf{s} + \mathbf{ds}, \theta^-) \cap (\mathbf{s}, \theta^-)^c\}$ is approximately equal to ds times the projection of $(\mathbf{s}, \theta^-)$ onto a plane to which $\mathbf{u}$ is the normal; therefore provided that $\mathbf{u}$ has a uniformly random orientation[1] relative to $(\mathbf{s}, \theta^-)$ the expected value of $V\{(\mathbf{s} + \mathbf{ds}, \theta^-) \cap (\mathbf{s}, \theta^-)^c\}$ is equal to

$$S\{(\mathbf{s}, \theta^-)\}ds/4 + O((ds)^2)$$

by Cauchy's Formula (where $S(\,.\,)$ denotes surface area). Therefore, expanding the exponential and reversing the order of expectation with respect to the orientaion of the line probe and integration over the distribution of Θ we have

$$\pi_{01} = \lambda\left[\int S\{(\mathbf{s}, \theta^-)\} dF(\theta)\right] ds/4 + O((ds)^2)$$
$$= \lambda\left[\int S(Y_0(\theta)) dF(\theta)\right] ds/4 + O((ds)^2).$$

It follows that the lengths of the intervals of intersection of the line probe with $X \setminus Y$ have an exponential distribution with parameter equal to a quarter of the mean grain surface area per unit volume (not discounting the hidden surfaces of overlapping particles). The critical requirement is the convexity of the grains; if $\mathbf{s} + \mathbf{ds} \in (\mathbf{x}, \theta)$

[1] For a fixed direction the distribution is still exponential but if the grains are not isotropic the parameter of the exponential distribution will vary with the direction of the line probe.

and $\mathbf{s} \notin (\mathbf{x}, \theta)$ then $\mathbf{s} - \varepsilon\mathbf{u} \notin (\mathbf{x}, \theta)$ for all $\varepsilon > 0$ and hence the "history" of the process Z prior to the point $\mathbf{s}$ has no effect on π_{01}. When considering π_{10} however we can no longer use the same technique because the rôles of the grains and the voids have effectively reversed and the voids are not convex. The distribution of the lengths of intersection of the line probe with Y is in fact equivalent to the distribution of the length of the busy period in a queue with random arrivals and an infinite number of servers (since overlapping is possible), where the service time distribution is given by the chord length distribution of the primary grain. However, this observation is not generally helpful since the chord length distribution will not be known unless we make specific shape assumptions about the primary grains, and even then it is often not tractable.

At first sight it might appear unimportant that we do not know the distribution of the lengths of the intercepts with Y since we could base our inference on the data which we know to be exponentially distributed; the sum of the observations is the sufficient statistic for the exponential distribution and estimators of volume fraction can be expressed in terms of the total length of the linear probes (which is fixed) and the exponentially distributed variables. However, the situation is considerably more complex since the unknown length distribution enters the data in that the *numbers* of intercepts (that is the $\{n_i\}$ and $\{m_i\}$) are random and their distribution depends on the distributions of the lengths of both sets of intercepts. Further complications that arise in the data are the various dependencies; for example we know that $|n_i - m_i| \leq 1$ for all i and also that the numbers of intercepts (and the lengths of intercepts) on different probes are correlated, because if a grain is "hit" by a probe then we can say something about the probabilty of it being hit by a parallel probe and about the distribution of the length of its intersection given the length of intersection with the first one. Indeed even the exponentially distributed intervals on different probes are correlated. We will return to these issues later but first we derive expressions for the covariance function of the Boolean model expressed in terms of intercept lengths and for the variance of an estimator of relative volume.

3.6.1 Covariance Function for the Exponential Model

Stoyan (1979) defines an exponential model to be a two phase spatial process in which the intercepts of a linear probe with one of the phases have an exponential length distribution. With the extra assumption that the grains are uniformly randomly orientated (that is, that the model is isotropic) the Boolean model with convex grains is an example of an exponential model; the intercepts of a line probe with the reference phase, $X \setminus Y$, have an exponential distribution (whose mean we shall denote by $1/\mu$) and we will denote by $F_Y(y)$ the distribution function of the intercept lengths with the phase of interest, Y (with corresponding density $f_Y(y)$ and mean $1/\tau$).

On any line probe we have alternating intervals of the two different phases and thus we can regard the process induced on the line probe as an alternating renewal process. Since for any two points we can always find a line probe that contains them we can deduce the covariance function from standard results in renewal theory. The fundamental result (see Cox, 1962) is that for an equilibrium alternating renewal process the probability of being in state 1 at time t given that the process is in state 1 at time 0 is given by

$$\frac{\mu_2}{\mu_1 + \mu_2} + \mu_1 \omega(t)$$

where

$$\omega^*(s) = \frac{1}{(\mu_1 + \mu_2)s} - \frac{\{1 - f_1^*(s)\}\{1 - f_2^*(s)\}}{\{1 - f_1^*(s) f_2^*(s)\}s^2},$$

$f_1(\,.\,)$ and $f_2(\,.\,)$ are the failure time densities for states 1 and 2 respectively, with means $1/\mu_1$ and $1/\mu_2$, and the notation $g^*(s)$ is used for the Laplace transform of the function $g(t)$. In our context $\mu_1 = \tau$, $\mu_2 = \mu$, $f_1(\,.\,) = f(\,.\,)$ and $f_2(\,.\,)$ is exponential with parameter μ, so that $f_2^*(s) = \mu/(\mu + s)$. Thus the $K(\,.\,)$ of § 3.5 can be expressed in terms of its Laplace transform, given by

$$K^*(s) = \frac{\mu}{(\mu+\tau)s} + \frac{\tau}{(\mu+\tau)s} - \tau\frac{\{1-f_1^*(s)\}\{1-f_2^*(s)\}}{\{1-f_1^*(s)f_2^*(s)\}s^2}$$

$$= \frac{1}{s} - \frac{\tau\{1-f^*(s)\}}{s\{s+\mu-\mu f^*(s)\}}.$$

If $f(\,.\,)$ is also exponential then

$$K^*(s) = \frac{1}{s} - \frac{\tau}{s(s+\mu+\tau)}$$

giving

$$K(t) = 1 - \frac{\tau}{\mu+\tau}\left(1-e^{-(\mu+\tau)t}\right).$$

Multiplying by π_0 and noting that $\pi_0 = \mu/(\mu+\tau)$ we see that

$$C(t) = \pi_0^2(1-e^{-(\mu+\tau)t}) + \pi_0 e^{-(\mu+\tau)t}$$

as stated in Stoyan (1979). Of course for the Boolean model $f(\,.\,)$ is not exponential and we return in §3.6.3 to the problem of finding $C(\,.\,)$ in this case.

3.6.2 Variance of Lineal Fraction

The aim of our lineal analysis is to estimate the volume fraction, π_0, of a specimen, X, occupied by the phase of interest, Y. The natural estimator is the lineal fraction, that is

$$\hat{\pi}_0 = \frac{1}{nl}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij},$$

and in this section we derive the formula for $\mathrm{var}(\hat{\pi}_0)$ in terms of the covariance function of the model. Rewriting $\hat{\pi}_0$ as $\bar{L}$, where

$$L_i = \frac{1}{l} \sum_{j=1}^{m_i} y_{ij}$$

we have

$$\mathrm{var}(\hat{\pi}_0) = \frac{1}{n} \mathrm{var}(L_i) + \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{cov}(L_i, L_j).$$

Because of the stationarity and isotropy of the model i and j can only enter the expression for $\mathrm{cov}(L_i, L_j)$ in terms of $|i-j|$. Therefore we can write the variance of $\hat{\pi}_0$ as

$$\frac{\sigma^2}{n} \left\{ 1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) c_k \right\}$$

where $\sigma^2 = \mathrm{var}(L_i)$ and $\sigma^2 c_k = \mathrm{cov}(L_i, L_j)$ for $|i-j| = k$. It remains to find expressions for σ^2 and c_k.

Without loss of generality we choose a Cartesian frame of reference such that the line T_i has end points $(0, id)$ and (l, id). Then we can write

$$L_i = \frac{1}{l} \int_0^l Z(x, id)\, dx$$

where $Z(x, y) = 1$ if the point (x, y) is in Y and zero otherwise. Hence we have

$$E[L_i L_j] = \frac{1}{l^2} E\left[\int_0^l Z(u, id)\, du \int_0^l Z(v, jd)\, dv \right]$$

$$= \frac{1}{l^2} E\left[\int_0^l \int_0^l Z(u, id) Z(v, jd)\, du dv \right]$$

$$= \frac{1}{l^2} \int_0^l \int_0^l E[Z(u, id) Z(v, jd)]\, du dv$$

$$= \frac{1}{l^2} \int_0^l \int_0^l C\left([(u - v)^2 + (kd)^2]^{1/2} \right) du dv$$

where $k = |i - j|$. A straightforward transformation of variables allows the reduction to a single integral and subtracting $E(L_i)E(L_j)$ gives

$$\sigma^2 c_k = \frac{2}{l^2}\int_0^l (l - x)C\left(\{x^2 + (kd)^2\}^{1/2}\right)dx - \pi_0^2 \qquad k = 1, \ldots, n-1$$

$$\sigma^2 = \frac{2}{l^2}\int_0^l (l - x)C(x)dx - \pi_0^2 .$$

3.6.3 Modelling Intercept Lengths

We now have an expression for the variance of the volume fraction estimator, $\hat{\pi}_0$, in terms of the covariance function of the model, $C(\,.\,)$; we also have an expression for the covariance function but which requires the length distribution of the intercepts of a line probe with Y and subsequent inversion of a Laplace transform. We have mentioned already that the required length distribution is equivalent to the busy period of an M/G/∞ queue whose service time distribution is the distribution of the lengths of random (*IUR*) chords of the primary grain of the Boolean model, but this would appear not to be very helpful since there are few results in the literature pertaining to this situation (not surprisingly, since most queueing problems are concerned with a finite service capacity) and those which exist certainly do not look like being useful in our context (see, for example, Ramalhoto, 1984, Shanbhag, 1966). Indeed, even for the simpler M/G/1 queue, which has received far more attention (see Cox & Smith, 1961, Ch. 5), the required distribution is given by the functional equation

$$f^*(s) = g^*(\mu + s - \mu f^*(s))$$

where $g(\,.\,)$ is the chord length distribution. Clearly equations of this kind rarely yield tractable solutions and even if an approximation could be obtained we would then have

to put it into the expression for $K^*(s)$ given in §3.6.1 and invert the resulting expression. Moreover, since we would almost certainly have to model the distribution of chord lengths in order obtain a solution it seems that a more fruitful line of approach would be to model directly the distribution of lengths of intersection with Y. However, before we consider possible models we look at one other approach, based on a sample estimate of $f^*(s)$. Since the Laplace transform of the probability density function of a random variable, Z, is simply $E[e^{-sZ}]$ we can use the data collected on the line probes to form an unbiased estimator of $f^*(s)$ given by

$$\hat{f}(s) = \frac{1}{n_y} \sum_{i=1}^{n_y} \exp(-sy_i)$$

where the $\{y_i\}$ are simply a relabelling of the $\{y_{ij}\}$ introduced in §3.6, after removal of any intercepts that include an end point of a line probe, since these lengths will be censored and will not have the correct distribution. Using this estimator gives

$$K^*(s) = \frac{1}{s} \left\{ \frac{s + (\mu - \tau)[1 - \Sigma\exp(-sy_i)/n_y]}{s + \mu[1 - \Sigma\exp(-sy_i)/n_y]} \right\}$$

which can be inverted only if we can find all of the singularities of the denominator. That is, we need to solve the transcendental equation

$$s = \mu\left(\Sigma\exp(-sy_i)/n_y - 1\right) .$$

For real s the only solution is at s=0 because as s increases the left hand side is monotone increasing and the right hand side is monotone decreasing. For complex s with $\mathrm{Re}(s) \geq 0$ there are no solutions because the real part of the right hand side is negative. However, for complex s with $\mathrm{Re}(s) \leq 0$ there are in general an infinite number of solutions with no obvious simple method available for evaluating them and so it appears that this method will not be helpful to us.

The approach we adopt, then, is to model directly the distribution of intercept lengths with Y by a standard parametric distribution. The most obvious candidates which are defined on the correct interval and which have relatively straightforward

density functions are the gamma, the Weibull, the lognormal and the inverse Gaussian distributions, although of course the list is not exhaustive. One point to notice is that *conditional on being greater than the maximum caliper diameter of any primary grain* the length of intersection of a line probe with Y is exponentially distributed. This follows from the same argument as was used for the intercepts with $X \setminus Y$ since now the history of the process on the line probe cannot affect the transtion probability, π_{10} (see fig. 3.6).

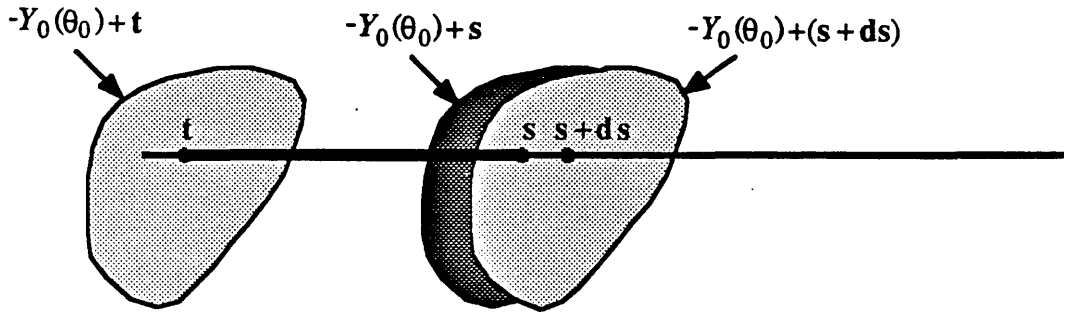


Fig. 3.6 Distribution of Long Intercepts with Boolean Model

The probability that **s+ds** is in $X \setminus Y$ given that **s** is in Y is equal to the probability that there exists a θ_0 such that at least one germ of the process with primary grain $Y_0(\theta_0)$ lies in $-Y_0(\theta_0)+s$ given that no such germs lie in $-Y_0(\theta_0)+(s+ds)$ or in $-Y_0(\theta_0)+t$, where **t** is the point at which the last transition from $X \setminus Y$ to Y occurred. However, conditional on |s-t| being greater than the maximum caliper diameter of a primary grain it is clear from the diagram that the required probability is independent of the history of the process prior to **s**. Therefore, by a similar argument to that used for the intercepts with $X \setminus Y$ we can see that the lengths of the intercepts with Y are exponential conditional on being greater than the maximum caliper diameter.

Thus we would like the distribution of these intercept lengths to have the same behaviour as an exponential distribution in its upper tail. In the lower tail the distribution will be very similar to the distribution of the length of a random chord of an individual grain, since intercepts through overlapping grains will generally be long; the particular characteristics of the chord length distribution will depend very much on the nature of the grains. Figure 3.7 shows simulations of the distribution of the length of intersection of a linear probe with various Boolean models with spherical primary grains.

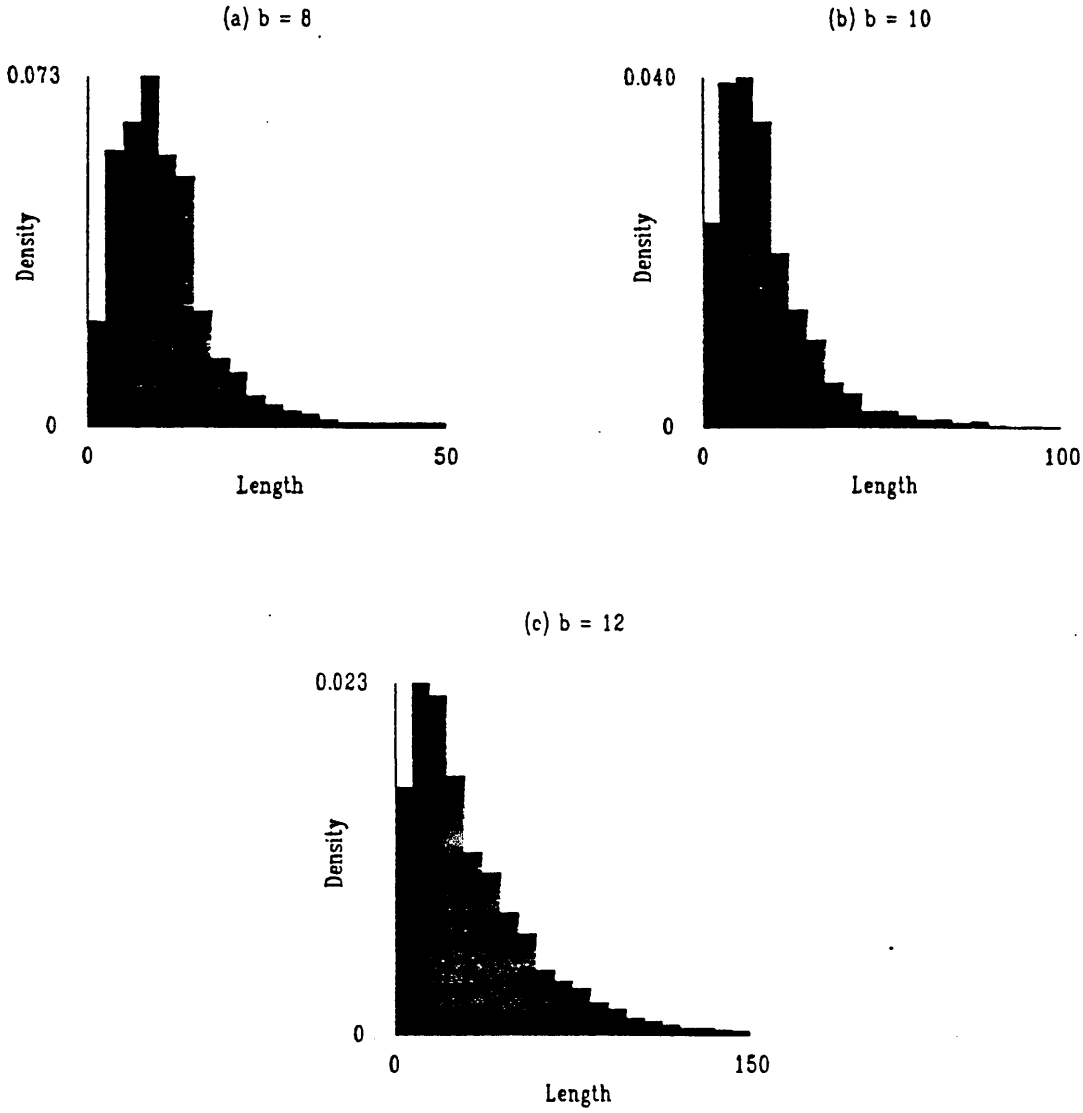


Fig. 3.7 Intercept Length Distributions for Boolean Models of Spheres

The graphs show empirical length distributions for intercepts made with a simulated Boolean model of spheres by independent *IUR* line probes. The sphere centres had intensity 0.001 and the radii were uniform on the interval $(0, b)$, where b is given for each graph. The sample sizes are necessarily random but are of the order of 10^4. Notice that the distribution tends towards an exponential as b increases reflecting the relatively high proportion of overlapping intervals and the fact that intervals have exponential lengths conditional on being longer than $2b$.

The four distributions mentioned above were fitted to each of the simulated data sets shown in fig. 3.7, the lognormal and gamma distributions being fitted by a quasi-Newton algorithm with the package MIX (Macdonald & Green, 1986) whilst the inverse Gaussian and Weibull distributions were fitted by maximum likelihood using a grid search. The parameter values obtained together with the likelihood ratio and (Pearson) χ^2 goodness-of-fit statistics for the grouped data are given in table 3.1. (For details of these standard goodness-of-fit tests see, for example, Cox & Hinkley, 1974, Ch.9.) The parameterisations used were

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \qquad r, \lambda > 0$$

$$f(x) = \frac{1}{x\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right] \qquad \mu \in \mathbb{R}, \sigma > 0$$

$$f(x) = abx^{b-1}\exp[-ax^b] \qquad a, b > 0$$

$$f(x) = \sqrt{\frac{r}{2\pi x^3}} \exp\left[-\frac{r}{2t^2 x}(x - t)^2\right] \qquad r, t > 0$$

for the gamma, lognormal, Weibull and inverse Gaussian distributions respectively, all of course being defined on $(0, \infty)$ since we are concerned with a positive-valued random variable.

Not only does the gamma distribution provide a very good fit in all of the cases, but also it is far superior to the other distributions, and, although this simulation alone cannot justify the use of the gamma distribution as a valid model in general, the evidence is very encouraging, since the gamma distribution is by far the most straightforward to handle of the four distributions considered, when it comes to working with Laplace transforms. Thus we continue the development using the gamma distribution as a model for the distribution of intercept lengths made with Y and discuss its limitations afterwards.

Distribution	Parameters	Likelihood ratio statistic	χ^2 statistic (Pearson)	d.f.
Gamma	r=2.49, λ=0.23	15.60	17.57	11
Lognormal	μ=2.16, σ=0.67	31.08	33.74	11
Weibull	a=0.02, b=1.6	31.61	46.85	11
Inverse Gaussian	t=10.6, r=15.2	56.52	59.35	11

(*a*) $b = 8$

Distribution	Parameters	Likelihood ratio statistic	χ^2 statistic (Pearson)	d.f.
Gamma	r=1.94, λ=0.11	7.85	8.68	10
Lognormal	μ=2.59, σ=0.74	15.27	15.17	10
Weibull	a=0.02, b=1.4	35.41	65.61	10
Inverse Gaussian	t=17.2, r=19.2	48.11	46.40	10

(*b*) $b = 10$

Distribution	Parameters	Likelihood ratio statistic	χ^2 statistic (Pearson)	d.f.
Gamma	r=1.51, λ=0.05	2.56	2.57	13
Lognormal	μ=3.14, σ=0.87	22.62	21.48	13
Weibull	a=0.03, b=1.0	40.62	36.79	13
Inverse Gaussian	t=32.2, r=25.8	67.82	73.84	13

(*c*) $b = 12$

Table. 3.1 Goodness of Fit of Standard Distributions to Simulated Intercept Data

The table gives fitted parameter values and goodness-of-fit statistics for four parametric distributions fitted to data generated as lengths of intersection of an *IUR* line probe with a Boolean model of spheres with spherical primary grains. The radii of the grains have a uniform distribution on (0, b) where b is given in the table. (The data are shown in fig. 3.7.) The intervals in the tails of the distributions were aggregated to avoid cells with very low observed frequencies although this had no effect on the relative performances of the different distributions and little effect on the statistics obtained when compared with the corresponding results calculated with the unaggregated data.

The gamma distribution has density

$$f_Y(y) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y} \qquad y \geq 0, \qquad r, \lambda > 0$$

where r will be referred to as the *index*, and has Laplace transform $\lambda^r/(\lambda+s)^r$, giving

$$K^*(s) = \frac{1}{s} - \frac{\lambda}{r} \left\{ \frac{1 - [\lambda/(\lambda+s)]^r}{s\left(s + \mu - \mu[\lambda/(\lambda+s)]^r\right)} \right\} .$$

In general there is not a closed form for the inversion of $K^*(\,.\,)$ but for integer r the second term is the ratio of two polynomials and therefore there is a simple inversion. Writing $R_r(s; \lambda) = \{(\lambda+s)^r - \lambda^r\}/s$ and $Q_r(s; \lambda, \mu) = \{(\lambda+s)^r(\mu+s) - \lambda^r\mu\}/s$ then

$$K^*(s) = \frac{1}{s} - \frac{\lambda}{r} \frac{R_r(s; \lambda)}{sQ_r(s; \lambda, \mu)}$$

and

$$K(t) = 1 - \frac{\lambda}{r} \sum_{i=1}^{r} \frac{R_r(a_i; \lambda)}{Q_r'(a_i; \lambda, \mu)} \left[\frac{\exp(a_i t) - 1}{a_i} \right]$$

where $a_1, \ldots, a_r$ are the roots of $Q_r(s; \lambda, \mu)$. (We note that this expression is always real-valued even when the roots of Q are not because the coefficients are real and therefore the roots occur in complex conjugate pairs, which ensures that the imaginary part of the sum is zero). The suggestion is to estimate the parameters of the model and then to calculate the covariance function, and hence the variance, for $[r]$ and $[r+1]$ (where $[\,.\,]$ denotes "the integer part of"), re-estimating λ each time conditional on the value of r being used. In the following sections we look at the estimation of the parameters and the performance of the procedure in practice.

3.6.4 Estimating Parameters

The direct application of standard techniques from statistical theory has to be treated with caution in the present context because of the unusual nature of the data. We have established that the marginal distribution of the lengths of intersection, $\{x_{ij}\}$, of a line probe with $X \setminus Y$ is exponential and we are assuming that the corresponding distribution for the intercepts, $\{y_{ij}\}$, made with Y is gamma. However, the joint distributions of both sets of intercept lengths derived from a set of parallel probes involves correlations between the intervals on different lines. This is intuitively obvious but to make the point more rigourously we recall once again the method of proof used to show that the marginal distribution of the $\{x_{ij}\}$ is exponential. In that proof we found the probability that the point **s+ds** lies in Y given that **s** lies in $X \setminus Y$ by considering the region in which a germ of the process could lie to satisfy that condition. If we now consider a set of parallel probes then for any given probe the above transition probability calculated conditional on the data contained in the other probes will be dependent on that data because the region in which a germ of the process can lie to satisfy the necessary conditions will depend on the data contained in the other probes (see fig. 3.8).

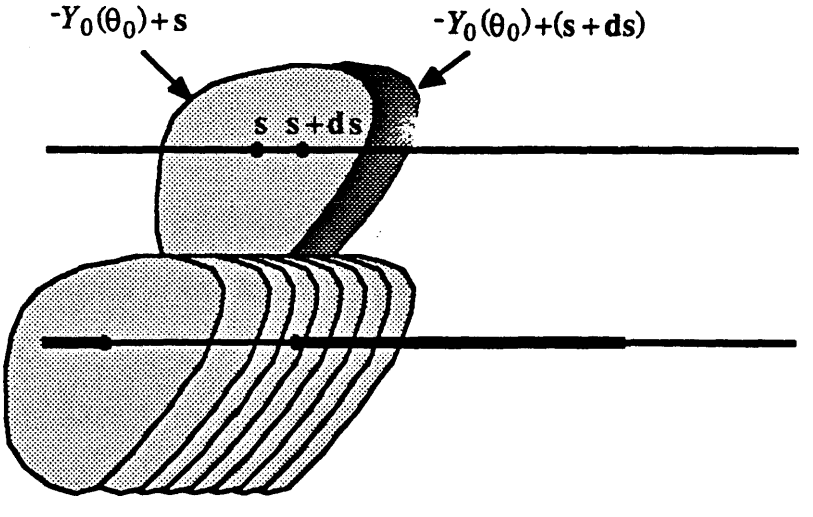


Fig. 3.8 Dependence of Lengths of Intervals on Parallel Probes.

We see that the region in which a germ of the Boolean model must lie in order to cause a transition on one line probe is dependent on the information carried on other probes.

Thus clearly the observations on different lines are not independent although those on the same line are. This is in contrast to the more usual situation encountered where observations on an individual experimental unit are correlated but the units themselves can be treated as independent. The effect in both cases is overdispersion in the population when the correlations are positive, as they would be in this case, and thus the variance in a given sample would tend to underestimate the population variance. This phenomenon is not the only unusual aspect of the data; the sample size is itself a random variable and certainly we would expect positive correlation between the numbers of intercepts on line probes near each other. A further complication is that the intercepts at the ends of the line probes represent incomplete observations and, if included, must be treated as censored data. Despite the complex dependencies in the data the overall picture may not be too bad, depending on the distance between line probes compared with the maximum caliper diameter of the primary grains. The important point is that the spatial extent of correlations will only be as great as the maximum caliper diameter of a primary grain and, because the primary grains are finite in size, regions of the sampling frame separated by a distance greater than the maximum caliper diameter will contribute independently, in some sense, to the whole data set. Recall that in §3.6.2 we had the expression

$$\mathrm{var}(\hat{\pi}_0) = \frac{1}{n}\mathrm{var}(L_i) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{cov}(L_i, L_j)$$

for the variance of the lineal fraction. Now if we increase n whilst keeping d, the distance between adjacent lines, constant then for each i in the second term on the right hand side there are only $O(1)$ nonzero terms; that is for $|i-j| > c$, for some fixed c, the data on probes i and j contribute independently to the whole data set. It is in this sense that we can feel justified in using the sample populations of intercept lengths as if they were simple random samples to estimate the marginal population distributions, provided that both d and n are reasonably large. Notice however that having done this to estimate the covariance function we are *not* ignoring the covariance terms in the

above expression when it comes to calculating the final expression for the variance. If we were simply to use $1/n$ times the sample variance of the L_i to estimate the variance of $\hat{\pi}_0$ we would be underestimating that variance with a relative error of $O(1)$.

We now consider some possible estimators of the parameters μ, λ and r in our model. The exponential parameter, μ, presents no great problems, even when we include the censored data in the estimation. If $x_1, \ldots, x_n$ are i.i.d. exponential observations with mean $1/\mu$ then the loglikelihood is simply

$$n\log\mu - \mu\sum x_i$$

which is maximised by

$$\hat{\mu} = n/\sum x_i.$$

We note that this is also the moment estimator of μ. When we include the censored observations the standard practice is to add to the loglikelihood the log of the survivor function for each censored observation. The survivor function is given by $\Pr(X > x)$ which for the exponential distribution is simply $\exp(-\mu x)$. Thus the loglikelihood becomes

$$n\log\mu - \mu(\sum x_i + \sum x_i^*)$$

where the x_i^* are the censored observations, and this is maximised by

$$\hat{\mu} = n/(\sum x_i + \sum x_i^*).$$

For the gamma distribution there is more scope because the moment estimators do not coincide with the maximum likelihood estimators. The maximum likelihood estimators do not have a simple closed form because the estimating equations involve the digamma function but they are not difficult to find numerically. The moment estimators based on the complete data only are very straightforward to obtain. With the parameterisation of the gamma distribution that we are using the mean and variance are given by r/λ and

r/λ^2 respectively; hence if m and s^2 are the sample mean and variance respectively then $(m^2/s^2, m/s^2)$ are the moment estimators of (r, λ). When we include the censored data the situation is a little less straightforward. Let Y have a gamma distribution with parameters r and λ as above and let the conditional distribution of Z given Y be the uniform distribution on the interval $(0, Y)$. Then the density of Z is given by

$$\frac{\lambda^r}{\Gamma(r)} \int_z^\infty y^{r-2} e^{-\lambda y} dy$$

and we have that

$$E(Z) = E(Y)/2.$$

Similarly we can show that $E(Z^2) = E(Y^2)/3$ and hence we look for weights a_n, b_n and c_n such that

$$a_n E[\sum y_i + 2\sum y_i^*] = r/\lambda$$

and

$$b_n E[\sum y_i^2 + 3\sum (y_i^*)^2] - c_n E[(\sum y_i + 2\sum y_i^*)^2] = r/\lambda^2,$$

where $y_1, \ldots, y_{nu}$ are the uncensored observations and $y_1^*, \ldots, y_{nc}^*$ are the censored observations. Clearly a_n must be $1/(nu+nc)$ whilst a suitable choice for b_n and c_n is

$$b_n = (1+nc/3(nu+nc)^2)/(nu+nc-1) \qquad c_n = 1/(nu+nc)(nu+nc-1).$$

Thus we can find moment estimators of r and λ based on the whole data set which only involve sums of observations and sums of squares of observations and are therefore very easily calculated. This is not necessarily a great advantage in these days of easy access to powerful computers but it may be useful in some circumstances and,

if nothing else, provides initial values for the numerical maximisation of the likelihood, which we turn to now.

Considering first just the complete observations, $y_1, \ldots, y_{nu}$, the loglikelihood is given by

$$nu(r\log\lambda - \log\Gamma(r)) - \lambda\sum y_i + (r-1)\sum\log y_i.$$

Thus the derivative of the loglikelihood involves the digamma function and, although there are approximations to it in the literature which might allow us to obtain closed form expressions for the maximum likelihood estimators, the loglikelihood is sufficiently well behaved to allow numerical maximisation by simple techniques, such as an iterative grid search, only requiring the calculation of the log-gamma function, which is a standard library call on most computers. To incorporate the censored observations we need to add in to the loglikelihood the log of the survivor function evaluated for each censored observation. The loglikelihood now becomes

$$(nu+nc)(r\log\lambda - \log\Gamma(r)) - \lambda\sum y_i + (r-1)\sum\log y_i + \sum\log I_{r,\lambda}(y_i^*),$$

where

$$I_{r,\lambda}(y_i^*) = \int_{y_i^*}^{\infty} u^{r-1}e^{-\lambda u}du\ .$$

The computational burden is now somewhat heavier with the necessity to perform a potentially large number of numerical integrations but even so, even with unsophisticated techniques such as Simpson's Rule for evaluating the integrals, a high degree of accuracy can be obtained very quickly and easily provided fast computers are available.

In §3.6.3 we noted that the Laplace transform of $K(\ .\ )$ can only be inverted easily for integer values of r and therefore we have to calculate the variance of $\hat{\pi}_0$ with r rounded to the integer nearest to the estimated value, or interpolate between the two values calculated with the integers below and above the estimated value. This means

that in practice we should not bother to calculate the joint maximum likelihood estimators of r and λ since we are going to round r to an integer value anyway. However, once we have fixed r, based on the moment estimators, we can then re-estimate λ conditional on the new value of r. Here we can again use either maximum likelihood or moment estimation and use just the uncensored data or the whole data set. The estimators are similar to those derived above and do not need to be given here but all have been calculated in the examples that follow, for illustrative purposes.

3.6.5 Simulated Examples

In this section we look at the application of the methods outlined in earlier sections to some simulated data sets. We compare the results obtained by the techniques of the preceeding sections with the theoretical values when the underlying model is a Boolean model of spheres with uniform radius distribution and look at the validity of some of the assumptions that have been made.

First we calculate the variance of $\hat{\pi}_0$ using the form of the covariance function derived under the gamma assumption, for a range of values of r, λ and μ, and for a lineal analysis consisting of 20 line probes of length 200 and separation 10. The values of r are integer to allow the inversion of the Laplace transform and for each value of r the values of λ used are such that the mean of the gamma distribution lies between 10 and 50. This choice of values is completely arbitrary (in fact being for convenience in producing images of the simulated models) but the scale of the observations is unimportant; of more importance are quantities such as the expected volume fraction occupied by Y. For each pair of values of r and λ a range of values of μ is considered corresponding to values of π_0 between 0.2 and 0.8. Some of the results are shown in table 3.2 and fig. 3.9. We notice that for fixed r and π_0 the variance increases as the mean of the gamma distribution increases (except in the extreme case of $\pi_0 = 0.8$ and $r = 5$). This corresponds to a reduction in the sample size (ie total number of intercept lengths) and also to an increase in the variance of both the exponential and gamma distributions. On the other hand if we fix π_0 and the mean of the gamma distribution and increase r then the variance decreases, corresponding to a decrease in the variance of the gamma distribution. Finally, for fixed values of r and λ the variance decreases, in general, as π_0 (and hence μ) increases although the value of π_0 for which the variance is maximum value is greater than zero, approaching zero as r and λ increase.

Index	Mean		
r	10	20	50
1	0.00148	0.00488	0.02296
3	0.00058	0.00195	0.01119
5	0.00044	0.00147	0.00841

(a) $\pi_0 = 0.2$

Index	Mean		
r	10	20	50
1	0.00099	0.00325	0.01751
3	0.00038	0.00111	0.00689
5	0.00028	0.00076	0.00495

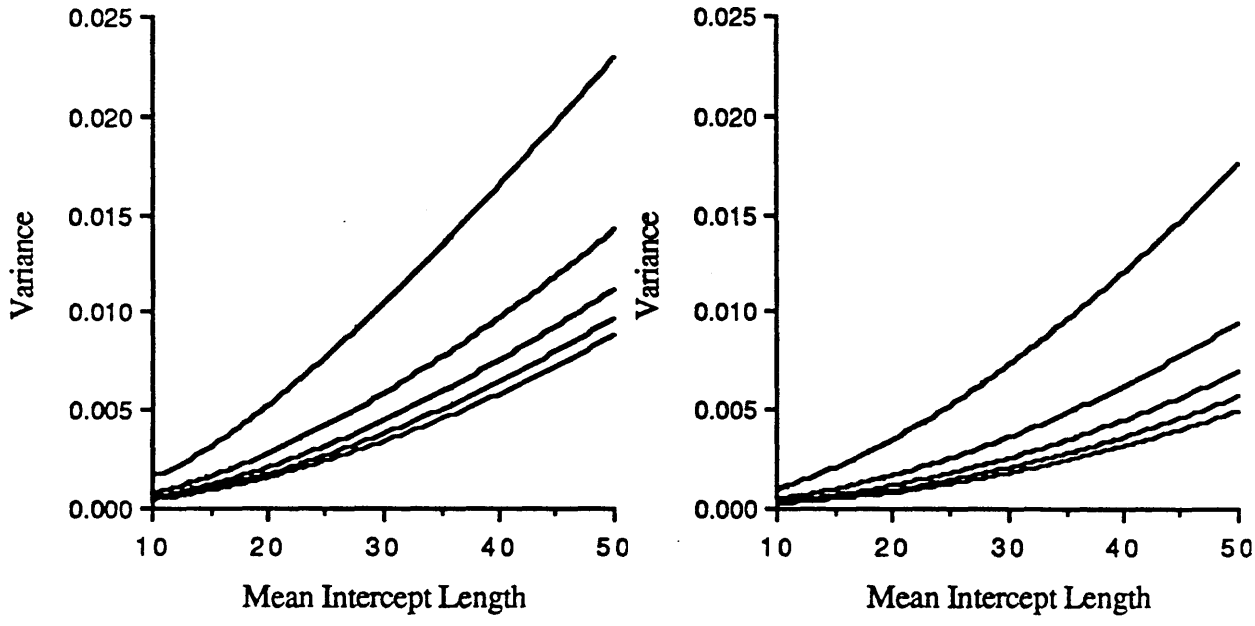
(b) $\pi_0 = 0.5$

Index	Mean		
r	10	20	50
1	0.00016	0.00041	0.00224
3	0.00009	0.00009	0.00037
5	0.00009	0.00005	0.00008

(c) $\pi_0 = 0.8$

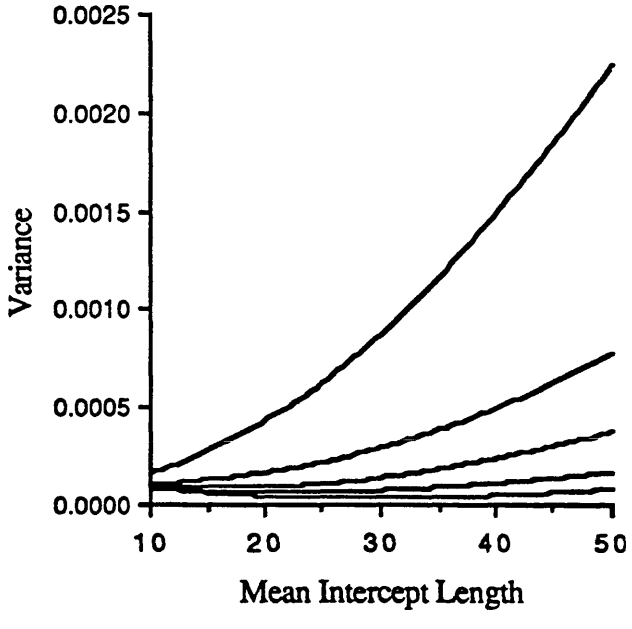
Table 3.2 Variance of Lineal Fraction

The table shows the variance of the lineal fraction when the intercept lengths with the phase of interest have a gamma distribution with index and mean given in the table. The intercept lengths with the reference phase have an exponential distribution and the volume fraction is π_0. The results are for a lineal analysis based on 20 line probes with separation 10 and length 200.



(a) $\pi_0 = 0.2$

(b) $\pi_0 = 0.5$



(c) $\pi_0 = 0.8$

Fig. 3.9 Variance of Lineal Fraction

The graphs show the variance of the lineal fraction when the intercept lengths with the phase of interest have a gamma distribution. The five lines on each graph correspond to $r = 1, 2, 3, 4, 5$ with $r = 1$ being the lowest and $r = 5$ being the highest. The intercept lengths with the reference phase have an exponential distribution and the volume fraction is π_0. The results are for a lineal analysis based on 20 line probes with separation 10 and length 200.

The increase in π_0 represents an increase in the expected sample size and therefore the general trend is not surprising. The results for small π_0 are not so easy to explain but anyway, we would have to be careful if we were to use the gamma model for small values of π_0 because this implies little overlapping of the grains of the Boolean model and less justification for the use of the gamma distribution.

Next we simulate Boolean models of spheres whose radii have a uniform distribution on the interval $(0, b)$, for $b = 6$, 8, 10 and 12, and whose centres have intensity 0.001 and we estimate the volume fraction by a lineal analysis with 20 parallel line probes of length 200 and separation 10. (In fact we simulate the induced process of circles in the plane with intensity $0.001b$ and radius density given by the well known formula of Wicksell's problem). For each of these simulations we estimate the parameters of the exponential and gamma distributions of intercept lengths and calculate the corresponding variance of the lineal fraction. We then compare the results with the known theoretical values. Table 3.3 gives the theoretical values of π_0 together with the theoretical variance of the lineal fraction; it also gives the actual value of the lineal fraction in the simulation together with the estimated variance under the gamma model. Table 3.4 gives the various estimates of the parameters of the model for the particular simulations and figure 3.10 demonstrates a graphical method for approximating the estimated variance. The values of the estimated variance in table 3.3 were calculated by linear interpolation from results computed for a large set of values of r, λ and μ.

Although the simulations cover only a limited range of examples of a particular form of the Boolean model it is heartening to see that in the four cases studied the gamma based model comes reasonably close to the theoretical value. Of course a Boolean model of spheres is a very special case but on the other hand it is one of the few for which an alternative estimator exists and the four simulations cover a wide range of conditions from a fairly sparse model ($b = 6$) to one with a high degree of overlapping ($b = 12$).

	b			
	6	8	10	12
π_0	0.202	0.415	0.649	0.836
$\text{var}(\hat{\pi}_0)$	0.00028	0.00059	0.00071	0.00046
observed $\hat{\pi}_0$	0.181	0.405	0.660	0.839
estimated $\text{var}(\hat{\pi}_0)$	0.00026, 0.00023	0.00056, 0.00046	0.00066, 0.00044	0.00074, 0.00023

Table 3.3 Theoretical and Estimated Values for Simulated Boolean Model

The table gives the theoretical values of the volume fraction, π_0, and the variance of the lineal fraction, $\text{var}(\hat{\pi}_0)$, for a lineal analysis of a Boolean model with spherical primary grains. The spheres have uniform radius distribution on the interval $(0, b)$ and their centres form a Poisson point process of intensity 0.001. The lineal analysis is by 20 lines of length 200 and separation 10. The third and fourth rows of the table give the values of the lineal fraction and its estimated variance for realisations of the Boolean model and lineal analysis. The estimated variance is calculated under the assumption that the intercepts of the line probes with the Boolean model have a gamma length distribution; the two values given are calculated with the index of the gamma distribution equal to the integer below and above the estimated value, respectively. The parameter estimates are given in table 3.4; those used for calculating the variance are the MLE of μ incorporating the censored data and the conditional MLE of λ calculated with the censored data included.

There is considerable variation in the estimates of the parameters and clearly where there is considerable dependence in the data we should be wary of treating the likelihood as if the observations were independent. To get an idea of how bad the maximum likelihood estimates might be a further study was carried out. For each of the cases in the simulation described above 50 independent realisations were generated and just the exponential parameter was calculated. Also the sample size was recorded for each realisation so that the effect of the variation in the sample size could be assessed. The reason for calculating only the exponential parameter, apart from computational cost, is that we know that the exponential distribution is marginally exact and, furthermore, we can calculate the true value of the parameter.

	b			
	6	8	10	12
Theoretical value	0.038	0.067	0.105	0.151
MLE - complete data only	0.041	0.071	0.114	0.177
MLE - censored data included	0.031	0.061	0.104	0.164

(a) Estimators of μ

	b			
	6	8	10	12
Moment based - complete data only	3.614 0.604	3.413 0.339	1.991 0.115	1.419 0.054
- censored data included	4.028 0.665	3.668 0.342	1.997 0.108	1.656 0.048
MLE - complete data only	3.614 0.604	3.120 0.317	2.489 0.114	1.596 0.061
- censored data included	3.021 0.492	3.109 0.294	2.246 0.119	1.553 0.048

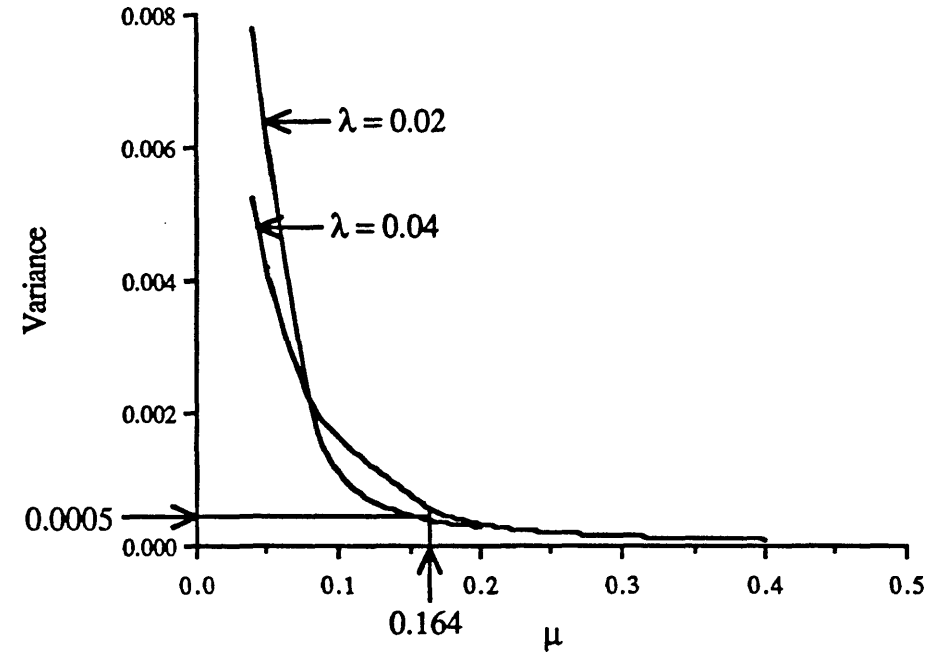
(b) Joint Estimators of r and λ

	b							
	6		8		10		12	
	$r=3$	$r=4$	$r=3$	$r=4$	$r=2$	$r=3$	$r=1$	$r=2$
Moment based	0.448	0.651	0.280	0.373	0.108	0.162	0.029	0.058
MLE	0.491	0.661	0.284	0.382	0.106	0.161	0.031	0.065

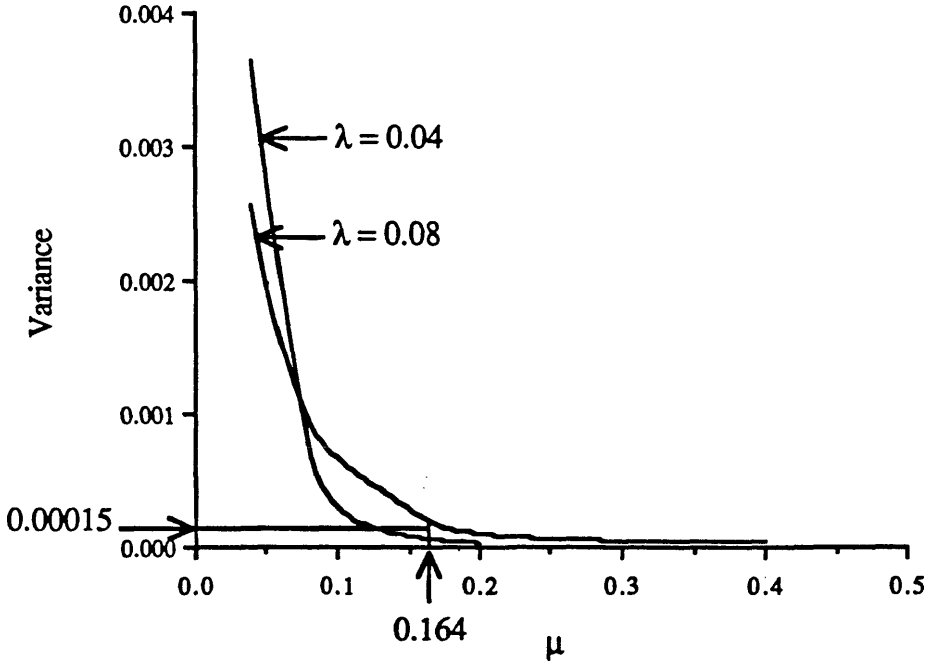
(c) Conditional Estimators of λ given r.

Table 3.4 Estimated Parameter Values for Simulated Boolean Models

The estimates in the table are for the simulations of the lineal analysis of a Boolean model of spheres described in table 3.3 and in the text. μ is the parameter of the exponential distribution of the intercept lengths made with the reference phase and λ and r are the parameters of the assumed gamma distribution for the intercept lengths with the phase of interest. For the parameters of the gamma distribution the maximum likelihood estimators were found numerically using the appropriate moment estimators as starting values. The estimation was based on the marginal distribution of the data and ignored the dependencies and the randomness of the sample size.



(a) $r = 1$



(b) $r = 2$

Fig. 3.10 Graphical Interpolation for Approximating Variance

The estimated variance is calculated for a large range of values of r, λ and μ. For a specific case we then plot the variance for the closest values of the parameters and interpolate between the plotted values. Illustrated is the simulation with b=12; for r=1 we have λ=0.031 and for r=2 we have λ=0.065, whilst μ=0.164 in both cases (see table 3.4).

The results of the simulations are shown in table 3.5. The observed variances are simply sample variances of the observed parameter estimates. The expected variances are the theoretical variances for the equivalent estimators given a fixed sample size of i.i.d. observations with the true parameter, where the sample size used is the mean of the observed sample sizes. These values can be obtained easily via the distribution of the maximum likelihood estimator of the mean parameter, which is $1/\mu$.

		b			
		6	8	10	12
		Values of $\hat{\mu}$			
True value		0.038	0.067	0.105	0.151
Observed	- complete data only	0.045	0.074	0.111	0.162
	- censored data included	0.033	0.063	0.100	0.152
		Standard deviation of $\hat{\mu}$			
Complete data only	- observed	0.0042	0.0070	0.0088	0.0202
	- expected	0.0037	0.0056	0.0089	0.0153
Censored data included	- observed	0.0033	0.0060	0.0076	0.0181
	- expected	0.0028	0.0048	0.0081	0.0143
		Sample size			
Complete data only	- mean	104	146	140	97
	- variance	73	85	48	71
Total sample size	- mean	136	169	154	104
	- variance	79	85	46	77

Table 3.5 Performance of Estimators of μ

The table gives the sample mean and standard deviation of estimates of the exponential parameter of the length distribution of intercepts of a line probe with the reference phase of a Boolean model of spheres, based on lineal analyses of 50 realisations of the Boolean model. The details of the model and the analyses are given in table 3.3 and in the text. The estimates are the standard MLEs for i.i.d. observations, both with and without censored data. The true values are calculated from known results about Boolean models.

The first striking thing is that, apart from the case b=12, the estimators appear to be biased. The differences between the true values and the means of the observed values are greater than 8 times the estimated standard error of the mean for b=6, 8 and 10, even when the censored data is included, whereas for b=12 the difference is less than one standard error. Further investigation shows that significant correlation exists between the sample size and the sum of the observations in several cases, the ratio of these quantities being the estimator in question. When the censored observations are included the correlation is significantly negative for b=6 and b=8 and significantly positive for b=12. This is explained by the fact that for sparse models (ie b=6 and b=8) an increase in the sample size is due to more grains hitting the line probes, with a resulting decrease in the total length of intersection with the reference phase. Conversely, when the volume fraction is high (b=12), and there is a high degree of overlapping of particles, an increase in sample size results from less overlaps being hit, which would be more likely to happen when particles are hit near their edges, giving a decrease in the total length of intersection with the phase of interest. However, this does not explain the bias in the estimators since we would normally expect the ratio of two negatively correlated random variables to be positively biased and *vice versa*; clearly there is scope for further investigation on this point.

Despite the mysterious bias we notice from the table that the estimators which incorporate the censored data are more efficient and have smaller bias than those that do not, as we would hope, but that both are less than efficient than would be the case for i.i.d. variables.

3.6.6 Limitations of the Method

The method we have developed for estimating the variance of the lineal fraction and applied in the preceeding section has some obvious limitations. First, it assumes that the specimen can be regarded as the realisation of a Boolean model with convex primary grains. Secondly, it assumes a gamma distribution for the lengths of the intercepts of an *IUR* line probe with the phase of interest. Furthermore, the estimation of the parameters of the distributions of intercept lengths with the two phases ignores the dependencies in the data and having estimated the parameters it is necessary to interpolate between results calculated for integer values of the index of the gamma distribution in order to make inversion of the Laplace transform possible.

As far as the assumption of a Boolean model is concerned this is an issue which arises in almost all branches of statistics. It is clear that there exist situations for which the Boolean model is appropriate (see Stoyan, Kendall & Mecke, 1987, Ch.3, for a list of situations in which the Boolean model has been applied and also a critical assessment of Boolean models) and we have mentioned earlier that Ripley (1988) has suggested some ideas for goodness-of-fit tests for random set models. However, this is not the subject of this thesis and since the validity of the assumption will depend on the context of the investigation being undertaken the responsibility for deciding whether the model is appropriate must lie with the experimenter. There are, of course, other models available and the development of a similar methodology for those models would be interesting. The most obvious classes of models are inhibiton (hard core) models, for densely packed, non-overlapping particles and cluster models, for sparsely distributed, highly irregular specimens, where the clusters of overlapping grains model the irregularly shaped patches of the phase of interest. However, it is clear that results for these models will be considerably less easy to obtain and the approach we have used may well not be justifiable if the model itself introduces a whole new set of computational problems. Indeed, even with the Boolean model of convex grains we have introduced some computational problems which detract from the appeal of the general scheme.

The gamma distribution assumption is certainly open to criticism. However, although we have no theoretical arguments to justify it, we have some encouraging empirical evidence, albeit for a particular sub-class of models. Further work in this area should include testing the assumption on a wider class of models and on real data sets for which a Boolean model of convex grains is appropriate. Of course we could start our modelling at the level of the intercept length distributions. The only point at which the Boolean model enters the calculations in the preceeding work is in the derivation of the exponential distribution for the lengths of intersection with the reference phase. This "exact" distributional property of the data is only as good as the Boolean model and an alternative is to model the two length distributions directly without relating them to any underlying random set model in the whole specimen. This would relieve us of the need to justify the Boolean model assumption and avoid the question of whether whatever distribution we use for the intercepts with the phase of interest is an accurate description of the "true" distribution (that is, the distribution for intercepts with a Boolean model). On the other hand we would lose the possibility of using the Boolean model for interpretion.

The parameter estimation problem is also worrying. The use of maximum likelihood estimation with a simplified form for the likelihood, which does not take account of the dependencies in the data, leads us to ask what the effect on the parameter estimates might be, and the answer to that question is not obvious, although the results of the simulations carried out to examine the behaviour of the estimator of the parameter of the exponential length distribution suggest the existence of some undesirable properties. It may be possible to develop techniques which diminish the effect of the dependencies. For example, suppose that we number the line probes 1, ... , n and split them into two groups, one with the odd numbered probes and the other with even numbered probes. We then perform the estimation procedure twice, once for each group, and combine the estimates. Since within each group every pair of probes must be at least twice as far apart as the separation distance in the original set we would hope that the dependencies would be considerably reduced.

The need to interpolate between results for integer values of the index of the gamma distribution may or may not have a significant effect. When the index is small the shape of the distribution may change markedly between the two different values and we should be cautious in interpreting the results, but for larger values of the index the two values for the variance between which the interpolation is done may be fairly close anyway (see table 3.3) and then we would not be too concerned with this limitation.

In conclusion, whilst noting that there are several limitations on the method we have also seen that it can produce good agreement with theoretical values in some cases. The advantage of using a simple model such as the one we have used is that we can obtain results on the basis of a small number of parameters, which we estimate from the data, and which may serve a rôle in a wider context. Moreover, the method is particular to a specific form of stereological study, namely lineal analysis, and avoids the need for a second test set to be applied to the specimen in studies for which it has been decided, perhaps on the basis of other considerations, to use that form of study.

3.6.7 Alternative Approaches

One of the main drawbacks of the approach taken above is that even for a simple physical model we have great difficulty in finding the corresponding statistical model, and indeed we ignored some of the dependencies that were known to exist in the data when estimating the parameters of the statistical model. Recent work by Diggle & Gratton (1984) suggests that in situations such as this an alternative to formal likelihood theory is to generate data from the physical model for different values of the parameters which describe it and to form an "empirical" likelihood in order to estimate values of the model which fit the sample best. This is particularly appealing in our case, provided that we are prepared to make fairly rigid assumptions about the nature of the primary grains of the Boolean model, since the simulation of a Boolean model is very easy and relatively quick. For example, supposing that we assume the Boolean model to have spherical primary grains whose centres have intensity λ and whose radii have density $f_R(r; \theta)$, then a knowledge of λ and θ clearly tells us all we need to know about the model. Given that we have some idea about the values of λ and θ we simulate Boolean models for a range of values of the parameters and collect observations of the type in our original sample (eg intercept lengths on linear probes). For each value of the parameter vector this process is repeated many times and an empirical density is built up for the observations in the sample, either by kernel density estimation or simply by forming a histogram. Thus a likelihood is evaluated for the sample at the different parameter values and after some form of smoothing it is maximised.

The technique is certainly appealing in its simplicity and can be seen to be quite effective in some examples. However, it is not clear how badly it is affected by dependencies of the type encountered in our context and how well it performs on a multi-dimensional parameter space. It also requires quite severe restrictions to be placed on the Boolean model to be realisable in practice. To illustrate how it works in a one-dimensional case the technique was performed on simulated data similar to that used in the previous section.

A realisation of a Boolean model of spheres was generated with the sphere centres having intensity 0.001 and uniform radius distribution on the interval (0, 10). A lineal analysis was performed using 20 lines of length 200 and separation 10 and the lengths of the complete intercepts with the phase of interest were grouped into 10 classes of equal width. Then for each of a range of values of λ, including 0.001, 100 further realisations were generated with intensity λ, and the observations for each set of realisations were grouped using the same intervals as for the sample. The resulting multinomial densities were used to evaluate the likelihood for the sample at each value of λ. The results of two independent repetitions of this procedure (out of a large number carried out) are shown in figure 3.11.

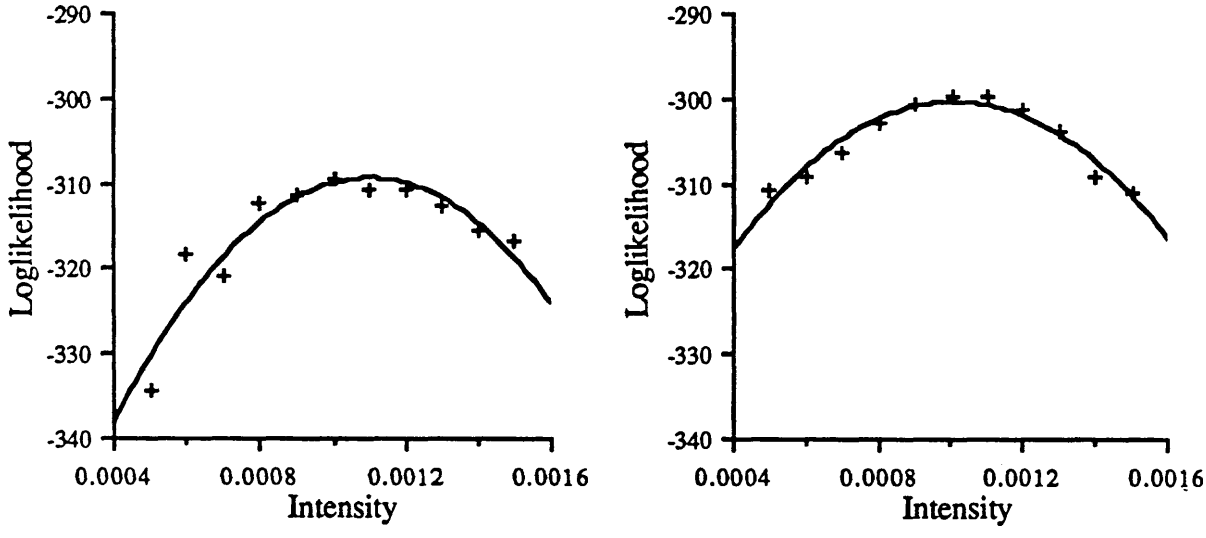


Fig. 3.11 Empirical Loglikelihood for the Intensity of a Boolean Model

The loglikelihoods in the figure are constructed from large numbers of simulations of the model for each of a set of parameter values, the simulated data for each value being grouped to form an estimated multinomial density of intercept lengths. The two graphs represent two repetitions of the procedure using independent simulated samples but the same estimated density. The true value of the intensity is 0.001 and the estimates from the quadratic approximation (fitted by least squares) are 0.0011 and 0.0010 respectively.

There are several drawbacks to this procedure, some of which we have noted already. Obviously the restrictions on the Boolean model necessary to make realisations practical may be unacceptable in some circumstances. Also there is no obvious way of incorporating the censored data into the analysis; this may result in throwing away a large amount of information. Thirdly, as with the previous approach, it is not clear what is the effect of the dependencies in the data and how to allow for them in the analysis. Results from simulations with different distances between the line probes show that the procedure breaks down as the lines become very close together; more seriously, the estimated density does not seem to improve as the size of the simulations is increased.

In the cases we have been looking at, where the primary grains are spherical, it is often possible to estimate the parameters of the Boolean model directly. For example, for the case used in the simulations in §3.6.5 where the radii of the spheres are uniformly distributed on $(0, b)$ and their centres are a realisation of a Poisson process of intensity λ, the expected value of the lineal fraction is $1-\exp\{-\pi\lambda b^3/3\}$ and the parameter of the exponential distribution of intercept lengths with the reference phase is $\pi\lambda b^2/3$; therefore we can estimate λ and b from the observed lineal fraction and the estimated exponential parameter. Furthermore, the covariance function has a simple form allowing easy calculation of the estimated variance of the lineal fraction. If such strong assumptions about the form of the physical model are felt to be justified then clearly the estimation problem is a much simpler one than those outlined above; all that is required is the total length and the number of complete intercepts with the reference phase.

3.7 Stereology in a Wider Context

In the introduction to Chapter 2 we noted that our ultimate interest in any stereological study is almost certain to be centred on a wider population than the specimens under study, probably through a model of some kind in which we wish to fit parameters and/or test hypotheses. In this section we look at the way the ideas of the previous sections might fit into such a wider context.

The analysis of the microstructure of concrete and cement is an active area of research of obvious importance in which stereology plays an important rôle. There are many questions of interest to materials scientists relating to various characteristics of these materials but here we just consider two questions of possible interest:

1) What is the proportion of Calcium Hydroxide ($Ca(OH)_2$) in the cement?

2) How heterogeneous is the $Ca(OH)_2$ in the cement?

The first question is concerned with the mean amount of a particular phase whereas the second is related to the variability of that phase throughout the population. Here the word population can be interpreted in a very general way; it might well refer to a certain production method rather than a deterministic, physical batch of cement and in this sense the ideas of model based stereology are particularly relevant. However, to find answers to the questions by statistical methods we have to be more specific in order to define a model. With the model we can translate the questions into problems of esimation, inference and hypothesis testing. The model and the questions to be answered will give some idea of how to design the sampling scheme but the model will itself be dependent on the nature of the observations that are possible to obtain.

If the population is a finite, physical batch of cement, for example a building or a road, we might be interested in forming a model which includes the spatial component of the situation. For example the second question might reduce to estimating the parameters in a function which describes the proportion of $Ca(OH)_2$ as a function of location.

This approach would not be appropriate in the more abstract setting where the "population" refers to a production process, and we consider that situation now. The

starting point will be a series of samples from the process each of which has a certain proportion of $Ca(OH)_2$, which we must estimate by stereological means. The quantities we are interested in ultimately are the mean proportion of $Ca(OH)_2$ and a measure of the variability of this proportion between samples. Apart from the usual sampling error an extra component of variability is introduced via the stereological estimation. It is to cope with this that the methods of the previous section are used.

We take n blocks of cement from the process and we let v_i be the proportion of $Ca(OH)_2$ in block i (i=1, ... ,n). We analyse m parallel planar sections from each block with a lineal analysis consisting of l parallel linear probes on each section. The sampling mechanism is summarised in figure 3.12. It can be assumed that the blocks are sufficiently small and well spaced that the material within a block can be considered to be homogeneous whilst different blocks are independent.

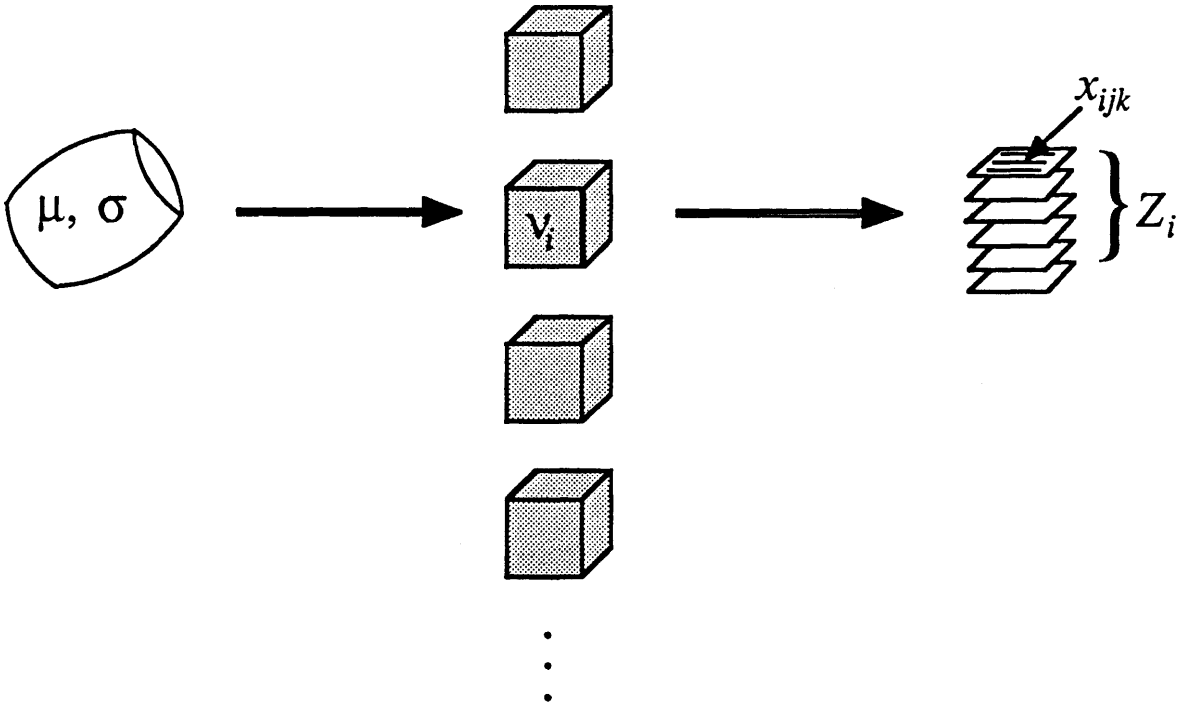


Fig. 3.12 Sampling Scheme for Cement Production Problem

The quantities of interest are the proportion, μ, of $Ca(OH)_2$ in cement made by a certain process and the heterogeneity of the $Ca(OH)_2$ phase, represented by σ (i.e. some measure of dispersion). Each sample has a proportion v_i of the phase of interest and the data consist of lengths of intercepts made with linear probes on planar sections through the sample. The total intercept length for a sample is represented by Z_i.

The data consists of observed intercept lengths on the linear probes; we let x_{ijk} be the total length of intersection of $Ca(OH)_2$ with line k on plane j in block i. Writing

$$Z_i = \sum_{j=1}^{m}\sum_{k=1}^{l} x_{ijk}$$

we have

$$E(Z_i|\nu_i) = \nu_i \qquad \text{var}(Z_i|\nu_i) = \sigma_i^2$$
$$E(\nu_i) = \mu \qquad \text{var}(\nu_i) = \sigma^2.$$

This is the familiar *random effects model*; within this model we are interested in μ and σ^2 whilst $\{\nu_i\}$ and $\{\sigma_i^2\}$ are nuisance parameters. In a standard application we would have b blocks and k independent observations in each block and the covariance matrix for the observations would be block diagonal, with the submatrix corresponding to block i having $\sigma^2+\sigma_i^2$ on the leading diagonal and σ^2 everywhere else. Interest in μ would normally involve estimation by weighted least squares or quasi-likelihood and interest in σ^2 would indicate using components of variance. The unknown $\{\sigma_i^2\}$ would be estimated by the sample variances within blocks.

In our context we know that each x_{ijk} is marginally unbiased for ν_i but observations from the same block are correlated, with a correlation structure reflecting their spatial arrangement, and so the usual estimator of within blocks variance cannot be used here. However, we have developed an estimator of σ_i^2 in §3.6 and therefore we can work in terms of the $\{Z_i\}$ directly.

An approach that is often used for this type of situation is an *empirical Bayes* approach. This implies making distributional assumptions about the data in the form of a *hierarchical model* in which the parameters of each stage are assumed to have a distribution parameterised by "super parameters" of the next stage. The empirical part comes from introducing empirical estimates of some of the parameters, rather than estimating the full parameter set from the likelihood.

Suppose that $Z_i|\nu_i$ has a Normal distribution with mean ν_i and variance σ_i^2 and that ν_i has a Normal distribution with mean μ and variance σ^2. Then, assuming the $\{Z_i\}$ to be independent, given $\{\nu_i\}$, and the $\{\nu_i\}$ to be independent, the joint distribution of $Z_1, \ldots, Z_n$ is multivariate Normal with mean $(\mu, \ldots, \mu)^T$ and diagonal covariance matrix with $\sigma^2+\sigma_i^2$ as the ith element on the diagonal. We replace $\{\sigma_i^2\}$ by estimates obtained by the techniques of the previous section and then estimate μ and σ^2 by standard maximum likelihood.

Chapter 4 Conclusions

We have demonstrated in this thesis that there is an important rôle to be played by statistical theory in stereology, both in the analysis and design of experiments. The difficulty is in finding a compromise between realism and tractability; whereas results concerning mean values of estimators can be obtained under very general conditions we find that the geometric aspect of stereology makes the derivation of second order results very difficult without the imposition of strict assumptions. Furthermore, sampling designs are restricted by the physical destruction of specimens necessary to make measurements. Against this background we have assessed the relative efficiencies of various estimators of volume, both absolute and relative, in a wide variety of contexts.

In Chapter 2 we looked at the estimation of volume from a design-based point of view, in which the randomisation results solely from the sampling design. The main emphasis was on systematic sampling schemes and the use of unbiased counting rules in situations where the phase of interest is the union of disjoint particles. We found that systematic and stratified sampling schemes can be shown to be more efficient than completely random ones, under very mild conditions, and that therefore the restrictions that need to be placed on sampling schemes to make them realisable in practice are, in fact, desirable anyway. This improved efficiency is a result of the principle of stratification and not the principle of antithetic variates, as has been suggested (Mattfeldt, 1987) (see pp. 20-22). The problem still remains as to how to deal satisfactorily with the estimation of the variance of such estimators. The key is in the covariogram and more work is needed in this area.

When the phase of interest is embedded in an opaque reference phase we find that the same techniques are applicable but that the loss of efficiency is more than we might expect. The behaviour of systematic and stratified schemes is more erratic and we should be aware of the possibility of periodicities inflating the variance when the phase of interest is composed of several disjoint parts; also the effect of the variation in the number of sections hitting the phase of interest may be significant when assessing

the relative efficiency of estimators.

For specimens containing a phase of interest composed of many disjoint particles the recently developed disector of Sterio (1984), based on unbiased sampling of particles, is appropriate. We examined the disector in some detail, taking a different approach from that of Sterio, which enabled us to examine its use in the context of finite population sampling theory (see pp. 32-69). Not only were we able to achieve realisable sampling schemes but we were able to assess the relative efficiencies of different schemes and estimators, giving some indications as to how we can use the disector effectively. By contrast, we found that the selector has many drawbacks, being a victim of blind faith in a single principle, namely unbiasedness, to the exclusion of all others (see pp. 70-79).

Chapter 3 is concerned with volume estimation from a model based point of view. We looked principally at Boolean models and derived an estimator for the variance of the lineal fraction when analysing a Boolean model of convex grains by a grid of parallel lines. The main drawbacks of the work are the assumptions that need to be made and we would like to be able to generalise the results to incorporate non-convex primary grains and non-Poisson point processes for the germs, but nevertheless the Boolean model is capable of representing many diverse spatial patterns and has been used in many situations. The key to the calculation of the variance is the covariance function and we have discussed alternative methods of estimating the covariance function, both within the framework of the Boolean model and more generally.

This thesis has shown some ways in which statistical theory and techniques can be applied to stereology. In all cases modelling assumptions need to be made but this is always the case in statistics. It is up to the stereologist to decide which assumptions are appropriate and acceptable in any particular situation.

REFERENCES

Baddeley, A.J., Gundersen, H.J.G. & Cruz Orive, L.M. (1986) Estimation of surface area from vertical sections. *J. Microsc.*, **111**, 219-223.

Besag, J.E. (1974) Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* B, **36**, 192-236.

Besag, J.E. (1986) On the statistical analysis of dirty pictures. *J. R. Statist. Soc.* B, **48**, 259-302.

Christakos, G. (1984) On the problem of permissible covariance and variogram models. *Water Resources Research*, **20**, 251-265.

Cochran, W.G. (1977) *Sampling Techniques*. Wiley, New York.

Coleman, R. (1979) *An Introduction to Mathematical Stereology*. Memoirs No.3, Dept. of Theoretical Statistics, University of Aarhus, Denmark.

Cox, D.R. (1962) *Renewal Theory*. Methuen, London.

Cox, D.R. & Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman & Hall, London.

Cox, D.R. & Smith, W.L. (1961) *Queues*. Methuen, London.

Cressie, N. (1985) Fitting variogram models by weighted least squares. *Mathematical Geology*, **17**, 563-586.

Crofton, M.W. (1885) Probability. *Encyclopaedia Britannica*, IXth edition.

Cruz Orive, L.M. (1980) Best linear unbiased estimators for stereology. *Biometrics*, **36**, 595-605.

Cruz Orive, L.M. (1987) Particle number can be estimated using a disector of unknown thickness: the selector. *J. Microsc.*, **145**, 121-142.

Cruz Orive, L.M. & Weibel, E.R. (1981) Sampling designs for stereology. *J. Microsc.*, **122**, 235-257.

Davy, P. & Miles, R.E. (1977) Sampling theory for opaque spatial specimens. *J. R. Statist. Soc.* B, **39**, 56-65.

Diggle, P.J. & Gratton, R.J. (1984) Monte Carlo methods of inference for implicit statistical models, *J. R. Statist. Soc.* B, **46**, 193-227.

Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, **PAMI-6**, 721-741.

Gundersen, H.J.G. (1977) Notes on the estimation of the numerical density of arbitrary profiles: the edge effect. *J. Microsc.*, **111**, 219-223.

Gundersen, H.J.G. (1986) Stereology of arbitrary particles. *J. Microsc.*, **143**, 3-45.

Gundersen, H.J.G. & Jensen, E.B. (1987) The efficiency of systematic sampling in stereology and its prediction. *J. Microsc.*, **147**, 229-263.

Gundersen, H.J.G. & Østerby, R. (1981) Optimizing sampling efficiency of stereological studies in biology: or "Do more less well!". *J. Microsc.*, **121**, 65-73.

Isham, V. (1981) An introduction to spatial point processes and Markov random fields. *International Statistical Review*, **49**, 21-43.

Kellerer, A.M. (1986) The variance of a poisson process of domains. *J. Appl. Prob.*, **23**, 307-321.

Kendall, D.G. (1974) Foundations of a theory of random sets. In: *Stochastic Geometry*, pp. 322-376, (eds. Harding, E.F. & Kendall, D.G.). Wiley, New York.

Macdonald, P.D.M. & Green, P.E.J. (1988) *MIX, An Interactive Program for Fitting Mixtures of Distributions, release 2.3*. Ichthus Data Systems, Ontario.

Matérn, B. (1960) *Spatial Variation*. Medd. Statens Skogsforskningsinst., **49**, 1-144.

Matérn, B. (1985) Estimating area by dot counts. In: *Contributions to Probability and Statistics in Honour of Gunnar Blom*, pp. 243-257, (eds. Lanke, J. & Lindgren, G.). University of Lund, Lund.

Matheron, G. (1975) *Random Sets and Integral Geometry*. Wiley, New York.

Mattfeldt, T. (1987) Volume estimation of biological objects by systematic sections. *J. Math. Biol.*, **25**, 685-695.

Miles, R.E. (1978a) The importance of proper model specification in stereology. In: *Geometrical Probability and Biological Structures: Buffon's 200th Anniversary* (eds. Miles, R.E. & Serra, J.), *Springer Lecture Notes in Biomathematics*, **23**, 115-136.

Miles, R.E. (1978b) The sampling, by quadrats, of planar aggregates. *J. Microsc.*, **113**, 257-267.

Miles, R.E. (1987) Preface to *ISS commemorative-memorial volume* (ed. Miles, R.E.), *Acta Stereologica*, **6/suppl.II**, 5-10.

Miles, R.E. & Davy, P. (1976) Precise and general conditions for the validity of a comprehensive set of stereological fundamental formulae. *J. Microsc.*, **107**, 211-226.

Miles, R.E. & Davy, P. (1977) On the choice of quadrats in stereology. *J. Microsc.*, **110**, 27-44.

Ramalhoto, M.F. (1984) Bounds for the variance of the busy period of the M/G/∞ queue. *Adv. Appl. Prob.*, **16**, 929-932.

Ripley, B.D. (1988) *Statistical Inference for Spatial Processes.* Cambridge University Press, Cambridge.

Royall, R.M. (1976) Current advances in sampling theory: implications for human observational studies. *Am. J. Epidemiol.*, **104**, 463-474.

Russo, D. (1984) Design of an optimal sampling network for estimating the variogram. *Soil Sci. Soc. Am. J.*, **48**, 708-716.

Santaló, L. (1976) *Integral Geometry and Geometric Probability.* Addison-Wesley, Reading.

Serra, J. (1982) *Image Analysis and Mathematical Morphology.* Academic Press, London.

Shanbhag, D.N. (1966) On infinite server queues with batch arrivals. *J. Appl. Prob.*, **3**, 274-279.

Sterio, D.C. (1984) The unbiased estimation of number and sizes of arbitrary particles using the disector. *J. Microsc.*, **134**, 127-136.

Stoyan, D. (1979) On the accuracy of lineal analysis. *Biom. J.*, **21**, 439-449.

Stoyan, D., Kendall, W.S. & Mecke, J. (1987) *Stochastic Geometry and Its Applications*. Wiley, Chichester.

Weibel, E.R. (1970) An automatic sampling stage Microscope for stereology. *J. Microsc.*, **91**, 1-18.

Warrick, A.W. & Myers, D.E. (1987) Optimization of sampling locations for variogram calculations. *Water Resources Research*, **23**, 496-500.

Wicksell, S.D. (1925) The corpuscle problem I. *Biometrika*, **17**, 84-89.

Wicksell, S.D. (1926) The corpuscle problem II. *Biometrika*, **18**, 152-172.