MONTE CARLO METHODS IN FOREST INVENTORY

BY

SHEN-THEN CHANG

Bachelor of Science in Agriculture
National Taiwan University
Taipei, Taiwan, Republic of China
1974

Master of Science
University of Idaho
Moscow, Idaho
1979

Doctor of Philosophy
Oklahoma State University
Stillwater, Oklahoma
1984

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
Master of Science
July, 1987

MONTE CARLO METHODS IN

FOREST INVENTORY

Thesis Approved:

J P Chandler
Thesis Adviser

Thomas B. Lynch

R. E. Hedrick

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to express sincere gratitude to my thesis adviser, Dr. J. P. Chandler, for his suggestion, assistance, patience, and encouragement through this study. I also wish to express my deep appreciation to other committee members: Dr. G. E. Hedrick, and Dr. T. Lynch for critical reviewing and suggestion concerning manuscript of this thesis.

I wish to express my gratitude to the Department of Computing and Information Science of Oklahoma State University and Dr. D. D. Fisher for providing me teaching assistantship.

Thanks to Rick Matzen for the good time we shared and tremendous help he has offered to me when we were in the same office.

I wish to express my gratitude to my parents, Mr. and Mrs. Chan-Chi Chang, and my brother and sisters for their help in every aspect.

My wife, Agnes, and my son, Brian, are the ones who suffered the most during my study. Thanks for their encouragement and understanding.

TABLE OF CONTENTS

LIST OF TABLES

Table

page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The volume of a tree usually is estimated by measuring its cross-section area at intervals along the stem, then estimating in these intervals by making parametric assumptions about their shapes. Assumed shapes include cones, cylinders, paraboloids, and neiloids. The disadvantage of measuring volume in this way is that the accuracy of volume estimation depends on the appropriateness of the assumed shapes and the interval length that is chosen. Among many formulae that have been used, Smalian's formula is the most popular one due to its simplicity in application. However, tree volume estimated by this method always is biased positively unless the interval length chosen is very short (Husch et al, 1982, and Brickell, 1984).

Another way of measuring volumes of trees involves development of taper equations and/or volume equations through regression analysis, with a few easily measured parameters such as tree height and diameters as unknowns. The problems associated with this method are: some trees usually must be felled in order to obtain accurate measurments of parameters to develop taper equations, different tree species require different equations, and equations

constructed from data in one area may not be applicable in other areas.

With all these difficulties involved in tree volume estimation, it is natural to look for methods that are easy to use, general for a wide variety of tree species, and unbiased. With the aid of computer, Monte Carlo methods provide this alternative.

The Monte Carlo method can be defined as representing the solution of a problem as a paramter of a hypothetical population, and using a random sequence of numbers to construct a sample of this hypothetical population, from which statistical estimates of the parameter can be obtained (Halton, 1970).

An estimator t of a parameter T is an unbiased estimator of T if

$$E[t] = T.$$

That is, t is an unbiased estimator of T if "on the average" its value is equal to T (Hines and Montgomery, 1980). Not all Monte Carlo methods are unbiased, for example, weighted Monte Carlo integration results in a biased estimator of the parameter (Rubinstein, 1981). Many Monte Carlo methods, however, are unbiased. It is those unbiased Monte Carlo methods that are to be considered as alternatives to conventional tree volume estimation methods (Gregoire et al., 1985).

The Monte Carlo method comes in various forms, depending on the problems to be solved. Although similar

estimates may be reached by different Monte Carlo methods, the variance associated with the estimator obtained from different Monte Carlo methods may vary greatly. Usually one would like to obtain an estimate of the parameter with variance as small as possible so that the estimation is as as close to the real solution as one can get.

The purpose of this project is to investigate different Monte Carlo methods for tree volume estimation in forest inventory. Specifically, this project intends to:

1. illustrate the unbiasedness of various Monte Carlo methods in volume estimation;
2. find some variance reduction techniques that result in small variances for volume estimation;
3. find estimation methods that have absolute errors comparable to or less than that from use of Smalian's formula with the same number of measurements.

CHAPTER II

MONTE CARLO METHODS

The Monte Carlo method is a group of methods that approximately solve mathematical or physical problems by simulation using random quantities. Every Monte Carlo computation that leads to quantitative results may be regarded as estimating the value of a multiple integral. To simplify the matter, only a one-dimensional integral is used in the following discussions. It also is assumed that the bound of the $\int_a^b f(x)dx$ always follows the relation

$$0 <= a <= b.$$

Given a function f(x), one may find the value of the definite integral $I = \int_a^b f(x)dx$ in different ways. If $\int_a^b f(x)dx$ has an analytic solution, then one can find the solution by direct substitution. If it is very difficult to find the analytic solution of $\int_a^b f(x)dx$ or $\int_a^b f(x)dx$ has no analytic solution, one can use numerical methods to find the value.

However, if one can obtain only empirical values about f(x) at different x without knowing f(x), or if x is a vector, then to find $\int_a^b f(x)dx$ by conventional methods is very difficult, if not impossible. Monte Carlo methods can make these integration problems easier by converting the problem of integration to the problem of estimation of an

unknown parameter. In the following discussion, the unknown parameter is denoted as I. The various estimators of the unknown parameter are denoted as I1, I2,...

The Hit-or-Miss Monte Carlo Method

The simplest Monte Carlo method is so called the hit-or-miss Monte Carlo method (Rubinstein, 1981). The value of the one dimensional integral $I = \int_a^b f(x)dx$, assuming f(x) is bounded by 0 <= f(x) <= c, equals to the expected value of the estimator I1

$$E[I1] = c(b - a)(1/n) \sum_{i=1}^{n} g(r[2i-1], r[2i])$$

where r[2i-1] and r[2i] are random numbers distributed uniformly between (a, b) and (0, c), respectively, and g(r[2i-1], r[2i]) is one if f(r[2i-1]) >= r[2i], or zero if f(r[2i-1]) < r[2i]. The graphical representation of the hit-or-miss Monte Carlo is shown in Figure 1.

In other words, one takes n points at random in the area (a, b)(0, c), and counts the proportion of those points which lie below the curve y = f(x). This is sampling from the binomial distribution with the probability of success p = I/[c(b - a)]. The variance of the estimator I1 is

$$VAR(I1) = (I/n)(c(b - a) - I).$$

The hit-or-miss Monte Carlo method is easy to understand and easy to apply. However, it has the disadvantage that its estimation has the largest variance among all Monte Carlo methods.

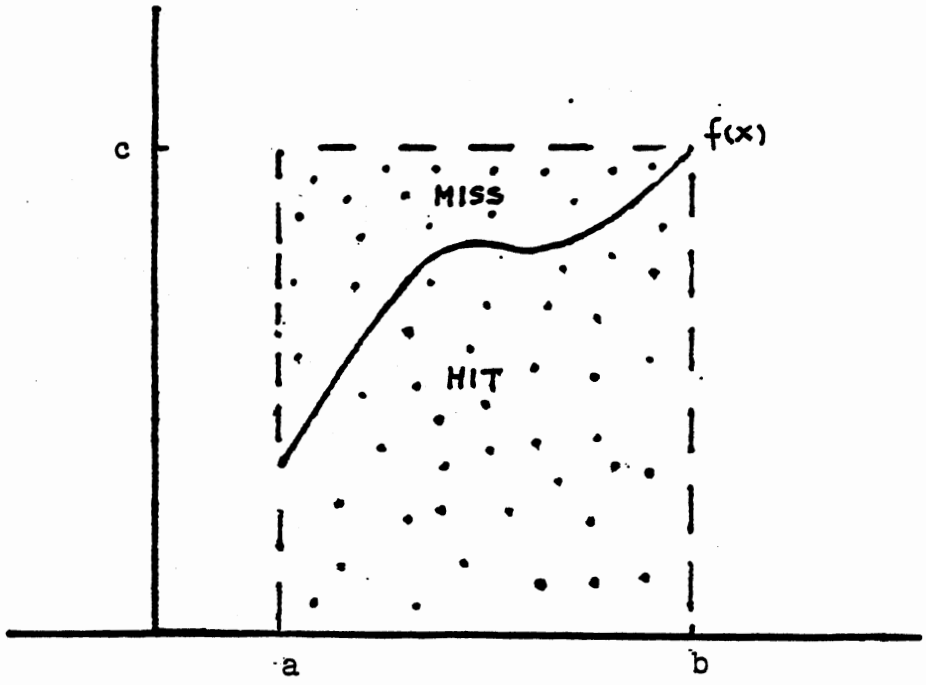

Figure 1. Graphical Representation of the Hit-or-miss Monte Carlo Method.

The Sample-Mean Monte Carlo Method

Another way of computing the integral $I = \int_a^b f(x)dx$ is to represent it as an expected value of some random variable (Hammersley and Handscomb, 1964, and Rubinstein, 1981). This Monte Carlo method is called the sample-mean or crude Monte Carlo method. By the mean-value theorem of calculus, computing the integral

$$I = \int_a^b f(x)dx$$

is equivalent to calculating

$$I = f(c)(b - a), \quad \text{for some } c,\ a <= c <= b.$$

The value of f(c) is

$$f(c) = E\left[(1/n)\sum_{i=1}^{n} f(c_i)\right]$$

where c is distributed uniformly between (a, b).

Therefore, the integral can be estimated as

$$I2 = (b - a)\left[(1/n)\sum_{i=1}^{n} f(c_i)\right].$$

Then,

$$I = E[I2] = (b - a)E\left[\left(\sum_{i=1}^{n} f(c_i)\right)/n\right]$$

$$= (b - a)(1/n)E\left[\sum_{i=1}^{n} f(c_i)\right], \quad a <= c_i <= b.$$

The variance of I2 is

$$VAR[I2] = (1/n)\left[(b - a)\int_a^b f(x)^2 dx - I^2\right].$$

This quantity is less than VAR[I1] since

$$VAR[I1] - VAR[I2] =$$

$$(I/n)[c(b - a) - I] - (1/n)\left[(b - a)\int_a^b f(x)^2 dx - I^2\right]$$

$$= (I/n)c(b - a) - ((b - a)/n)\int_a^b f(x)^2 dx$$

$$= ((b - a)/n)\left\{cI - \int_a^b f(x)^2 dx\right\}.$$

Since $c >= f(x)$ for all x between (a, b), therefore

$$cI - \int_a^b f(x)^2 dx \geq 0.$$

The comparison between the hit-or-miss and the sample-mean methods illustrates a general principle of Monte Carlo work: if one can replace an estimate by an exact value at any point of a Monte Carlo evaluation, the sampling error in the final result will be reduced.

Variance Reduction

In the application of Monte Carlo methods to many problems, it has been found that the sample size required to attain the desired level of accuracy is so large that if purely random sampling is used the cost of computation will be prohibitive. Fortunately, a few techniques for variance reduction are available. When those variance reduction techniques are applied properly the sample size requirement can be reduced without sacrificing the accuracy.

Variance reduction can be seen as a means to use known information about a problem. In fact, if nothing is known about the problem in hand, variance reduction cannot be achieved. Variance reduction cannot be obtained from nothing; it is merely a way of using all of the available information. The more that is known about the problem, the more effective are the variance reduction techniques that can be employed. The following discussion will concentrate in three variance reduction techniques, namely, correlated sampling or use of control variates, importance sampling, and use of antithetic variates.

Control Variates

In estimating the parameter $I = \int_a^b f(x)dx$, sometimes one can find a function $g(x)$ which approximates $f(x)$ and has an has an analytic solution. Then, since

$$I = \int_a^b f(x)dx$$
$$= \int_a^b [f(x) - g(x)]dx + \int_a^b g(x)dx$$

one can estimate I by

$$I = E[I3] = E[(1/n)(b - a) \sum_{i=1}^{n} \{f(x_i) - g(x_i)\} + G],$$

where $G = \int_a^b g(x)dx$, and $a <= x_i <= b$.

The value of the first term on the right-hand side of the equation can be obtained by the sample-mean Monte Carlo, and that of the second term by direct computation. This technique is known as correlated sampling or sampling with control variates (Hammersley and Handscomb, 1964, Kahn and Marshall, 1953).

The sample variance associated with I3 is

$$VAR[I3] = VAR[f(x) - g(x)] - VAR[G]$$
$$= VAR[f(x) - g(x)]$$

since G has variance of 0. If $g(x)$ approximates $f(x)$ such a way that $|f(x) - g(x)|$ varies less than $|f(x)|$, then the variance of I3 also will be much smaller than the variance of the estimator calculated without the control function $g(x)$. For example,

$$F = \int_0^1 \exp(-2x)dx$$

calculated with the sample-mean Monte Carlo with ten random numbers has mean of 0.4323 and of variance 0.0601, as shown

in Table I. The result from use of control variates with

$$g(x) = 1 - x$$

as control function has mean of 0.4226 and variance of 0.0055. The variance is reduced about 11 fold. Graphically (Figure 2), the variance from the sample-mean Monte Carlo estimation of F arises from the variation of f(x) as x runs over $0 <= x <= 1$. When the control function is introduced, the resulting function $[f(x) - g(x)]$ varies much less when x is in the range (0, 1), consequently, the smaller variance for $[f(x) - g(x)]$.

In selecting the control function, one must be sure that the control function chosen is simple enough to be integrated analytically. On the other hand, the control function must mimic f(x) and absorb most of its variation. In this example, g(x) absorbed much of the variation of f(x), and g(x) itself is a simple function to be integrated analytically.

There are various ways of looking at this method. For example, when estimating an unknown parameter s by means of an estimator s1, one may look for another estimator s2 which has a strong positive correlation with s1 and whose expectation is a numerically known quantity e. One then samples s1 and s2 simultaneously, using the same random variables, and use

$$s1 - s2 + e$$

as the estimator of s. Thus

$$s1 = (b - a)(1/n)(\sum_{i=1}^{n} f(x_i)),$$

TABLE I

EVALUATION OF THE FUNCTION f = exp(-2x) BY THE SAMPLE-MEAN MONTE CARLO METHOD AND BY THE CONTROL VARIATES METHOD WITH g(x) = (1 - x) AS CONTROL FUNCTION

x	f(x)	g(x)	f(x) - g(x)
0.073	0.8642	0.9270	-0.0628
0.411	0.4396	0.5890	-0.1492
0.826	0.1916	0.1740	0.0176
0.669	0.2624	0.3310	-0.0686
0.438	0.4164	0.5620	-0.1456
0.212	0.6544	0.7880	-0.1336
0.357	0.4897	0.6430	-0.1533
0.892	0.1680	0.1080	0.0600
0.734	0.2304	0.2660	-0.0356
0.138	0.7588	0.8620	-0.1032

True value of $\int_0^1 f(x)dx = 0.4323$

Estimation of $\int_0^1 f(x)dx$ by the sample-mean Monte Carlo method:

Mean = 0.4476
Variance = 0.0601

Estimation of $\int_0^1 f(x)dx$ with use of control variates:

Mean = 0.4246
Variance = 0.0055

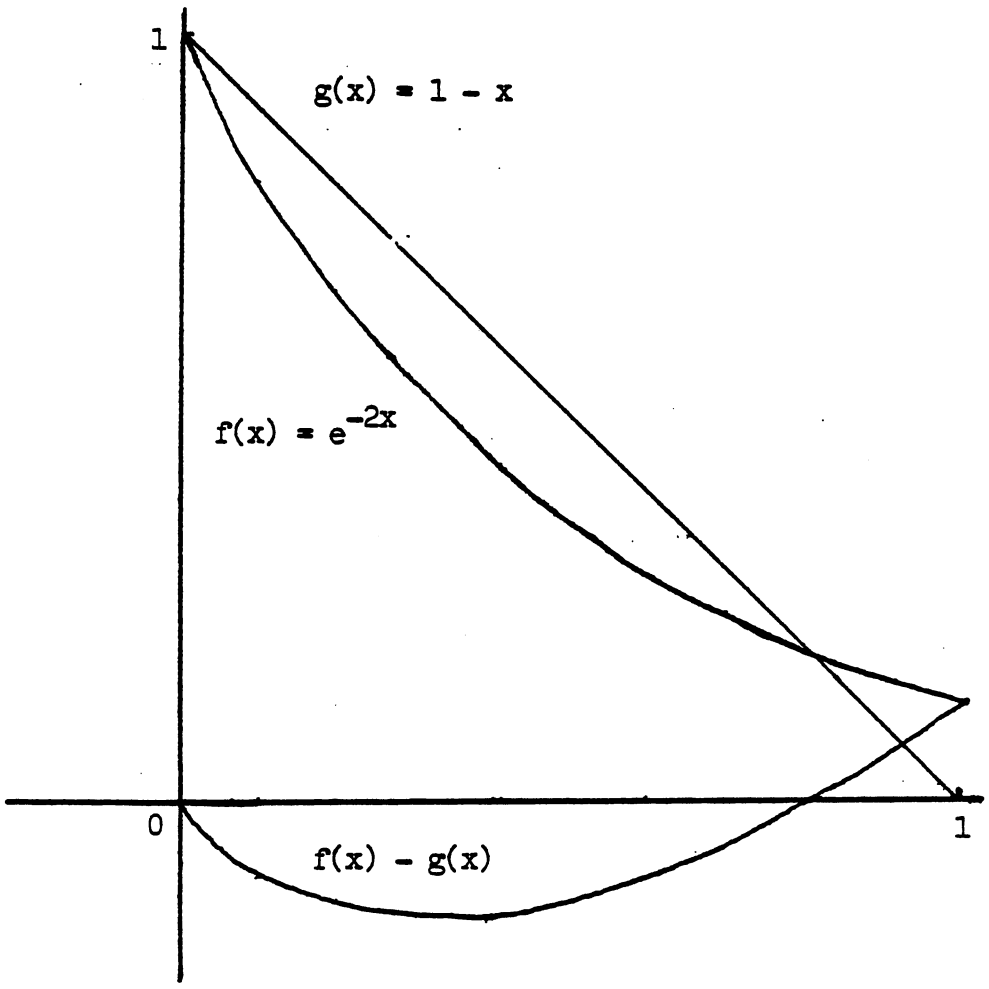

Figure 2. Graphical Representation of Monte Carlo Method with Control Variates.

$$s2 = (b - a)(1/n)\left(\sum_{i=1}^{n} g(x_i)\right), \quad a <= x_i <= b,$$

and

$$s = E[s1 - s2 + e]$$
$$= (b - a)(1/n)\sum_{i=1}^{n} E[f(x_i) - g(x_i)] + e.$$

Then

$$VAR[s] = VAR[s1 - s2 + e]$$
$$= VAR[s1] + VAR[s2] - 2COV[s1, s2].$$

If the positive correlation between f(x) and g(x) is strong enough to make 2COV[s1, s2] term greater than VAR[s2], then variance reduction is achieved.

Importance Sampling

Alternatively, if one chooses

$$G = \int_a^b g(x)dx = 1,$$

then

$$I = \int_a^b f(x)dx = \int_a^b [f(x)/g(x)](g(x)dx)$$
$$= \int_a^b [f(x)/g(x)]dG.$$

If g(x) is positive and does not change sign in (a, b), then G is a cumulative distribution function (cdf) between a and b, and g(x) is a probability density function (pdf). By the weighted-mean value theorem

$$I = [f(c)/g(c)]\int_a^b dG,$$

for some c in (a, b).

In order to estimate the integral, a sample of $c_1, \ldots, c_n$ is taken from the pdf g(x).

The estimate of I, I4, then can be obtained by

$$I4 = (1/n)\sum_{i=1}^{n}[f(c_i)/g(c_i)]G$$
$$= (1/n)\sum_{i=1}^{n}[f(c_i)/g(c_i)] \qquad a <= c_i <= b$$

since G = 1. This variance reduction technique is called importance sampling.

The sample variance of I4 can be found as

$$VAR[I4] = (1/n)\int_a^b [f(x)^2/g(x)]dx - I^2 .$$

The main idea of importance sampling is to concentrate the distribution of the sample points in the parts of the interval that are of most importance instead of spreading the sample uniformly. As far as the result is not biased, one compensates for distorting the distribution by taking f(x)/g(x) in place of f as estimator.

As an example, suppose one wishes to approximately compute the intcgral

$$F = \int_0^{\pi/2} \sin(x)dx,$$

The exact value of this integral is 1. The value of f(x) varies from 0 to 1 when x run from 0 to $\pi/2$. When estimated with the sample-mean Monte Carlo method, the estimate has mean of 1.0102 and variance of 0.2633 (Table II).

If $g(x) = 8x/\pi^2$ is used as importance function, the cumulative probability function G(x) is

$$G(x) = u = \int_0^x g(x)dx = \int_0^x (8x/\pi^2)dx = 4x^2/\pi^2 .$$

By using inverse transform on a random number u to find the value of x_i, one can calculate $f(x_i)/g(x_i)$. Estimate from with importance sampling has mean of 1.0045 and variance of 0.0227 (Table II). The resulting function f(x)/g(x) has very small variation for x running between 0 and 1. Much of

TABLE II

EVALUATION OF THE FUNCTION $f(x) = \sin(x)$ BY THE SAMPLE-MEAN MONTE CARLO METHOD AND BY IMPORTANCE SAMPLING WITH $g(x) = 8x/\pi^2$ AS IMPORTANCE FUNCTION

Random Varible u 1/	$f(u\pi/2)$	Random Variate x 2/	f(x)	g(x)	f(x)/g(x)
0.791	0.9466	1.3970	0.9849	1.1324	0.8697
0.338	0.5063	0.9132	0.7915	0.7402	1.0693
0.402	0.5903	0.9959	0.8393	0.8072	1.0398
0.211	0.3402	0.7215	0.6605	0.5848	1.1294
0.596	0.9362	1.2127	0.9366	0.9828	0.9530
0.643	0.8468	1.2596	0.9520	1.0210	0.9324
0.093	0.1456	0.4790	0.4609	0.3883	1.1870
0.137	0.2135	0.5814	0.5492	0.4713	1.1653
0.960	0.9980	1.5391	0.9995	1.2475	0.8012
0.724	0.9075	1.3366	0.9727	1.0834	0.8978

True value of $\int_0^{\pi/2} f(x)dx = 1$

Estimation of $f(x)dx$ by the sample-mean Monte Carlo method:

Mean = 1.0102
Variance = 0.2633

Estimation of $\int_0^{\pi/2} f(x)dx$ by importance sampling:

Mean = 1.0045
Variance = 0.0227

1/ Random variable u distributes uniformly between (0, 1).

2/ x is obtained by inverse transform from

$$u = \int_0^x 8t/\pi^2 \, dt.$$

the varation in f(x) is absorbed by the importance function g(x) (Figure 3).

Suppose that one can find a function g(x) such that f(x)/g(x) = c, where c is a constant. Then c equals 1/I. Applying this to the variance equation yields VAR(I4) = 0. It appears that one has a perfect Monte Carlo method, giving the exact answer every time. This is unfortunately useless, since to sample f(x)/g(x) one must know g, and to determine g(x) (= f(x)/I) one must know I, and if one knows I there is no need for Monte Carlo methods to estimate it.

Although one cannot find a function to make f(x)/g(x) be a constant, one might be able to find some function g(x) such that [f(x)/g(x)] is as close to a constant as one can make it, then the variance can be drastically reduced.

One may ask that, given a function g(x) approximating the function f(x), which method one should use to estimate $\int_a^b f(x)dx$. It has been shown (Halton, 1965) that

$$\mathrm{VAR}[I3] - \mathrm{VAR}[I4] = n\,\mathrm{COV}[(f(x) - g(x))^2/g(x),\ g(x)].$$

If the approximation of f(x) by g(x) is absolutely uniform, i.e. |f(x) - g(x)| is approximately constant, correlated sampling (sampling with control variates) is more efficient; while if the approximation is relatively uniform, i.e. |f(x) - g(x)| is approximately proportional to |g(x)|, then importance sampling is preferable.

Monte Carlo computation with importance sampling is especially relevant for integrands that "peak", so that f(x)

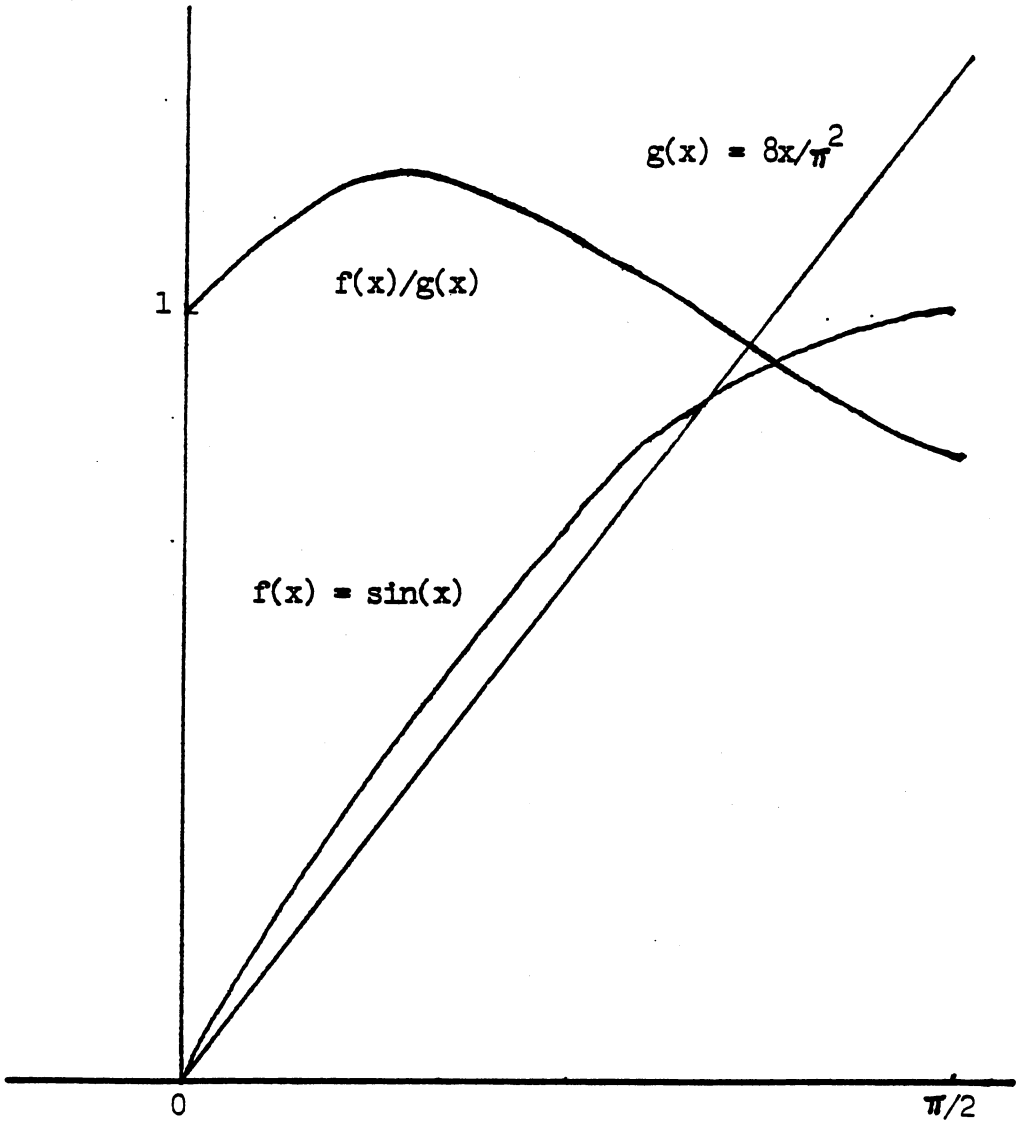

Figure 3. Graphical Representation of Monte Carlo Method with Importance Sampling.

has extreme high values at certain x's (Figure 4). For such a function, one sample in a peak swamps several samples elsewhere. If one can find a g(x) that approximates the shape of f(x) and makes f(x)/g(x) unpeaked, then importance sampling can reduce variance and improve accuracy. On the other hand, if f(x) "dips", importance sampling would not help much in variance reduction (Figure 5). Similarly, if a function g(x) is chosen as importance function such that f(x)/g(x) is a peaked function, one should consider either looking for another importance function, or combining this importance sampling with some other variance reduction techniques. If the resulting f(x)/g(x) is a function that dips, this importance function should be considered as acceptable.

One may argue that it is possible to pick up two functions that have the same variance but one dips and another one peaks. While this is true, it is not "fair" to compare their variance only, since their means may differ considerably. Therefore, one should compare variation in relative terms, and coefficient of variation is a good measure of relative variation. Then, the function that dips usually would have smaller coefficient of variation than the function that peaks if indeed their means are statistically different.

Antithetic Variates

The principle of using antithetic variates in Monte

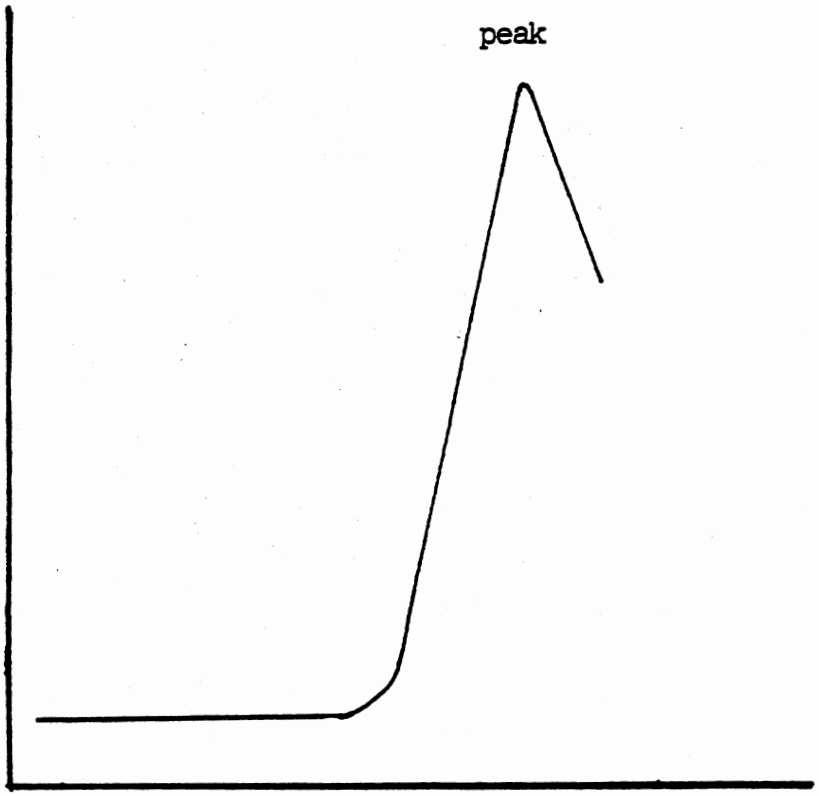

Figure 4. A Function That Peaks.

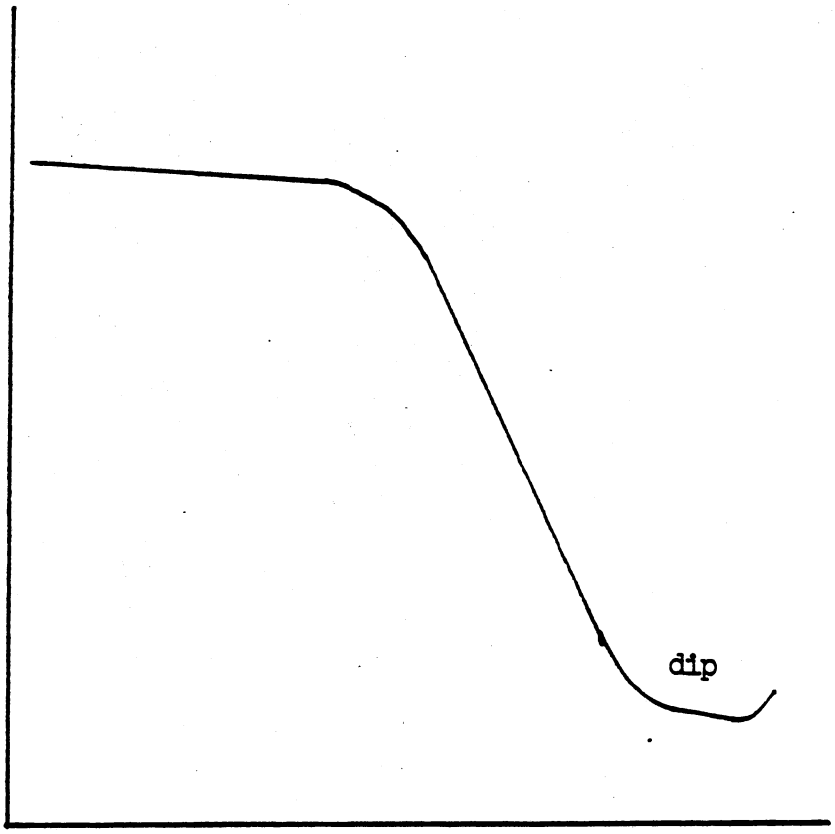

Figure 5. A Function That Dips.

Carlo methods is that one seeks two unbiased estimators t1 and t2 with strong negative correlation, for some unknown parameter I. Then

$$I5 = (t1 + t2)/2$$

also is an unbiased estimator of I. The sampling variance of I5 then is

$$VAR[I5] = VAR[(t1 + t2)/2]$$
$$= (VAR[t1] + VAR[t2])/4 + COV[t1, t2]/2.$$

If the covariance is strongly negative, the method of antithetic variates can be very effective in variance reduction (Hammersley and Morton, 1955). In fact, if f(x) is a continuous monotonic function with continuous first derivative, the variance of the estimator from use of antithetic variates is guaranteed to be less than half of the magnitude of that of estimator from the sample-mean Monte Carlo method (Rubinstein, 1981).

Considering the integral

$$I = \int_a^b f(x)dx,$$

which equals to

$$I = (1/2)\int_a^b [f(x) + f(a+b-x)]dx.$$

To estimate I, one can taken a sample of size n from the uniform distribution between (a, b) and find

$$I5 = (b-a)(1/2)(1/n)\sum_{i=1}^{n}[f(x_i) + f(a+b-x_i)],$$

where $a <= x_i <= b$.

The reasoning for using antithetic variates is similar to that for using control variates and importance sampling: achieving variance reduction by combination of functions.

In using antithetic variates one derives another function g(x) such that g(x) is symmetric to f(x) along the line x = a + (b - a)/2. If f(x) is monotonic or tends to be monotonic, then [f(x)+g(x)]/2 usually will have smaller variation then f(x) due to compensation. For example,

$$f(x) = (x^3 + x^2 + 1)/(x^2 + 1)$$

is almost monotonically increasing between (0, 5) except in the range (0, 1) (Figure 6). By generating

$$g(x) = f(5 - x)$$

$$= (101 - 65x + 14x^2 - x^3)/(x^2 - 10x + 26)$$

and then combining f(x) and g(x), the resulting function [f(x)+g(x)]/2, has much smaller variation in (0, 5) than does (Figure 6). Using the sample-mean Monte Carlo method on [f(x)+g(x)]/2, the variance will be drastically reduced comparing to that of the sample-mean Monte Carlo method on f(x) alone.

Correlated sampling and importance sampling depend on knowing an "easy" function g(x) which approximates the "difficult" function f(x). However, this easy function usually is difficult to find, and one must be content with some compromise. When the conditions are fairly loose it is easier to reach a good compromise. This is the case with antithetic variates. In practice, it is relatively easy to find negatively correlated unbiased estimator of a parameter usually easier than it is to find an equally satisfactory control function or importance function. The antithetic variates method therefore tends to be more efficient in

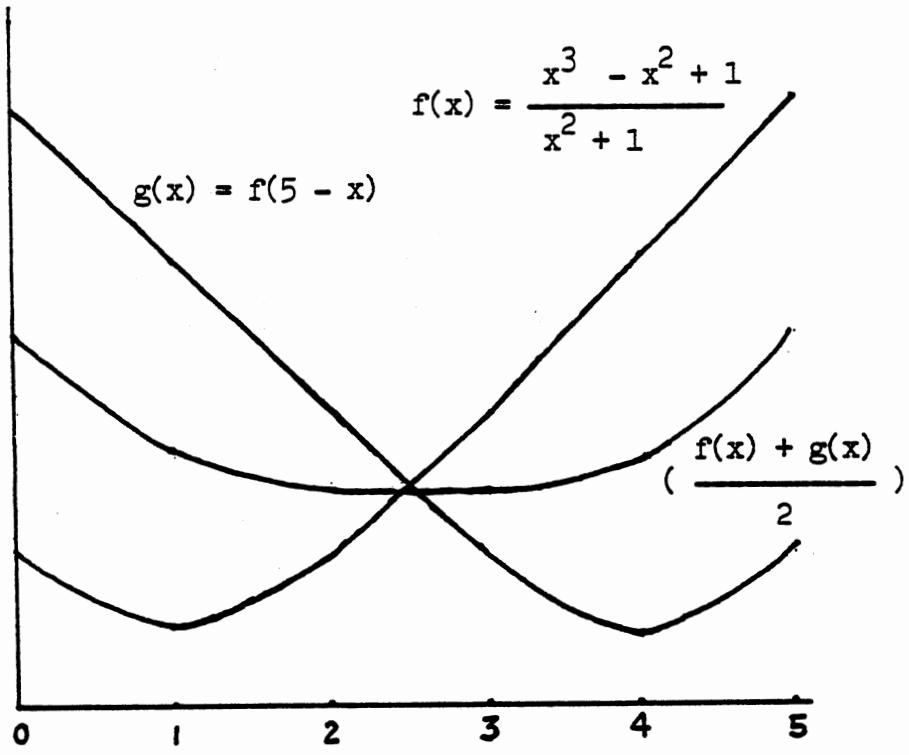

Figure 6. Graphical Representation of Monte Carlo Method with Antithetic Variates.

practice, or, at least easier to apply.

CHAPTER III

TREE VOLUME ESTIMATION

Assume that the relationship between tree height and diameter can be expressed as some function

$$y = f(x),$$

where x is height along the tree,

y is the diameter at height x.

Then the volume of a tree can be found by integrating the square of the function f(x):

$$V(H) = \pi k \int_0^H [f(x)]^2 dx, \qquad (1)$$

where V(H) is total volume of the tree,

H is the height of the tree, and

k is the scale factor.

This method of tree volume calculation is called the cross-section method (Figure 7).

Alternatively, tree height x can be expressed as a dependent variable of the diameter y

$$x = g(y),$$

and tree volume can be calculated by the shell method

$$V(B) = 2 \pi k \int_0^B yg(y)dy,$$

where B is the diameter at the base of a tree (Figure 8).

Unfortunately, the function f(x) or g(y) usually is not known. Therefore, the volume of the tree cannot be obtained

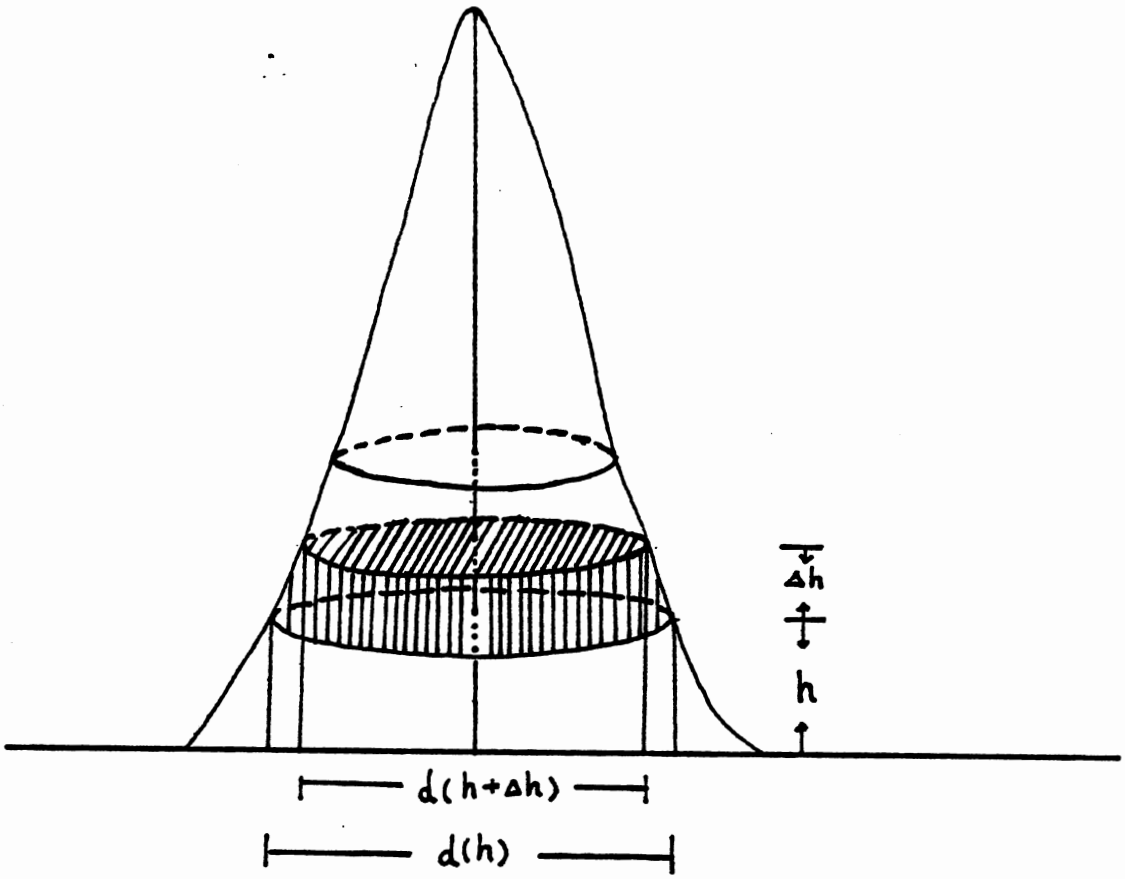

Figure 7. The Cross-section of a Solid.

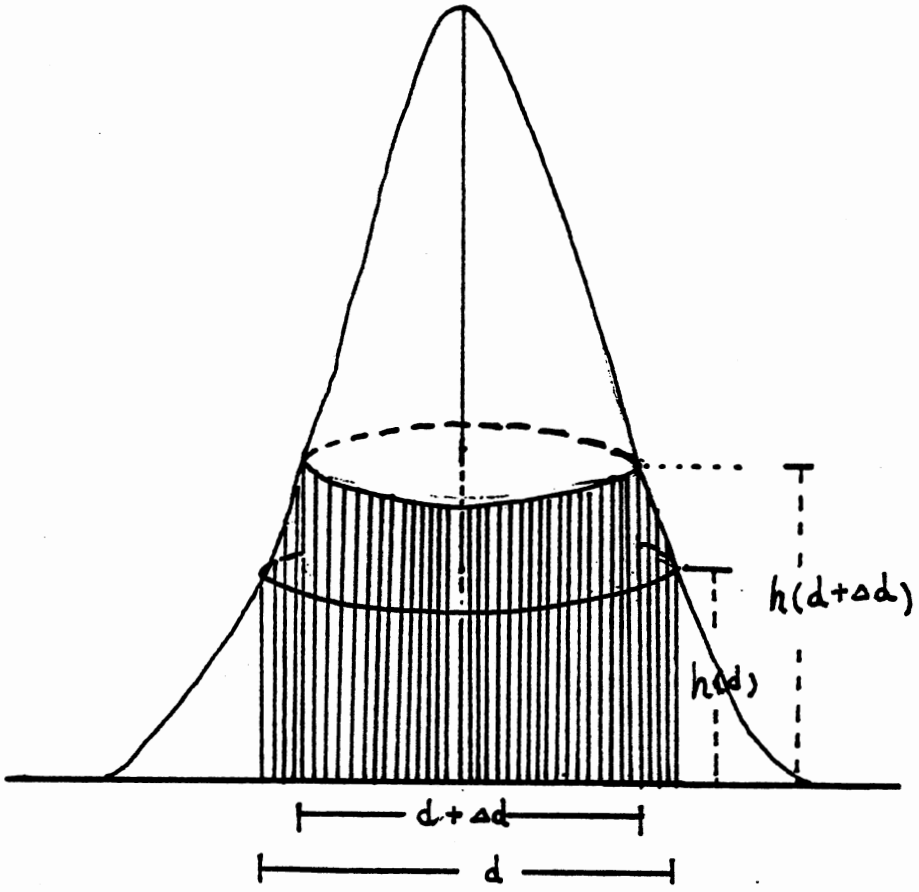

Figure 8. The Shell of a Solid.

from direct integration.

Methods that are used most often to obtain an estimate of tree volume are the integration method based on the trapezoidal rule (Smalian's formula) and construction of taper and/or volume equations based on a few easily measured parameters.

Smalian's formula essentially is a numerical method of integrating a function with the trapezoidal rule. In estimating the volume of a tree, Smalian's formula works as follows. Starting from the base of the tree, the stem is divided into intervals of equal length. The cross-section areas of both ends of each interval is estimated by taking measurements of the diameters, assuming the stem cross-section is always a circle. These cross-section areas are used to estimate the volume of the stem in this interval. Accumulation of the volume of all intervals results in total tree volume.

Alternatively, one can estimate the volume V of a tree by the Monte Carlo method since the numerical value of the height x, or g(y), and the diameter f(x), or y, can be obtained by direct measurement. The volume V then can be estimated without bias by

$$V(H)=\pi k(H/n)\sum_{i=1}^{n} f(x_i)^2, \qquad 0 <= x_i <= H,$$

with the cross-section method, or by

$$V(B)=2\pi k(B/n)\sum_{i=1}^{n} y_i\, g(y_i), \qquad 0 <= y_i <= B,$$

with the shell method.

Advantages of Monte Carlo Methods in Volume Estimation

The Monte Carlo method should be a good choice for tree volume estimation for several reasons. Although there is some definite relationship between diameter and height of a tree, it is very difficult to express this relation in an exact mathematical formula. Without an exact formula, it is impossible, or at least very difficult, to calculate the volume analytically. Even if a formula is established, it tends to be complex if it models the tree taper closely, and this can make computation difficult. With the Monte Carlo method, one always can estimate the tree volume without an exact mathematical formula.

Unbiased Monte Carlo methods such as use of antithetic variates, use of control variates, importance sampling, etc..., can be applied easily. Conventional methods of tree volume measurement are biased. The widely used Smalian's formula is known to be positively biased. The only way to reduce this bias is by increasing the sample size (Brickell, 1984).

Using taper or volume equations to measure the volume of trees also has difficulties. Taper and volume equations developed through regression are often biased when applied to local subpopulations or to populations other than the one on which the sample is based. Worse yet, this bias cannot be reduced by increasing the sample size.

Volume estimated with the unbiased Monte Carlo method on a single tree is not biased. When volume estimation by Monte Carlo methods applies to a large population, the unbiased estimation from the Monte Carlo method leads to cancellation of errors incurred on each tree and gives a total error of estimation close to zero.

In using the Monte Carlo method, if somehow one can reduce the variance of estimation on each tree, then in a large forest the estimation of the total volume should be the true total volume with very small variation. With improved sampling techniques, one should be able to achieve high accuracy in volume estimation with a reduced number of samples, thus saving labor cost in measuring tree volume in a forest.

The Power Function Model of Solids of Revolution

If the formula

$$V(H) = \pi K (H/n) \sum_{i=1}^{n} f(x_i)^2, \qquad 0 <= x_i <= H,$$

or

$$V(B) = 2\pi K (B/n) \sum_{i=1}^{n} y_i \, g(y_i), \qquad 0 <= y <= B,$$

is used to estimate the volume of a tree, the variance associated with the estimator will be quite large. To reduce the variance, one needs some information about the tree to be estimated. Usually, trees with excurrent form (single stem) can be approximated by a power function model of solids of revolution. The power function model of solids

of revolution has the form

$$s(x) = D(1 - x/H)^p \qquad (2)$$

where s(x) is the diameter at height x,

H is height of a tree,

x <= H,

D is diameter at breast height (dbh).

The power p in (2) determines the form (or shape) of a tree. It represents a cylinder if p = 0, a paraboloid if p = 0.5, a cone if p = 1, and a neiloid if p = 1.5.

Real trees do not assume any of these simple forms, but instead, a combination of them. Generally it is assumed that the bottom part is approximately a neiloid, the middle section approximately a paraboloid, and the top part approximately a cone.

For trees of excurrent form, they can be approximated quite well by power function model of solids of revolution. There are taper equations established through regression bearing similarity to the power function model. For example equations by Ormerod (1963), and volume ratio models by Burkhart (1975) and by Cao (1977).

The close approximation of the power function model of solids of revolution to trees with excurrent form suggests ways to estimate true tree volume with Monte Carlo methods along with variance reduction techniques.

Assuming the true taper equation of a tree is f(x), then V(H) in (1) is the volume of the tree. To achieve variance reduction, one can use the power function model of

solids of revolution either as control variates or as importance functions for importance sampling, depending on degree of accuracy required. Since most excurrent trees have forms between paraboloid and cone, one can approximate the real tree taper with power function model by properly choosing the value of p in formula (2).

The Cross-section Method

Using importance sampling in tree volume estimation is first proposed by Gregoire et al (1985). In their original work a taper equation similar to that given by Ormerod is used as the importance function. Van Deusen and Lynch (1986) show that using power function model of solids of revolution as an importance function with p fixed to 0.5, trees with various shapes all can be estimated without bias, although to obtain the same degree of accuracy requires more samples for those trees with shapes quite different from the shape of the importance function.

Apparently, S(x)/S(H) from the power function model is a pdf, where

$$S(x) = \pi k \int_0^x [D(1 - t/H)^p]^2 dt.$$

From (1),

$$V(H) = \pi k \int_0^H f(x)^2 dx$$

$$= \pi k \int_0^H [f(x)^2 / g(x)] dG(x),$$

where

$G(x) = S(x)/S(H)$, and

$g(x)dx = dG(x)$.

One can estimate the true volume V(H) by V

$$V = (\pi k/n) \sum_{i=1}^{n} f(x_i)/g(x_i),$$

where x is a random variable sampled from pdf g(x). Each x can be generated as follows:

1. generate a random variable u_i, distributed in (0, 1) uniformly.
2. use the inverse transform to find x_i from

$$u_i = \int_0^x g(t)dt.$$

To estimate the volume of a real tree, $f(x_i)$ is taken directly from the tree.

It will now be proved that any function with the form

$$y = K(H - x)^b \tag{3}$$

chosen as importance function shall be the same as choosing the solids of revolution (2) with b = p.

It is obvious that the pdf from (2) for the cdf

$$G(x) = S(x)/S(H)$$

is $dG(x)/dx = d[S(x)/S(H)]/dx$

$$= [dS(x)/dx]/[S(H)]$$

$$= [(\pi D^2/H^{2P})(H - x)^{2P} dx]/[(\pi D^2/H^{2P})\int_0^H (H - x)^{2P} dx]$$

$$= [(H - x)^{2P} dx]/[\int_0^H (H - x)^{2P} dx]. \tag{4}$$

The pdf q(x) from (3) for the cdf

$$Q(x) = Y(x)/Y(H)$$

is $dQ(x)/dx = d[Y(x)/Y(H)]/dx$,

where

$$Y(x) = \pi\int_0^x [K(H - t)^{2b}]\, dt,$$

and

$$Y(H) = \pi\int_0^H [K(H - t)^{2b}]\, dt.$$

Therefore,

$$q(x) = d[Y(x)/Y(H)]/dx$$

$$= [dY(x)/dx]/[Y(H)]$$

$$= [(\pi K^2)(H - x)^{2b}dx]/[(\pi K^2)\int_o^H (H - x)^{2b}dx]$$

$$= [(H - x)^{2b}dx]/[\int_o^H (H - x)^{2b}dx],$$

which is the same as (4) derived from solids of revolution if b = p.

The equivalence of (2) and (3) means that if a taper equation can be converted to the solids of revolution (2) with a difference only in the constant term, then using this taper equation as an importance function is the same as using the solids of revolution (2). It also means that unless trees to be measured have shapes quite different from power function model of solids of revolution, using (2) as importance function can improve the accuracy of tree volume estimation considerably without knowing "too much" about the tree, i.e. the taper function of the tree. Thus in their work, Gregoire et al (1984) can just choose (2) as their importance function. Using Monte Carlo method with the power function model (2) as an importance function one can do without taper equation completely for any tree species in the case of tree volume estimation.

Since solids of revolution generally approximate true tree taper of excurrent trees quite closely, it also should be able to be used as control function in estimating tree volume. To estimate tree volume with solids of revolution as control function, one calculate $f(x_i)$ and $s(x_i)$ at the

same time by generating a random variable x in (0, H). The volume V(H) then can by estimated without bias by

$$V = (H/n)\sum_{i=1}^{n}[f(x_i) - s(x_i)] + \int_0^H s(x)dx.$$

Since most trees of economic importance already have taper equations close to true tree shape developed, one may, instead of using hypothetic taper equations such as power function model of solids of revolution, use those eqations as control functions to get a better estimate of tree volume.

The Shell Method

The relationship between diameter and height also can be expressed as x = g(y), where y is diameter, and g(y) is height for the point with diameter y. The volume of a tree can then be calculated by shell method as

$$V(B) = 2\pi k\int_0^B yg(y)dy.$$

Since g(y) usually is not known, one cannot calculate the volume of a tree by direct integration. By using the sample-mean Monte Carlo, the volume can be estimated as

$$V = (2\pi Bk/n)\sum_{i=1}^{n} y_i\, g(y_i), \qquad 0 <= y_i <= B.$$

As stated before, this estimator has a large variance associated with. By using antithetic variates, one can reduce the variance. But one can do more by finding an "easy function" and using it either as importance function or as control function in variance reduction. The natural choice again is the power function model of solids of revolution.

An equation that expresses the height as dependent variable of diameter for the power function model of solids of revolution is

$$h(y) = H(1 - (y/B)^q), \quad q = 1/p,$$

which is just the inverse of (1). The volume of a tree up to diameter y is

$$\begin{aligned} V(y) &= 2\pi k\int_0^y H(1 - (t/B)^q)tdt \\ &= 2\pi kH\int_0^y (1 - (t/B)^q)tdt \\ &= 2\pi kHy^2[1/2 - (y^q/((q+2)B^q)]. \end{aligned}$$

And the volume of the tree is

$$V(B) = 2\pi kHB^2[1/2 - 1/(q+2)].$$

The ratio

$$V(y)/V(B) = (y/B)^2[((q+2)B^q - 2y^q)/(qB^q)]$$

is a cdf. And

$$\begin{aligned} r(y) &= (2\pi KH(1 - (y/B)^q)y)/(2\pi kHB^2(1/2 - 1/(q+2))) \\ &= (1 - (y/B)^q)y/(B(1/2 - 1/(q+2))) \qquad (4) \end{aligned}$$

is the pdf associated with V(y)/V(B).

One can generate random variates y_i from $V(y_i)/V(B)$ by generating a random variable u_i uniformly distributed in (0, B), then using the inverse transform to find y_i. This y_i is used to obtain values $g(y_i)$ and $r(y_i)$. One then takes their ratio. Therefore, the estimate of a tree's volume by the shell method Monte Carlo calculation with importance sampling is

$$V = (2\pi k/n)\sum_{i=1}^{n} g(y_i)y_i/r(y_i),$$

when formula (4) is used as importance function.

One can choose the best form as importance function

by properly choosing the value of q. This however may not be economical, since to do so complicates the process of the inverse transform due to the difficulty of finding the value y . If the shape of the power function model of solids of revolution is set to be a parabola, which is a good enough choice for practical purposes, (Gregoire et al, 1986), the entire process can be simplified.

When r(y) is set to be a parabola, q has the value of two. Substitute it into V(y)/V(B), one gets

$$V(y)/V(B) = -(y^4/B^4 - 2y^2/B^2).$$

The diameter y then can be easily found as

$$y = B\sqrt{1 - \sqrt{(1 - u_i)}} ,$$

where u_i is a random variable uniformly distributed between 0 and 1.

The pdf then can be expressed as

$$r(y) = 4y(1 - y^2/B^2)/B ,$$

which is also a simple function once y is generated from the cdf by inverse transform.

Critical Height Sampling

Another tree volume estimation method is called critical height sampling, it was first proposed by Kitamura (1964). The principle of critical height sampling is as follows.

Each tree has an associated distribution with cdf y^2/B^2. Here y is the diameter at some height on the tree. B is the diamter at the base of the tree. One can generate

a diameter y_i randomly by $y_i = B\sqrt{u_i}$, and using this y_i to find the corresponding height $g(y_i)$; here u_i is a random number with a uniform distribution between 0 and 1. The expected value of g(y) is

$$E[g(y)] = \int_0^B g(y)(2y/B^2)dy.$$

The volume of a tree calculated by the shell method is

$$V = 2\pi k\int_0^B yg(y)dy.$$

Taking the ratio of the previous formula and this one yields

$$E[x]/V = 1/(\pi kB^2).$$

Rearranging this equation, one gets

$$V = \pi kB^2 E[g(y)],$$
$$= (\pi kB^2/n)\sum_{i=1}^{n} g(y_i).$$

Therefore, one can estimate the volume of a tree by generating a number of critical heights and multiplying their average by the basal area of that tree.

The sample variance of V tends to be large since this is the sample-mean Monte Carlo of critical height sampling, therefore the less accurate of the estimator. Unlike the cross-section method and the shell method, there is no known "easy function" that can be applied to variance reduction. Therefore, use of antithetic variates is the only variance reduction technique applicable to critical height sampling.

There is an intimate relation between critical height method and the shell method. In fact, one can derive the formula for critical height sampling from the shell method. The volume of a tree calculated by the shell method is

$$V = 2\pi K \int_0^B y g(y) dy.$$

By the weighted mean-value theorem of integration, this equation can be written as

$$V = 2\pi K g(c) \int_0^B y dy, \qquad 0 <= c <= B.$$

The integral on the right-hand side of the above equation is $B^2/2$. Therefore, the above equation can be wriiten as

$$V = (\pi K B^2) g(c).$$

In terms of expectation, g(c) is E[g(y)], where y is in (0, B). Then the above equation can be written as

$$V = (\pi K B^2) E[g(y)],$$

which is the same as the critical height equation derived in the beginning of this section.

CHAPTER IV

DESCRIPTION OF THE SIMULATION PROGRAM

The main purpose of this project is to find a better Monte Carlo method to replace the widely used, but biased, Smalian's formula for tree volume estimation. The selection criteria are that the method(s) chosen should perform better than Smalian's formula in terms of absolute error when volume is estimated with the same number of samples taken from each tree, and the sample size should be small. This is because increasing sample points increases the accuracy of result from Smalian's formula faster than that of Monte Carlo methods. If a large number of samples are taken, then the result from Smalian's formula will be better. Therefore if the Monte Carlo method evaluated are to be considered as possible alternatives of Smalian's formula, they should produce smaller absolute errors than that from Smalian's formula for smaller number of samples per tree.

In this study, tree volume estimation of loblolly pine is simulated. The "true" tree is represented by the taper equation developed by Max and Burkhart (1976) for natural stand-outside bark. The equation has the form

$$d^2/D^2 = b1(h/H - 1) + b2(h^2/H^2 - 1) +$$

$$b3(a1 - h/H)^2 I1 + b4(a2 - h/H)^2 I2$$

where $I1 = 1$ if $h/H <= a1$,

$I2 = 1$ if $h/H <= a2$,

$b1 = -2.4602$, $b2 = 0.9751$,

$b3 = -0.7044$, $b4 = 131.4842$,

$a1 = 0.8026$, $a2 = 0.0867$.

The "true" volume of a tree is obtained by integrating this equation through the entire height of the tree.

All Monte Carlo computations in this study can be classified into three categories: 1) computations based on the cross-section method, 2) computations based on the shell method, and 3) critical height sampling.

The simulation study is conducted on trees of hypothetical sizes. Diameters at breast height ranged from 0.1524 m (6 in) to 0.4826 m (19 in). Tree heights range from 12.192 m (40 ft) to 30.480 m (100 ft). Input tree sizes are given in Table III.

Smalian's Formula

Tree volume estimated with Smalian's formula is used in comparison with results from Monte Carlo computations. The algorithm for estimating volume with Smalian's formula is as follows.

a. for n sample points on the stem, divide the stem height by n+1, the result is the interval length;
b. for each section of the stem, calculate its volume by finding the cross-section area of the section

TABLE III

SIZES OF TREES USED IN SIMULATION STUDY

DBH (METERS)	HEIGHT (METERS)						
	12.192	15.240	18.288	21.336	24.384	27.432	30.480
0.1524	1	1	1	1			
0.1778	1	1	1	1			
0.2032	1	1	1	1	1		
0.2286	1	1	1	1	1		
0.2540		1	1	1	1		
0.2794		1	1	1	1	1	
0.3048			1	1	1	1	
0.3302			1	1	1	1	
0.3556			1	1	1	1	
0.3810			1	1	1	1	
0.4064			1	1	1	1	
0.4318			1	1	1	1	1
0.4572			1	1	1	1	1
0.4826			1	1	1	1	1

top and the section bottom, add them together, dividing the result by 2, then multplied with the section length.

c. sum up volume of all sections to get volume estimation.

Monte Carlo Computation of the Cross-section Method

Under this category, importance sampling, importance sampling with antithetic variates, and use of control variates are evaluated. Crude Monte Carlo method and crude Monte Carlo with antithetic variates are include for comparison purpose. Solids of revolution of (2) are used either as importance function or as control function. Another calculation with taper equation from Kozak et al (1969) as control function is also programmed to compare if using established taper equations as control function can improve the estimation. This taper equation has the form

$$y^2/D^2 = b1(x/H - 1) + b2(x^2/H^2 - 1),$$

y : diameter at height x;

b1 : regression coefficient, -2.6447 is used here;

b2 : regression coefficient, 1.2453 is used here.

The regression coefficients b1 and b2 are from Cao et al (1980) for loblolly pine.

The Crude Monte Carlo and the Crude Monte Carlo With Antithetic Variates

The algorithm for the crude Monte Carlo method is as follows.

a. Generate a sequence of n random numbers u_i, each one is distributed uniformly in (0, 1).

b. use $x_i = H*u_i$ as a height to find out the cross-section area $A_i = \pi[f(x_i)]^2$ at this height.

c. compute $V = H*(1/n)\sum_{i=1}^{n} A_i$.

In the computation of crude Monte Carlo method with antithetic variates, one takes n/2 random samples instead of n random samples. For each random number u_i generated, one calculates another random number v_i by $v_i = 1 - u_i$. The algorithm for this estimation is as follows.

a. Generate n/2 random numbers u_i, each one is distributed uniformly in (0, 1).

b. Derive another n/2 random numbers v_i from u_i.

c. Calculate the cross-section area A_{u_i} and A_{v_i} from from $H*u_i$ and $H*v_i$, respectively.

d. Calculate the estimate by $(H/n)*(\sum_{i=1}^{n/2}[A_{u_i} + A_{v_i}])$.

Importance Sampling Without and With Antithetic Variates

Equation (2) for solids of revolution is used as an

importance function. The power in the equation changes automatically in the program so that one can select the appropriate form as importance function by studying the simulation results.

The algorithm for importance sampling is as follows.

a. Generate a sequence of n random numbers u_i.

b. Obtain the random variate x by taking inverse transform on

$$u_i = V(x)/V(H) = 1 - (1 - x_i/H)^{2P+1}.$$

c. Compute $(\pi/n)\sum_{i=1}^{n} f(x_i)^2/g(x_i)$.

In the computation of importance sampling with antithetic variates, instead of taking n random samples, one takes n/2 random samples. For each random number u_i, one generates another random number v_i by $v_i = 1 - u_i$. The estimate is calculated as

$$(\pi/n)[\sum_{i=1}^{n/2} f(x_i)/g(x_i) + f(y_i)/g(y_i)].$$

Here x_i and y_i are generated by inverse transform from u_i and v_i, respectively. The algorithm is as follows.

a. Generate n/2 random numbers u_i.

b. Derive n/2 random numbers v_i from u_i.

c. Derive x_i from u_i and y_i from v_i by inverse transform the same as that in importance sampling calculation.

d. Calculate the estimate by

$$(\pi/n)[\sum_{i=1}^{n/2} f(x_i)^2/g(x_i) + f(y_i)^2/g(y_i)].$$

Use of Control Variates

The algorithm for control variates sampling is as follows.

a. Generate n random numbers u_i between (0, H).

b. Calculate the estimate by

$$(1/n)\sum_{i=1}^{n}[f(u_i) - g(u_i)] + G$$

where G is the integral of either solids of revolution of equation (2) or the taper function by Kozak et al (1969) with coefficient from Cao et al (1980).

Monte Carlo Computation of the Shell Method

Under this category, only importance sampling and importance sampling with antithetic variates are programmed In these calculations, one estimates a tree's volume in the same way as volume estimation in the cross-section method except that the pdf now is

$$(1 -(y/B)^{q})y/[B\ (0.5 - 1/(q+2))],$$

and the cdf is

$$(y/B)^{2}[((q+2)B^{q} - 2y^{q})/(qB^{q})].$$

The procedure of finding estimate is the same as that under the cross-section category.

In the simulation study, only the importance function that assumes the shape of paraboloid is tested because assuming tree taper other than paraboloid adds a tremendous computation burden due to the difficulty of finding random

variates from the inverse transform.

Critical Height Sampling

Under this category two methods are studied: critical height sampling and critical height sampling with antithetic variates.

The algorithm for critical height sampling is as follows.

1. Generate n random variables u that are uniformly distributed in (0, 1).
2. Calculate $V = (B/n)\sum_{i=1}^{n} H*u$.

The algorithm for critical height sampling with antithetic variates is as follows.

1. Generate n/2 random numbers u that are uniformly distributed in (0, 1)
2. Derive another n/2 random numbers v_i in (0, 1) by $v_i = 1 - u_i$.
3. Calculate $V = (B/n)[\sum_{i=1}^{n/2} H*u_i + \sum_{i=1}^{n/2} H*v_i]$.

CHAPTER V

RESULTS AND DISCUSSION

The error from Smalian's method, as shown in Table IV, is positively biased. The error however decreases from above 14 % for 4 sampling points per tree to about 1.3 % for 16 samples per tree. On the other hand, the simulation study shows that Monte Carlo methods tested are all very close to the "true mean", their signed errors are very close to zero and have signs on both sides of zero (Table V - Table XIV). Apprarently, this is due to the fact that they all estimate the true mean without bias. Increasing sample size however does not reduce absolute error or coefficient of variation so fast.

Critical height sampling, with or without the use of antithetic variates, exhibits a very large coefficient of variation and large absolute error (Table V and Table VI). This indicates that volume estimates scatter widely around the true mean. Critical height with antithetic variates reduces variance considerably, but the result is not as good as expected by Van Deusen and Lynch (1986). Since a method that results in smaller variance is desirable, critical height sampling apparently should not be the choice.

The sample-mean Monte Carlo of the cross-section method

TABLE IV

ERRORS FROM SMALIAN'S METHOD
WITH DIFFERENT SAMPLES

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	SMALIAN'S VOLUME	% SIGN ERROR	% ABSOLUTE ERROR
4	70.1088	+14.0219	14.0219
6	66.6867	+ 8.4563	8.4563
8	64.7598	+ 5.3225	5.3225
10	63.5182	+ 3.3032	3.3032
12	62.7802	+ 2.1030	2.1030
14	62.4688	+ 1.5965	1.5965
16	62.2979	+ 1.3185	1.3185

TABLE V

RESULTS OF CRITICAL HEIGHT SAMPLING

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
4	- 0.3306	54.3363	183.0602	15.9252	100.5098
6	-11.1243	42.3366	204.0801	30.4307	122.9002
8	- 2.8266	33.6566	246.7488	47.8993	118.1132
10	5.9634	36.7064	202.8155	47.2608	123.0994
12	- 1.6100	31.0280	209.2961	75.0054	124.0201
14	0.3606	27.4318	175.2700	80.1576	118.5715
16	- 8.9875	20.8463	185.4969	74.8796	126.4535

TABLE VI

RESULTS OF CRITICAL HEIGHT SAMPLING
WITH ANTITHETIC VARIATES

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
4	-0.1662	33.0617	107.4214	0.9388	44.7579
6	0.6980	21.1585	110.2519	14.4007	53.5561
8	6.2068	21.4950	110.5417	7.6648	50.9656
10	1.2610	19.6775	95.4558	24.0288	53.2715
12	-1.6980	18.7067	78.5120	20.6668	52.2430
14	-4.1458	19.2832	91.3613	22.3987	54.7235
16	-1.0309	11.9846	79.2977	24.0127	52.4792

is poor as expected. The absolute error is almost 25 %, and the coefficient of variation is almost 70 % for 4 samples per tree (Table VII). The use of antithetic variates along with the sample-mean method reduces the absolute error to 13 % and coefficient of variation to 12 % for 4 samples per tree (Table VIII). This improvement is not enough, however. The absolute error is in the same order as the absolute error from Smalian's formula for small number of samples.

Monte Carlo computation of the cross-section method with a power function model of solids of revolution as control function is intermediate in performance when judged from the size of absolute error and/or that of coefficient of variation (Table IX). The best form of solids of revolution used as control function appears to be a neiloid, when the power of control function is 1.5.

Using the equation from Kozak et al as control function does not improve the accuracy of estimation. This indicates that taper equations developed through regression do not approximate tree better than the power function model of solids of revolution (Table X).

Comparing the result from using equation by Kozak et al and that from using equation of solids of revolution shows that choosing a cone as control function is comparable to using equation by Kozak et al. In both cases, the absolute errors are below 10 % with only 4 sample points for each tree. This is much better than the estimate from Smalian's formula with the same sample points, which is 14 % for 4

TABLE VII

THE SAMPLE-MEAN MONTE CARLO INTEGRATION
OF THE CROSS-SECTION METHOD

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V	AVERAGE C.V.
4	0.0234	25.4886	154.0277	28.8274	75.8004
6	6.6026	25.0540	144.1209	28.2300	75.1239
8	0.9913	22.9408	123.5812	30.3454	71.2806
10	-5.1870	19.7725	126.8572	41.8779	74.5864
12	1.4208	19.9306	112.4694	44.9995	76.2166
14	5.0915	17.5392	106.3077	51.0371	75.5293
16	2.1676	16.5638	117.2920	43.7578	75.4770

TABLE VIII

THE SAMPLE-MEAN MONTE CARLO INTEGRATION OF
THE CROSS-SECTION METHOD WITH
ANTITHETIC VARIATES

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
4	3.6052	12.7692	53.7102	0.0662	11.9070
6	-1.5775	9.0129	48.7669	0.3503	8.7695
8	2.2367	9.3529	45.7606	1.2311	11.7828
10	0.9626	7.7534	42.3030	1.5596	15.1243
12	-0.7140	6.8227	43.2093	0.7453	14.0822
14	0.1396	8.3192	43.2541	1.9731	15.2742
16	-0.0673	4.6711	37.3379	1.2015	17.4252

TABLE IX

MONTE CARLO INTEGRATION OF THE CROSS-SECTION METHOD WITH THE POWER FUNCTION MODEL OF SOLIDS OF REVOLUTION AS CONTROL VARIATES

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
POWER OF CONTROL FUNCTION:			0.50		
4	0.0510	12.9101	86.9156	3.1899	27.6776
6	1.9026	14.2923	82.7106	7.5094	27.0647
8	4.7948	13.8640	80.2082	7.2651	32.1145
10	0.7463	10.9052	62.8898	8.8415	28.0142
12	-0.2095	8.8652	66.0231	11.0528	33.8105
14	-1.1355	7.9342	65.5141	14.0954	34.0433
16	1.0131	9.6288	59.5235	12.9284	32.1395
POWER OF CONTROL FUNCTION:			1.00		
4	-2.5154	9.1138	59.1412	1.0492	9.2465
6	1.1493	9.5893	70.9799	1.0087	17.7769
8	-2.0285	7.2777	64.5510	1.9211	13.7950
10	-1.1058	7.3887	58.4549	2.7504	17.1443
12	1.0962	7.7291	59.2532	3.2262	19.3620
14	-1.2035	5.9181	52.4751	3.4989	20.3671
16	1.1430	6.3165	60.3064	3.3403	23.2527
POWER OF CONTROL FUNCTION:			1.50		
4	-1.1881	6.7679	68.0724	0.6346	12.2316
6	-1.6221	5.4457	52.5173	0.9210	9.6192
8	-0.1552	6.5231	58.2526	0.7783	17.2974
10	-2.5444	5.5578	54.6269	2.0791	14.9870
12	1.2104	6.7368	53.7797	2.2826	17.6643
14	1.9555	6.0064	46.1821	2.7623	18.8679
16	-0.7081	4.4110	48.8928	2.0710	17.0177

TABLE X

MONTE CARLO INTEGRATION OF THE CROSS-SECTION METHOD WITH KOZAK'S TAPER EQUATION AS CONTROL VARIATES

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
4	-0.3832	9.5803	58.8018	1.2695	20.4936
6	0.6424	8.9704	51.4510	2.4010	22.2000
8	-1.4345	7.1755	53.2778	4.0905	22.2166
10	-0.3647	6.3586	44.1845	6.4869	23.9959
12	1.8815	5.7552	41.9341	5.8769	22.5226
14	1.0891	5.0571	39.2983	6.6373	22.7172
16	-0.8969	4.3373	40.1608	8.4911	22.9995

sample points per tree.

In the cross-section method category, the results from computation with importance sampling again show that neiloid approximates the "true" taper better. But more importantly, the absolute error is reduce to around 7 % with only 4 samples per tree (Table XI). The better performance of importance sampling over use of control variates agrees to the common observations (Chandler, personal communication).

Importance sampling with antithetic variates pushes the accuracy of estimation a little further. With only 4 sampling points per tree, the absolute error is less than 5 % when the importance function is taken as a paraboloid and/or as a cone (Table XII). The average coefficient of variation also reduces drastically to about 5 % for 4 samples per tree.

Under the shell method category, importance sampling with parabola as importance function produces some results that even are worse than that from using control variates in the cross-section method category. The absolute error for 4 sampling points per tree almost reaches 27 % and the coefficient of variation in average is more than 60 % (Table XIII).

However, when antithetic variates are used along with importance sampling for the shell method, the absolute error with only 4 sampling points per tree reduces to less than 3 % (Table XIV). The average coefficient of variation is also the lowest among the methods tested.

TABLE XI

MONTE CARLO INTEGRATION OF THE CROSS-SECTION METHOD WITH THE POWER FUNCTION MODEL OF SOLIDS OF REVOLUTION AS IMPORTANCE FUNCTION

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
POWER OF IMPORTANCE FUNCTION : 0.50					
4	-0.6144	19.5816	89.1140	6.4361	37.9588
6	2.3159	15.0946	99.3568	11.1219	44.7046
8	0.1160	15.3664	78.4560	14.1368	40.3010
10	-3.8599	11.7319	71.5367	15.1853	38.1370
12	-1.3573	10.3638	72.9158	19.3912	43.9520
14	1.2066	9.4311	69.4515	21.6013	42.8508
16	0.9140	8.6344	70.6537	24.1000	43.7809
POWER OF IMPORTANCE FUNCTION : 1.00					
4	3.3191	11.5582	44.5834	2.7306	20.0826
6	-0.2532	7.8672	43.1789	2.3648	18.5178
8	-0.4305	5.8194	40.6707	3.0328	19.0031
10	0.4569	6.5420	41.7937	4.6276	19.7033
12	0.8698	6.4426	39.5331	7.0296	21.4502
14	0.1656	5.2021	34.9420	6.2122	20.9824
16	-1.2374	4.1886	39.1443	7.5008	19.8556
POWER OF IMPORTANCE FUNCTION : 1.50					
4	-0.0877	7.0261	31.0907	0.6331	12.7184
6	-1.1952	4.4989	42.1810	1.0032	13.5584
8	0.8322	4.4245	27.9177	4.0627	15.2728
10	-0.3281	3.8094	29.3392	4.1930	15.6856
12	-0.6910	3.7543	29.3065	5.2342	14.6767
14	-0.2736	3.9924	24.9385	3.6149	14.5018
16	-0.1845	3.7784	24.8913	4.3547	15.3071

TABLE XII

MONTE CARLO INTEGRATION OF THE CROSS-SECTION METHOD WITH ANTITHETIC VARIATES WITH THE POWER FUNCTION MODEL OF SOLIDS OF REVOLUTION AS IMPORTANCE FUNCTION

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715

SAMPLE NO	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
POWER OF IMPORTANCE FUNCTION : 0.50					
4	0.6216	4.9214	27.3534	0.0000	3.6565
6	-0.4670	4.0706	22.1692	0.3890	6.4980
8	0.3385	4.2360	19.7239	0.3077	6.9861
10	-0.1372	2.7676	17.5637	0.3847	6.6655
12	0.2177	3.7323	17.8787	0.9195	7.6327
14	0.8079	3.6680	19.6268	0.8194	9.2423
16	1.1712	3.5670	18.2547	1.0220	9.0073
POWER OF IMPORTANCE FUNCTION : 1.00					
4	-0.6771	4.6289	21.0848	0.0000	4.0805
6	0.0552	4.6134	17.0013	0.0641	5.4656
8	0.4491	4.3512	17.1598	0.1675	6.3829
10	-0.7855	3.0961	16.4919	0.0738	6.9532
12	0.1005	2.9057	14.7674	0.4980	7.5944
14	-0.2065	2.2341	13.2704	0.2572	7.5892
16	-0.9380	2.4440	14.9195	0.2805	7.3238
POWER OF IMPORTANCE FUNCTION : 1.50					
4	-0.8151	8.6854	37.2952	0.0000	8.7511
6	1.2986	6.7652	31.8811	0.0821	12.5717
8	-0.2861	6.6258	39.8862	0.3993	12.8469
10	-0.7454	6.5528	31.2080	0.7236	11.9297
12	-0.4254	5.7949	36.8361	0.5390	13.1096
14	-0.1996	5.2241	33.3232	1.1818	13.0455
16	0.8360	5.8027	26.9155	0.8659	14.1200

TABLE XIII

MONTE CARLO INTEGERATION OF THE SHELL METHOD WITH THE POWER FUNCTION MODEL OF SOLIDS OF REVOLUTION AS IMPORTANCE FUNCTION

NUMBER OF TREES ESTIMATED : 62
TOTAL VOLUME : 61.48715
POWER OF IMPORTANCE FUNCTION : 0.5

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
4	0.6342	26.7090	134.4440	8.0078	63.5565
6	-4.8886	23.0473	109.1666	17.3826	66.1813
8	3.3089	20.9219	102.9185	34.1760	66.1697
10	-3.3124	17.6479	91.1343	24.5546	65.3309
12	4.3761	14.0668	109.4598	33.9363	64.9713
14	-1.1415	13.6759	96.8338	41.7433	66.4002
16	-1.6863	12.7039	105.1240	41.7700	65.4794

TABLE XIV

MONTE CARLO INTEGRATION OF THE SHELL METHOD WITH ANTITHETIC VARIATES WITH THE POWER FUNCTION MODEL OF SOLIDS OF REVOLUTION AS IMPORTANCE FUNCTION

NUMBER OF TREE ESTIMATED : 62
TOTAL VOLUME OF TREES : 61.48715
POWER OF IMPORTANCE FUNCTION : 0.5

SAMPLE NO.	% SIGN ERROR	% ABSOLUTE ERROR	MAXIMUM C.V.	MINIMUM C.V.	AVERAGE C.V.
4	-0.4689	2.7888	12.6008	0.0263	4.4062
6	0.6970	2.3182	9.8979	0.8079	4.9568
8	0.6053	2.3167	9.3685	0.6523	5.1810
10	-0.0156	1.9411	9.7623	1.6845	5.4183
12	0.1751	1.9290	9.2602	2.0066	5.4213
14	0.6010	1.7371	7.5846	3.0704	5.0439
16	0.4005	1.4576	8.1090	2.1863	5.3644

To understand why importance sampling does not reduce variance under the shell method so significantly as the same technique under the cross-section method, the functions f(x)/r(x) for both methods are sketched, where x is the random variate, f(x) the original integrand, and r(x) the probability density function. The importance function has the shape of a neiloid for the cross-section method, and the shape of a paraboloid for the shell method.

The graph of the f(x)/r(x) from the cross-section method exhibits very little variation (Figure 9). This graph also explains why importance sampling with antithetic variates for the cross-section method does not improve the accuracy greatly. The accuracy cannot be improved further because, first, there is not much variation in the function f(x)/r(x); and second, f(x)/r(x) is neither a monotonic nor nearly a monotonic function. The graph of the f(x)/r(x) from the shell method, on the other hand, shows great variation (Figure 10). However, it is a strictly monotonic decreasing function. This explains why use of antithetic variates along with importance sampling in the shell method is very effective in variance reduction.

Concern about the infinite variance of the function f(x)/r(x) from the shell method may arise. Figure 9 seems to indicate that when the random variate aproaches 0 the function f(x)/r(x) approaches infinity. If this is true then one has an unbounded integrand, or the worst "peak" one can expect from importance sampling. Fortunately, this

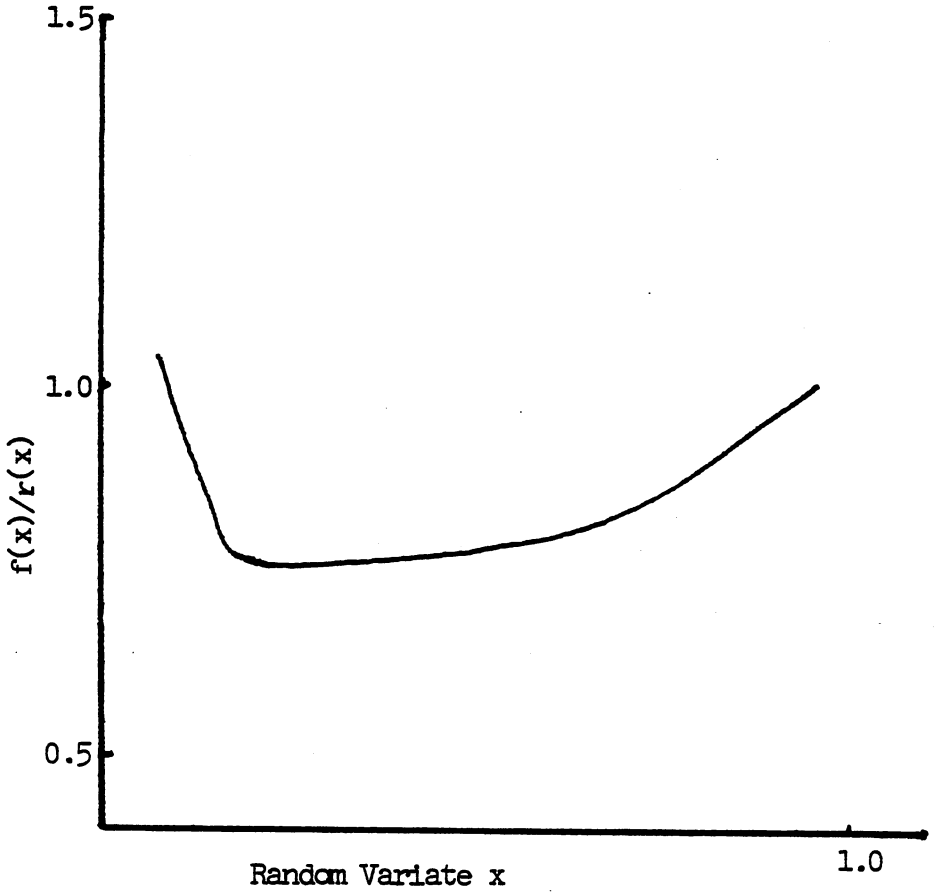

Figure 9. Graph of the Function f(x)/r(x) Resulted from Importance Sampling of the Cross-section Method with the Power Function Model of Solids of Revolution as Importance Function.

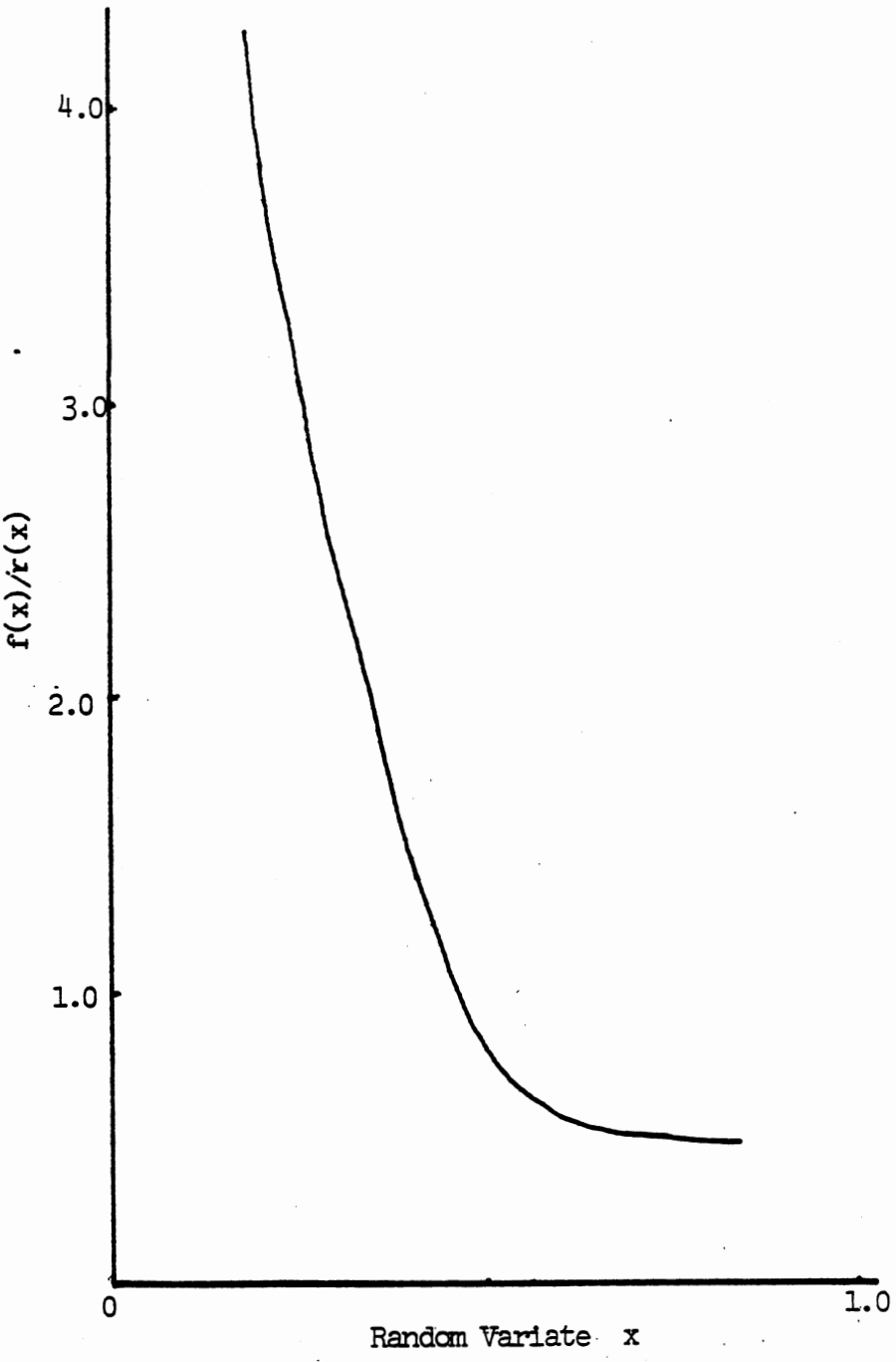

Figure 10. Graph of the Function f(x)/r(x) resulted from Importance Sampling of the Shell Method with the Power Function Model of Solids of Revolution as Importance Function.

is not the case. One can prove that the integrand is bounded when the random variate approaches 0.

Since the power function model of the solids of revolution is used as the importance function and the shape of the importance function is set to be a parabola, the pdf r(y) is

$$r(y) = (4y/B^2)(1 - y^2/B^2).$$

The random variate x is

$$x = y^2/B^2 = 1 - \sqrt{(1 - u)},$$

where u is distributed uniformly in (0, 1).

The pdf r now can be expressed in terms of the random variate x as

$$r(x) = 4\sqrt{x}(1 - x)/B.$$

The function f(y) is

$$f(y) = 2\pi kyh(y),$$

h is the height of a point with diameter y. In terms of the random variate x, f can be expressed as

$$f(x) = 2\pi kB\sqrt{x}h(x).$$

The ratio of the two functions is

$$f(x)/g(x) = 2\pi kB\sqrt{x}h(x)/[4\sqrt{x}(1 - x)/B]$$
$$= (kB\pi/2)h(x)/(1 - x).$$

When the random variate x approaches 0 the function f(x)/g(x) approaches $\pi kBH/2$, since the value of h(0) is H, the height of the tree.

The simulation results indicate that the best Monte Carlo method is importance sampling with antithetic variates under the shell method category. Importance sampling with

antithetic variates under the cross-section category is the second best method. Using control variates under the cross-section category is acceptable if high accuracy is not required.

In this simulation study, most of the Monte Carlo methods investigated show better performance in volume estimation than Smalian's formula does. Methods studied here all are unbiased. By choosing appropriate variance reduction techniques, the accuracy of estimation can be improved greatly.

It is also demonstrated in this study that the simple geometrical model, the power function model of solids of revolution, when used as importance function or as control function, provides as much information for excurrent trees as some taper equatons developed through regression.

CHAPTER VI

SUMMARY AND CONCLUSION

1. The objective of this study is to find some Monte Carlo methods that estimate the volume of trees without bias, and that out-perform Smalian's method when the number of samples taken from each tree is small.
2. All Monte Carlo simulations result in unbiased estimations of true tree volume. In contrast, Smalian's formula is known to be positively biased, and this is illustrated in this study.
3. Critical height sampling, with or without the use of antithetic variates, results in large absolute error and large average size of coefficient of variation, and is inferior to Smalian's formula.
4. The best Monte Carlo method simulated is importance sampling with antithetic variates under the shell method with the power function model of solids of revolution equivalent to a paraboloid as importance function.
5. Simulation results also conclude that the power function model of solids of revolution provides enough information about excurrent trees to be used as an auxillary function in volume estimation. Equation (2)

for solids of revolution can be used either as an importance function or as a control function.

6. Simulation results indicate that one should use importance sampling with antithetic variates under the shell method when very high accuracy is desirable. Importance sampling with antithetic variates under the cross-section method provides slightly less accurate estimation but it is easy to apply in real situation. Volume estimation with control variates can be used if accuracy is not critical.

BIBLIOGRAPHY

[1] Brickell, J. E. 1984. Stem analysis: a conventional approach to volume determination. Paper Presented at: Growth and Yield and Other Mensurational Tricks: A Regional Technical Conference, Logan, Utah, November, 1984.

[2] Burkhart, H. E. 1977. Cubic-foot volume of loblolly pine to any merchantable top limit. South J Appl For 1:7-9.

[3] Cao, Q. V., H. E. Burkhart, and T. A. Max. 1980. Evaluation of two methods for cubic-volume prediction of loblolly pine to any merchantable limit. Forest Sci 26: 71-80.

[4] Gregoire, T. G., H. T. Valentine, and G. M. Furnival 1985. Estimation of volume with a taper function. Society of American Foresters working groups in inventory, biometrics, and remote sensing: Use of Auxiliary Information in Resource Inventories. Blacksburg, VA, October, 1985.

[5] Gregoire, T. G., H. T. Valentine, and G. M. Furnival 1986. Estimation of bole volume by importance sampling. Can J For Res 16: 554-557.

[6] Halton, J. H. 1965. On the relative merits of correlated and importance sampling for Monte Carlo integration. Proc Cambridge Phil Soc 61: 497-498.

[7] Halton, J. H. 1970. A retrospective and prospective survey of the Monte Carlo method. SIAM Review 12: 1-61.

[8] Hammersley, J. M. and D. C. Handscomb. 1964. Monte Carlo Methods. Methuen & Co., London, 178 p.

[9] Hammersley, J. M. and K. W. Morton. 1956. A new Monte Carlo technique: antithetic variates. Proc. Cambridge Phil Soc 52: 449-474.

[10] Hine, W. W. and D. C. Montgomery. 1980. Probability and Statistics in Engineering and Management Science. 2nd ed. John Wiley & Sons, N.Y., 634 p.

[11] Husch, B., C. I. Miller, and T. W. Beers. 1982. Forest Mensuration. 3rd ed. John Wiley & Son, N.Y., 402 p.

[12] Kahn, H and A. W. Marshall. 1953. Methods of reducing sample size in Monte Carlo Computations. J Oper Res 1: 263-278.

[13] Kitamura, M. 1964. On an estimate of the volume of treesin a stand by the sum of critical heights. 73 Kai Niki Rin Ko. p.64-67.

[14] Kozak, A., D. D. Munro, and J. H. G. Smith. 1969. Taper functions and their application in forest inventory. Forest Chron 45: 278-283.

[15] Max, T. A. and H. E. Burkhart. 1976. Segmented polynomial regression applied to taper equations. Forest Sci 22: 283-289.

[16] Ormerod, D. W. 1973. A simple bole model. Forest chron 49: 136-138.

[17] Rubinstein, R. Y. 1981. Simulation and the Monte Carlo Method. John Wiley & Sons, N.Y., 278 p.

[18] Van Deusen, P and T. B. Lynch. 1986. Two methods for unbiased tree-volume estimation. In review.

VITA

Shen-Then Chang

Candidate for the Degree of

Master of Science

Thesis: MONTE CARLO METHODS IN FOREST INVENTORY

Major Field: Computing and Information Science

Biographical:

Personal Data: Born in Taipei, Taiwan, Republic of China, January 18, 1952, the son of Mr. and Mrs. Chan-Chi Chang.

Education: received Bachelor of Science in Agriculture degree in Plant Pathology from National Taiwan University in 1974; received Master of Science degree in Plant Science from University of Idaho in 1979; received Doctor of Philosophy degree in Plant Pathology from Oklahoma State University in 1984; completed requirements for Master of Science degree in Computing and Information Science in July, 1987.

Professional Experience: Research assistant, Department of Plant Pathology, National Taiwan University, 1976; graduate research assistant, Department of Plant Pathology, Oklahoma State University, 1981-1983; teaching assistant, Department of Computer and Information Sciences, Oklahoma State University, 1984-1986; assistant professor, Department of Computer Science, St. Cloud State University, Minnesota, 1986-present.