# A DESCRIPTION OF VIRGINIA PINE DIAMETER DISTRIBUTIONS

by

Frank Thomas Lloyd

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Forest Biometrics

APPROVED:

Chairman, Thomas C. Evans

Lawrence S. Davis

James Pickands, III

Emmett F. Thompson

May 1966

Blacksburg, Virginia

# TABLE OF CONTENTS

TABLES

# INTRODUCTION

For the forest manager and forest mensurationist the yield table is a familiar and useful concept. There are too many types and formats of yield tables to enumerate here, but they all possess one common property: a tabulation of some preferred volume measurement classified by useful and measurable forest stand attributes, notably age and site quality. Degrees of sophistication are apparent upon inspecting a typical collection of yield tables, for example, one might only tabulate cubic foot volume while another might tabulate board foot, cubic foot and cord volume. Indeed, they all attempt to relate volume to some elementary measures of stand structure, where stand structure here means measurable attributes that indicate volume, value, quality or associated factors. These tables have been developed for the primary and expressed purpose of aiding the forester in making growth and yield projections and predictions.

A widely used measure of stand structure is the stand table, which tabulates diameter frequencies by useful stand variables, such as, age and site index. Aided by these diameter frequencies, the forester can obtain estimates of timber quality and value because of the positive relationship of diameter with quality. Obviously, increasing complexity of an associated yield table, makes construction of accompanying stand tables extremely involved. Imagine a tabulation of diameter frequencies by stand age and site index, and then add a third variable, say density, and note the increased complexity of the analysis and tabulation.

In addition to indicating quality, a sufficiently developed stand table could possibly be used to estimate volume and totally replace the conventional yield table. Conceptually, using diameter frequency distributions in timber management is attractive for intensive management. The associated mapping of stand variables and detailed record keeping, gives the manager a logical and precise technique for making either inventory estimates or decision oriented predictions.

To resolve the problems of complexity associated with the tabular model of the stand table, an obvious alternative is a mathematical model. If it was possible to specify a mathematical approximation of the tabular model, then the stand table concept might more easily include a wider range of stand variables, and serve as a more flexible tool for making growth and yield projections. The mathematical model is especially appealing from the analysis and calulation point of view because of the wide acceptance and availability of high speed computers.

# OBJECTIVE

The objective of the study was to obtain a mathematical approximation for the distributions of breast height diameters of naturally regenerated, even-aged, pure Virginia pine. Problems associated with the objective fall into three general categories: (1) selection of a frequency estimator, and the estimation of the required parameters, (2) testing the predicted frequencies against observed frequencies, and (3) association of predicted frequencies with useful and identifying properties of the forest stands, such as, site quality, stand density and age.

# LITERATURE REVIEW

The concept of mathematical approximation of the frequency distributions of tree diameters is not new. The first efforts in American forestry, Meyer (1928, 1930), Schnur (1934), Schumacher (1928), attempted to fit the Gram-Charlier types A and B and Pearsonian Types I and III curves to even-aged species.

More recently, Hurst (1957) fitted five types of curves to four groups of observed diameter frequency distributions of Shorea Leprosula by the method of moments. The four groups represented four permanent plots, each plot being initially treated by a different silvicultural method and each plot measured at three successive and equal time intervals, thus resulting in twelve diameter frequency distributions. He investigated the parameters from each curve for correlation with time and silvicultural treatment. His analysis indicated that the Pearson three parameter type I curve did the best job of both fitting the data and producing parameter estimates that were significantly correlated with time and treatment.

Clutter and Bennett (1965) developed a model for describing the frequency distribution of the diameters in planted slash pine, using as their model the following modification of the beta distribution:

$$f(x_i) = \frac{(\alpha+\beta+1)!}{\alpha!\beta!} x_i^{\alpha}(1-x_i)^{\beta} \quad , \quad 0 \le x_i \le 1$$

$$= 0 \quad , \quad \text{otherwise}$$

where:

$$x_i = \frac{D_i - D_{min}}{D_{max} - D_{min}} = \text{coded tree diameter}$$

$D_i$ = uncoded diameter

$D_{max}$ = estimated maximum diameter in the stand

$D_{min}$ = estimated minimum diameter in the stand.

The equation $x_i = \frac{D_i - D_{min}}{D_{max} - D_{min}}$ is used merely to code tree diameters ($D_i$) so all diameters of the stand fall in the range of the beta distribution ($0 \le x_i \le 1$). $D_{max}$ and $D_{min}$ were estimated by multiple regression equations of age, number of trees per acre and site index at 25 years of age, and therefore, represent the average maximum or minimum diameter for a given combination of the three variables.

# TECHNIQUES AND PROCEDURES

The data for this study were obtained from the Southeastern Forest Experiment Station, Forest Service, United States Department of Agriculture. They consist of 161 plots of Virginia pine, located in the Piedmont of Maryland, Virginia, North Carolina and South Carolina. The plots were either 1/10 acre or 1/4 acre in size, depending on the density of the stand being sampled. Although plots containing 10 percent Virginia pine or more were originally measured, only "pure" stands with 80 percent or more Virginia pine were used in this analysis. The number of plots was further reduced by excluding those whose diameter distributions had been disturbed, e.g., plots which had been burned or partially cut. Only 51 plots were ultimately accepted.

The apparent ability of the beta distribution to describe diameter distributions of planted slash pine prompted its use as a plausible hypothesis for a similar test in the natural stands of Virginia pine under consideration here. In general, the procedure will consider: (1) methods of estimating the population parameters $\alpha$ and $\beta$ from observed plot frequencies, (2) tests of goodness of fit of the predicted frequencies, and (3) methods appropriate to using the distribution for growth and yield projections.

Two methods of estimating population parameters are plausible, the method of moments and the method of maximum likelihood. Sometimes, they yield identical estimators, but for $\alpha$ and $\beta$ of the beta distribution the two methods produce different estimators. The question logically arises as to which method to use. This can best be answered by investigating the merits of the two methods as they apply to the beta distribution.

In general, maximum likelihood estimates of distribution parameters have certain theoretical properties that make them more desirable than moment estimates. The variance of maximum likelihood estimates will always be at least as small as the variance of moment estimates of the same distribution parameter. Also, for large samples, a maximum likelihood estimator is asymptotically normally distributed and possesses minimum asymptotic variance among all asymptotically normally distributed estimators. Although maximum likelihood estimators possess theoretically desirable properties not associated with moment estimators, the possibility exists that there are some practical criteria that might justify using moment estimates.

If it appears plausible to use the beta distribution for the frequency estimator, one naturally asks if diameter at breast height is beta distributed. This question can best be answered by using the familiar chi-square goodness of fit test. Chi-square stipulates that expected frequencies must be calculated with maximum likelihood estimates of the parameters, although it is worthwhile to point out that this restriction is sometimes ignored and moment estimates are used. The expected frequencies are obtained by multiplying the number of trees per acre times the integral of each coded, 1-inch diameter class.

Once the distribution is defined and justified, its use as a tool for growth and yield projection requires evaluation of the correlation between the constants of the distribution and the stand variables which comprise the forester's control, notably age, site and density. The constants of interest for the beta distribution are the two parameters, $\alpha$ and $\beta$, and the limits of its range, 0 and 1. It is apparent that correlation of any useful and measurable stand variable with these constants would enable the forester to predict diameter distributions from the natural conditions encountered in managing a species.

In the earlier work aimed at using mathematical approximations for growth and yield projections, the researchers attempted to correlate stand variables to the distribution parameters. If high correlation appeared, they could then specify a particular distribution that was related to the stand variables they considered, by estimating the distribution parameters from the relationship they found. The difficulty with this approach has been finding frequency estimating functions that were sufficiently flexible to describe a wide variety of frequency distributions and at the same time possess distribution parameters that reflected properties in the diameter distribution that are related to the stand variables.

One would justly and logically expect to find correlation between stand variables and the maximum ($D_{max}$) and minimum ($D_{min}$) diameter of a stand of trees. For example, one would generally expect $D_{max}$ and $D_{min}$ to increase with increasing age. If this type of relationship could be described mathematically, then a forester could estimate the range ($D_{min}$ to $D_{max}$) for various combinations of stand variables and code the predicted range to the range of the beta distribution (0 to 1) by:

$$x_i = \frac{D_i - D_{min}}{D_{max} - D_{min}} , \qquad \text{(eq. 1)}$$

thus resulting in a useful tool for predicting diameter distributions. This would enable the forester to make growth projections by evaluating volume estimates over time. He would also be able to evaluate yield in terms of volume, quality, value, profit, costs, etc., for different but feasible combinations of controlling stand variables.

# RESULTS

## Estimating Parameters by Moments

Moment estimates of the parameters of the beta distribution are obtained by equating the population moments to the sample moments. The first and second population moments are (see Mood, 1963:114):

$$u_1 = \frac{\alpha+1}{\alpha+\beta+2} \qquad \text{(eq. 2)}$$

$$u_2 = \frac{(\alpha+1)(\alpha+2)}{(\alpha+\beta+2)(\alpha+\beta+3)} \qquad \text{(eq. 3)}$$

where $\alpha$ and $\beta$ are the unknown population parameters. The $n^{\underline{th}}$ sample moment is (see Elderton, 1953:15):

$$m_n = (f_1\, x_1^n + f_2\, x_2^n + \cdots + f_i\, x_i^n)/n \qquad \text{(eq. 4)}$$

where:

$f_i$ = number of trees in the $i^{\underline{th}}$ 1-inch diameter class

$x_i$ = mid-point value of $i^{\underline{th}}$ 1-inch diameter class

$n = \sum_{i=1}^{k} f_i$ = total number of trees per acre

$k$ = the number of 1-inch diameter classes.

Solving the first and second population moments (eq. 2 and 3) for $\alpha$ and $\beta$, and equating

$$m_1 = u_1$$

$$m_2 = u_2$$

results in the following moment estimates of the parameters:

$$\alpha = \frac{2m_1^2 - m_1 m_2 - m_2}{m_2 - m_1^2} \qquad \text{(eq. 5)}$$

$$\beta = \frac{m_1 + m_1 m_2 + 2m_2}{m_2 - m_1^2} \qquad \text{(eq. 6)}$$

where $m_1$ and $m_2$ are calculated by eq. 4 from the observed diameter frequency distributions.

## Estimating Parameters by Maximum Likelihood Using Newton-Raphson Technique

Maximum likelihood estimates of the parameters of the beta distribution are those estimates which maximize the likelihood function:

$$L = \prod_{i=1}^{k} \frac{(\alpha+\beta+1)!}{\alpha!\beta!} x^{\alpha f_i} (1-x)^{\beta f_i}, \qquad \text{(eq. 7)}$$

$0 < \alpha < \infty$,

$0 < \beta < \infty$.

For easier arithmetic, the likelihood function is often expressed as:

$$\log L = \alpha \sum_{i=1}^{k} f_i \log x_i + \beta \sum_{i=1}^{k} f_i \log (1-x_i)$$

$$+ n(\log (\alpha+\beta+1)! - \log \alpha! - \log \beta!) \quad \text{(eq. 8)}$$

where $x_i = \frac{D_i - D_{min}}{D_{max} - D_{min}}$. To find the maximum likelihood estimates of $\alpha$ and $\beta$, one takes the partial derivative of log L (eq. 8) with respect to $\alpha$ and with respect to $\beta$, equates the derivatives to zero and solves for $\alpha$ and $\beta$. The derivatives are:

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^{k} f_i \log x_i + n\left(\frac{\partial \log}{\partial \alpha} \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}\right). \quad \text{(eq. 9)}$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{k} f_i \log(1-x_i) + n\left(\frac{\partial \log}{\partial \beta} \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}\right).$$

(eq. 10)

Because of the difficulty in solving for $\alpha$ and $\beta$ in the two equations above, it is desirable to use an approximation that can be solved more easily (see Sokolnikoff, 1939:319-321). The approximations of the derivatives are:

$$\left.\frac{\partial \log L}{\partial \alpha}\right|_{\substack{\alpha+\Delta\alpha \\ \beta+\Delta\beta}} = \left.\frac{\partial \log L}{\partial \alpha}\right|_{\alpha,\beta} + \left.\frac{\partial^2 \log L}{\partial \alpha^2}\right|_{\alpha,\beta} \Delta\alpha + \left.\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right|_{\alpha,\beta} \Delta\beta = 0$$

(eq. 11)

and

$$\left.\frac{\partial \log L}{\partial \beta}\right|_{\substack{\alpha+\Delta\alpha \\ \beta+\Delta\beta}} = \left.\frac{\partial \log L}{\partial \beta}\right|_{\alpha,\beta} + \left.\frac{\partial^2 \log L}{\partial\alpha\partial\beta}\right|_{\alpha,\beta}\Delta\alpha + \left.\frac{\partial^2 \log L}{\partial\beta^2}\right|_{\alpha,\beta}\Delta\beta = 0$$

(eq. 12)

where:

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^{k} f_i \log x_i + n\left(\frac{\partial \log}{\partial \alpha}\Gamma(\alpha+\beta+2) - \frac{\partial \log}{\partial \alpha}\Gamma(\alpha+1)\right)$$

(eq. 13)

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{k} f_i \log(1-x_i) + n\left(\frac{\partial \log}{\partial \beta}\Gamma(\alpha+\beta+2) - \frac{\partial \log}{\partial \beta}\Gamma(\beta+1)\right) \qquad \text{(eq. 14)}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = n\left(\frac{\partial^2 \log}{\partial \alpha^2}\Gamma(\alpha+\beta+2) - \frac{\partial^2 \log}{\partial \alpha^2}\Gamma(\alpha+1)\right) \qquad \text{(eq. 15)}$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = n\left(\frac{\partial^2 \log}{\partial \beta^2}\Gamma(\alpha+\beta+2) - \frac{\partial^2 \log}{\partial \beta^2}\Gamma(\beta+1)\right) \qquad \text{(eq. 16)}$$

and

$$\frac{\partial^2 \log L}{\partial\alpha\partial\beta} = \frac{\partial^2 \log L}{\partial\beta\partial\alpha} = n\left(\frac{\partial^2}{\partial\alpha\partial\beta}\log\Gamma(\alpha+\beta+2)\right). \qquad \text{(eq. 17)}$$

The first and second derivatives of the logarithm of the gamma function in equations 13 and 17 above are called the digamma and trigamma respectively (see Pairman, 1919). The infinite series expansion formulas for the digamma and trigamma for different ranges of the values of the argument are given in the appendix. "Argument" in the above sentence means different values of $\alpha$ and $\beta$ or combinations of $\alpha$ and $\beta$ and some constant. For example, the argument for $\log \Gamma(\alpha+\beta+1)$, is the linear combination $\alpha+\beta+1$. It is also important to realize that the derivative of a function of a function, i.e., $f[g(x)]$ is obtained as follows:

$$\frac{d\ f[g(x)]}{dx} = \frac{d\ f[g(x)]}{d\ g(x)} \cdot \frac{d\ g(x)}{dx} .$$

Therefore,

$$\frac{\partial}{\partial\alpha} \log \Gamma(\alpha+\beta+2) = \frac{\partial}{\partial\beta} \log \Gamma(\alpha+\beta+2) \qquad \text{(eq. 18)}$$

$$\frac{\partial^2}{\partial\alpha^2} \log \Gamma(\alpha+\beta+2) = \frac{\partial^2}{\partial\beta^2} \log \Gamma(\alpha+\beta+2) = \frac{\partial^2}{\partial\alpha\partial\beta} \log \Gamma(\alpha+\beta+2). \qquad \text{(eq. 19)}$$

In order to start this iterative technique of calculating maximum likelihood estimates, it is necessary to define initial values of the parameters, call them $\alpha_o$ and $\beta_o$, such that:

$$0 < \alpha_o < \infty,$$

$$0 < \beta_o < \infty.$$

For this project, it was convenient to use the moment estimates of $\alpha$ and $\beta$ as the initial values (starting points). The process is started by equating:

$$\alpha_1 = \alpha_o$$

$$\beta_1 = \beta_o$$

and solving the following two equations:

$$\left.\frac{\partial^2 \log L}{\partial \alpha^2}\right|_{\alpha,\beta} \Delta\alpha + \left.\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right|_{\alpha,\beta} \Delta\beta = -\left.\frac{\partial \log L}{\partial \alpha}\right|_{\alpha,\beta} \quad \text{(eq. 20)}$$

$$\left.\frac{\partial^2 \log L}{\partial \beta \partial \alpha}\right|_{\alpha,\beta} \Delta\alpha + \left.\frac{\partial^2 \log L}{\partial \beta^2}\right|_{\alpha,\beta} \Delta\beta = -\left.\frac{\partial \log L}{\partial \beta}\right|_{\alpha,\beta} \quad \text{(eq. 21)}$$

for the two unknowns $\Delta\alpha_1$ and $\Delta\beta_1$, using $\alpha_1$ and $\beta_1$ wherever there is an $\alpha$ and $\beta$ in the equations. The second iteration will yield $\Delta\alpha_2$ and $\Delta\beta_2$ where the values of $\alpha$ and $\beta$ in equations 19 and 20 are $\alpha_2$ and $\beta_2$, where:

$$\alpha_2 = \alpha_1 + \Delta\alpha_1$$

$$\beta_2 = \beta_1 + \Delta\beta_1.$$

The $i^{th}$ iteration will yield the answers $\Delta\alpha_i$ and $\Delta\beta_i$ with:

$$\alpha_i = \alpha_{i-1} + \Delta\alpha_{i-1}$$

$$\beta_i = \beta_{i-1} + \Delta\beta_{i-1}.$$

The process is continued until $\Delta\alpha_i$ and $\Delta\beta_i$ both obtain some predetermined small value. In other words, the value of $\Delta\alpha_i$ and $\Delta\beta_i$ theoretically converge to zero, consequently the process must be terminated. Practically, the values of $\Delta\alpha_i$ and $\Delta\beta_i$ will not converge to absolute zero, but rather, the inaccuracies of the estimating procedure caused convergence to cease at values of the order 0.00000049 to 0.000000049.

The theoretical absolute convergence of the technique is obvious if one inspects the equations used. Clearly, the Taylor series approximation to the derivatives in equations 11 and 12 have been equated to zero in equations 20 and 21, thus, it follows that if the derivatives are to equal zero, $\Delta\alpha_i$ and $\Delta\beta_i$ must vanish. As already pointed out, absolute convergence did not occur, the reason being the inability of the divergent infinite series expansion approximations of the digamma and trigamma to give better than 8 place accuracy.

## Maximum Likelihood Estimates by Two Other Techniques

Two additional techniques for finding maximum likelihood estimates were investigated in this project. For a visual conception of how these two methods work, it helps to think of the likelihood function (eq. 8) as a function of its parameters $\alpha$ and $\beta$. This function can be pictured as a surface in three dimensions with an $\alpha$-axis, a $\beta$-axis and a likelihood axis. The maximum likelihood estimates are the values of $\alpha$ and $\beta$ that correspond to the highest (maximum) point on the likelihood surface. Thus, both techniques are simply "brute force" techniques of finding the highest point on the conceptualized likelihood surface.

The first technique evaluates the logarithm of the likelihood (eq. 8) at each point in a grid of values of $\alpha$ and $\beta$ located around and initial pair of values in the $\alpha\beta$-plane. The point corresponding to the largest value of the likelihood function is used as the center point of a second grid of points which is reduced in size and has the points closer together. This process is continued until the points in the grids obtain some predetermined distance between them.

The second technique is basically the same idea, except it alternately varies $\alpha$ and $\beta$. The process is started by first choosing initial values of $\alpha$ and $\beta$. Then allow one of the parameters to increment its value over a range around the initial value. The value of the parameter being varied is then held fixed at the point where the largest likelihood was calculated while the other parameter is varied in a similar manner. As the process continues, the range and increment of the parameters are reduced. Both of these techniques, plus the Newton-Raphson method, worked, and assuming programming for all three was relatively efficient, the Newton-Raphson method appeared to be the fastest.

Comparison of the estimates of $\alpha$ and $\beta$ by moments and maximum likelihood show that differences were relatively small (Table 1). Both the moment estimates and maximum likelihood estimates are asymptotically normally distributed for large samples, thus it is reasonable to assume for this work that the estimates of $\alpha$ and $\beta$ are approximately normally distributed. Therefore, an additional comparison of the two methods might be the asymptotic variances, best calculated by:

$$S_{\hat{\alpha}}^2 = \sum_{i=1}^{51} \frac{(\hat{\alpha}_i - \bar{\hat{\alpha}})^2}{50}$$

and

$$s_{\hat{\beta}}^2 = \sum_{i=1}^{51} \frac{(\hat{\beta}_i - \bar{\hat{\beta}})^2}{50}$$

Table 1. Estimates of Alpha and Beta

| | Alpha | | Beta | |
|---|---|---|---|---|
| Plot | Mom. Est. | Ma. Li. Es. | Mom. Est. | Ma. Li. Es. |
| 1 | 3.003 | 2.700 | 2.204 | 1.909 |
| 2 | 2.798 | 2.528 | 6.206 | 5.646 |
| 3 | 0.787 | 0.887 | 1.965 | 2.068 |
| 4 | 1.076 | 0.962 | 1.068 | 0.990 |
| 5 | 0.893 | 0.887 | 0.785 | 0.793 |
| 6 | 1.854 | 1.761 | 3.001 | 2.782 |
| 7 | 1.029 | 0.847 | 1.892 | 1.668 |
| 8 | 0.792 | 0.808 | 0.884 | 0.884 |
| 9 | 1.639 | 1.751 | 6.991 | 7.122 |
| 10 | 2.202 | 2.104 | 4.428 | 4.236 |
| 11 | 2.665 | 2.714 | 4.277 | 4.249 |
| 12 | 2.268 | 2.232 | 2.730 | 2.669 |
| 13 | 1.842 | 1.649 | 2.450 | 2.128 |
| 14 | 1.131 | 1.183 | 0.264 | 0.300 |
| 15 | 1.042 | 0.944 | 1.596 | 1.436 |
| 16 | 1.124 | 0.984 | 1.145 | 1.033 |
| 17 | 1.660 | 1.585 | 1.818 | 1.760 |
| 18 | 1.899 | 1.873 | 2.186 | 2.113 |
| 19 | 0.888 | 0.921 | 1.049 | 1.019 |
| 20 | 1.012 | 1.009 | 1.826 | 1.786 |
| 21 | 2.248 | 2.244 | 2.062 | 2.076 |
| 22 | 2.423 | 2.555 | 5.551 | 5.683 |
| 23 | 1.489 | 1.554 | 3.682 | 3.722 |
| 24 | 1.742 | 1.677 | 1.742 | 1.729 |
| 25 | 1.563 | 1.554 | 3.297 | 3.248 |
| 26 | 1.361 | 1.423 | 2.788 | 2.841 |

Table 1. Continued.

| | Alpha | | Beta | |
|---|---|---|---|---|
| Plot | Mom. Est. | Ma. Li. Es. | Mom. Est. | Ma. Li. Es. |
| 27 | 0.823 | 0.936 | 1.864 | 1.993 |
| 28 | 0.944 | 0.858 | 0.780 | 0.794 |
| 29 | 0.644 | 0.748 | 2.680 | 2.867 |
| 30 | 0.569 | 0.583 | 1.314 | 1.356 |
| 31 | 1.955 | 1.706 | 2.110 | 1.913 |
| 32 | 0.618 | 0.664 | 0.926 | 0.937 |
| 33 | 1.650 | 1.535 | 2.649 | 2.480 |
| 34 | 1.369 | 1.343 | 1.672 | 1.614 |
| 35 | 0.529 | 0.504 | 1.028 | 1.036 |
| 36 | 1.111 | 1.164 | 2.727 | 2.658 |
| 37 | 0.795 | 0.872 | 3.321 | 3.472 |
| 38 | 1.669 | 1.439 | 2.805 | 2.509 |
| 39 | 0.923 | 0.990 | 2.267 | 2.306 |
| 40 | 0.968 | 0.946 | 2.384 | 2.316 |
| 41 | 1.249 | 1.109 | 2.849 | 2.618 |
| 42 | 0.407 | 0.297 | 0.638 | 0.680 |
| 43 | 0.484 | 0.515 | 1.919 | 1.940 |
| 44 | 2.280 | 2.168 | 1.978 | 1.845 |
| 45 | 1.866 | 1.769 | 2.570 | 2.413 |
| 46 | 0.843 | 0.856 | 1.107 | 1.104 |
| 47 | 1.323 | 1.463 | 4.392 | 4.576 |
| 48 | 0.554 | 0.615 | 1.613 | 1.697 |
| 49 | 1.100 | 1.057 | 2.508 | 2.447 |
| 50 | 1.434 | 1.454 | 3.185 | 3.101 |
| 51 | 0.851 | 0.927 | 1.679 | 1.682 |

where

$\hat{\alpha}_i$ = an estimate of $\alpha$ for the $i^{\underline{th}}$ plot

$\hat{\beta}_i$ = an estimate of $\beta$ for the $i^{\underline{th}}$ plot

$$\bar{\hat{\alpha}} = \sum_{i=1}^{51} \hat{\alpha}_i/51$$

$$\bar{\hat{\beta}} = \sum_{i=1}^{51} \hat{\beta}_i/51 \text{ (Table 2).}$$

## Generating Expected Frequencies

The expected frequenceis were obtained by multiplying the integral for each coded 1-inch diameter class of the beta distribution times the number of trees expected or desired per acre. Since the integral of the beta distribution can not be obtained in closed form except for integer values, of $\alpha$ and $\beta$, it is necessary to use a method of numerical integration (see Adams and White, 1961:554-555). The integral is equal to the area under the curve and is calculated by summing the areas of very narrow trapazoids. Since the beta distribution as it is used here, is a probability density, the entire integral from 0 to 1 will equal 1. The width of the trapaziods used in the project was approximately equal to 0.01.

Table 2.

Asymptotic Variances

| | Moment | Max. Like. |
|---|---|---|
| Alpha | 0.41627 | 0.36409 |
| Beta | 1.88753 | 1.85824 |

The method seems good for approximating the integral of the beta distribution. The worst answer out of 102 integrations was 0.997. The absolute value of most of the errors was in the range 0.0001 to 0.000001.

## Chi-Square Goodness of Fit

The chi-square values indicate that diameter at breast height is not beta distributed (Table 3). It is of interest to note that chi-square values were also calculated using the moment estimates of the parameters and the resulting statistics were very close to the ones using the maximum likelihood estimates. Although theoretically this is not a sound comparison, it does give one an intuitive feeling that the two methods do not differ drastically.

## Estimation of $D_{max}$ and $D_{min}$

Initial plotting of the minimum diameter on stand variables age, site index and number of trees per acre suggested that the correlation would be slight. This indication plus the fact that 44 of the 51 plots had minimum diameters of one or two inches discouraged further trials. Since Virginia pine stands within the age range of the variable typically contain small trees in intermediate and surpressed crown classes, the low correlation was not surprising.

Table 3. Chi-Square Values

| Plot | Chi-Square | d.f.* | Plot | Chi-Square | d.f.* |
|---|---|---|---|---|---|
| 1 | 148.21 | 4 | 27 | 107.94 | 6 |
| 2 | 20086.33 | 10 | 28 | 73.38 | 6 |
| 3 | 137.30 | 5 | 29 | 7.38 | 3 |
| 4 | 18.35 | 6 | 30 | 78.59 | 4 |
| 5 | 11.86 | 6 | 31 | 44.30 | 6 |
| 6 | 203.44 | 6 | 32 | 30.12 | 4 |
| 7 | 84.86 | 6 | 33 | 42.75 | 9 |
| 8 | 31.08 | 5 | 34 | 33.95 | 2 |
| 9 | 506020.33 | 8 | 35 | 36.07 | 5 |
| 10 | 206.84 | 2 | 36 | 180.28 | 11 |
| 11 | 263.33 | 3 | 37 | 37.68 | 3 |
| 12 | 43.55 | 5 | 38 | 102.82 | 6 |
| 13 | 191.91 | 7 | 39 | 37.40 | 8 |
| 14 | 34.55 | 4 | 40 | 18.90 | 5 |
| 15 | 47.76 | 4 | 41 | 73.14 | 6 |
| 16 | 28.49 | 5 | 42 | 54.06 | 5 |
| 17 | 58.64 | 7 | 43 | 47.06 | 7 |
| 18 | 49.75 | 3 | 44 | 92.07 | 7 |
| 19 | 42.90 | 5 | 45 | 92.80 | 4 |
| 20 | 12.18 | 4 | 46 | 8.72 | 3 |
| 21 | 97.67 | 5 | 47 | 673.54 | 5 |
| 22 | 1108.59 | 5 | 48 | 5.38 | 3 |
| 23 | 260.70 | 6 | 49 | 36.16 | 3 |
| 24 | 26.64 | 3 | 50 | 678.34 | 3 |
| 25 | 85.80 | 5 | 51 | 216.39 | 3 |
| 26 | 59.15 | 7 | | | |

* Degrees of freedom

Similar plotting of maximum diameter exhibited a slight curvilinear response to density as measured by number of trees per acre and interactions of age with density and age with site index. A multiple regression analysis was run, therefore, according to the model:

$$\hat{Y} = \hat{\beta}_o + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5 + \hat{\beta}_6 X_6$$

where

$\hat{Y}$ = maximum tree diameter

$X_1$ = number of trees per acre

$X_2$ = age

$X_3$ = site index

$X_4 = X_1 X_2$

$X_5 = X_2 X_3$

$X_6 = \log_{10} X_1$

The resulting equation produced an $R^2$ of 0.615, where:

$$R^2 = \frac{\text{Sum of squares of regression}}{\text{Total sum of squares}}.$$

That is, the regression of the $X_i$'s explained 61.5 percent of the variation of $\hat{Y}$. Although this is significant for many problems in forestry, it is not high enough for this problem because only a few of the 51 estimates fell into the observed diameter classes.

# DISCUSSION

This project has failed to produce a mathematical approximation for the diameter frequency distributions of natural stands of Virginia pine. It has, however, defined some of the problem areas that must be resolved before a technique of this type can be used as a prediction and decision making tool for these and similar stands. It also has given an indication of the relative importance of the problems involved.

The most important single contribution has been the successful application of the Newton-Raphson technique to obtaining maximum likelihood estimates of the parameters of the beta distribution. Heretofore, a solution of this kind for $\alpha$ and $\beta$ has not been available. Of the three methods for finding maximum likelihood estimates, the Newton-Raphson technique appeared to be more efficient than either of the "brute force" methods, in that for equal decimal precision, it produced answers more rapidly than the others.

A comparison of the parameter estimates calculated by moments and by maximum likelihood, indicated that the two techniques produced estimates that were not drastically different. But it is equally clear that the comparison does not reflect the full impact of the differences upon any other alternative criterion such as plot volume.

Lacking a clear definition of a preferred criteria, Table 4 below, compares observed and expected frequencies for selected plots covering a range of the stand variables in age, site and density. Apparently, although not conclusively, the manner of estimation has only a trivial effect upon estimated frequencies. If the effect persists with other criteria, the moment estimates might be preferred simply because they are easier to calculate.

As with the earlier work (e.g. Meyer and Hurst, op. cit.), the most difficult problem has been associating the constants of distributions with stand variables which can be controlled or measured in the normal course of management. The earlier workers concentrated more specifically on the association of distribution parameters with stand variables. For planted slash pine, Clutter and Bennett (op. cit.) developed a different approach which limited the diameters to a predictable range ($D_{min}$ to $D_{max}$) and coded the diameters from 0 to 1, so the beta distribution could be used. In effect, the procedure suggests that the parameters, $\alpha$ and $\beta$, may be constant over the range of stand variables, and that the ultimate utility and flexibility of the distribution as a prediction device may depend upon the correlation of $D_{min}$ and $D_{max}$ with age, site, density, etc. Intuitively, this hypothesis appears supportable for young planted

Table 4. Expected and Observed Frequencies

**Plot 1**

| Dbh | Observe | Moment | Max.Li. |
|---|---|---|---|
| 0.5 | 10 | 9 | 11 |
| 1.5 | 30 | 95 | 101 |
| 2.5 | 360 | 269 | 263 |
| 3.5 | 420 | 420 | 400 |
| 4.5 | 450 | 424 | 414 |
| 5.5 | 160 | 254 | 269 |
| 6.5 | 90 | 48 | 61 |

**Plot 7**

| Dbh | Observe | Moment | Max.Li. |
|---|---|---|---|
| 0.5 | 68 | 45 | 54 |
| 1.5 | 96 | 111 | 116 |
| 2.5 | 104 | 144 | 141 |
| 3.5 | 136 | 149 | 142 |
| 4.5 | 184 | 132 | 126 |
| 5.5 | 128 | 104 | 98 |
| 6.5 | 44 | 64 | 65 |
| 7.5 | 8 | 28 | 32 |
| 8.5 | 12 | 5 | 6 |

**Plot 20**

| Dbh | Observe | Moment | Max.Li. |
|---|---|---|---|
| 0.5 | | | |
| 1.5 | | | |
| 2.5 | 56 | 71 | 70 |
| 3.5 | 172 | 164 | 162 |
| 4.5 | 220 | 192 | 191 |
| 5.5 | 156 | 171 | 171 |
| 6.5 | 116 | 120 | 122 |
| 7.5 | 52 | 58 | 60 |
| 8.5 | 16 | 11 | 12 |

**Plot 27**

| Dbh | Observe | Moment | Max.Li |
|---|---|---|---|
| 0.5 | | | |
| 1.5 | 40 | 93 | 82 |
| 2.5 | 280 | 191 | 186 |
| 3.5 | 230 | 224 | 226 |
| 4.5 | 160 | 216 | 223 |
| 5.5 | 170 | 184 | 189 |
| 6.5 | 180 | 136 | 139 |
| 7.5 | 70 | 84 | 83 |
| 8.5 | 30 | 37 | 35 |
| 9.5 | 10 | 6 | 6 |

Table 4. Continued.

| Plot 40 | | | | Plot 48 | | | |
|---|---|---|---|---|---|---|---|
| | Frequencies | | | | Frequencies | | |
| Dbh | Observe | Moment | Max.Li. | Dbh | Observe | Moment | Max.Li. |
| 0.5 | | | | 0.5 | | | |
| 1.5 | | | | 1.5 | 390 | 390 | 375 |
| 2.5 | 90 | 78 | 87 | 2.5 | 540 | 560 | 566 |
| 3.5 | 200 | 213 | 216 | 3.5 | 540 | 499 | 510 |
| 4.5 | 240 | 270 | 263 | 4.5 | 340 | 351 | 356 |
| 5.5 | 270 | 256 | 248 | 5.5 | 180 | 179 | 176 |
| 6.5 | 220 | 199 | 194 | 6.5 | 30 | 39 | 36 |
| 7.5 | 150 | 126 | 126 | | | | |
| 8.5 | 40 | 60 | 64 | | | | |
| 9.5 | 0 | 18 | 20 | | | | |
| 10.5 | 10 | 2 | 2 | | | | |

pine largely because such stands are very homogeneous. However, natural stands are not so homogeneous and typically display changes in skew with age. Thus, the failure of correlation to exist between the parameters and stand variables used in this analysis might be due to the small sample and the restricted ranges of the stand variables.

The second category of problems mentioned in the introduction is the comparison of expected frequencies with observed frequencies. Chi-square goodness of fit was used to test if the "actual" diameter distributions were beta distributed. By "actual" diameter distributions, is meant that the range of diameters was not estimated by $D_{min}$ and $D_{max}$, but rather, the observed values were used. Although most of the chi-square values were highly significant, one should not conclude that the beta distribution is not potentially useful. Before drawing conclusions, one should investigate the effect of a larger plot size would have on the chi-square values. A question of greater importance one should ask is, if chi-square is really the proper test to use for this problem being considered? If the model is to be used to predict volume, value, etc., then it logically follows that the model should be tested for its precision in estimating those quantities. Also, the model should be tested in its entirety, i.e., with the estimates of $D_{min}$ and $D_{max}$ rather than the observed values.

Estimation of the range ($D_{min}$ to $D_{max}$) could not be obtained for the Virginia pine data used in this analysis. An estimate of $D_{min}$ was not attempted because 86 percent of the plots had their minimum diameters in the one and two inch diameter classes. This response was not surprising, for it seems to be a property of natural stands of Virginia pine to retain small, suppressed trees. To eliminate this problem, it may be necessary to establish some criteria which could be used to determine the trees to be measured and those not to be measured. It might be desirable to simply truncate the distribution because of a lack of interest in trees smaller than some threshold diameter.

It is surprising that the upper limit of the range, $D_{max}$, displayed no correlation with stand variables, for it is to be expected that tree diameter is functionally related to age, to site and to stand density. Again, the small sample and limited range of age, site and density seemed to be causing the difficulty.

# CONCLUSIONS

Although the research reported here has not satisfied its primary objective of a mathematical expression of diameter distributions in natural stands of Virginia pine, several constructive conclusions and recommendations are possible. All of them are offered with the perspective that research in the interest of advancing growth and yield techniques in forestry is needed, that the study has been exploratory to begin with, and that the contributions presented here, however conclusive or trivial, are the beginnings of a foundation for further research work.

Certain deficiencies in the study are apparent, if not obvious. The major deficiency in the statement of the problem and its conclusion is the lack of a rigorous criterion, or criteria, for judging the success of a fitted distribution. Chi-square was used in this project to test the success of the model, but as the work progressed, it became increasingly clear that chi-square was not the best criterion for the purpose of this work. It is the utilitarian purpose of the model that more aptly determines its success, for it is the ability to describe with sufficient precision the attributes of the forest stand that justly determines whether or not the model is useful. Therefore, validation of the model should be connected with the forester's interest in volume, value, quality, cost estimation, return, profit and any other attributes useful in management of timber.

By far the most important deficiency appears in imperfections of the data. Although the data were originally collected to study the growth and silvicultural characteristics of natural Virginia pine in the mid-Atlantic states, two problems pertaining to this research are distinguishable. First, plots were measured in stands containing as little as 10 percent Virginia pine, while for this project it was necessary to have "pure" stands which are defined as those having 80 percent or more Virginia pine. Second, plots should be located in stands that are previously selected on the basis of possessing undisturbed distributions, and sufficiently cover the range of the stand variables. Elimination of plots from the original 161 because of contamination with other species and disturbances of the distributions resulted in a sample of only 51 plots. This is a plausible reason for the surprisingly low correlation of the maximum tree diameter with age, site, and density.

A second fault in the data was the lack of information on crown position of the trees. The prolonged period of establishment of natural stands and the relatively short life of Virginia pine results in stands that retain suppressed and intermediate trees. This may explain, at least in part, why the minimum tree diameter was not related to the stand variables. A possible solution to the problem would be to admit to analysis only those trees in the dominant and codominant crown classes.

There is also the unresolved problem of what plot size to use. The heterogeneity or clumpling of trees of similar sizes in natural stand of Virginia pine suggests that a larger plot might better represent the diameter distributions.

## ACKNOWLEDGMENTS

I sincerely thank for his assistance and guidance throughout my past three years of study at Virginia Polytechnic Institute. His comments and suggestions about my work, my thesis and my general preparation for a job are an unmeasurable part of my education, without which, the intellectual void that results from inexperience would be so very much larger.

I have appreciated very much the constructive comments in reference to my program of study and this thesis from

I also want to thank for his detailed reading of this thesis and his help with many of the statistical problems associated with it.

This work could not have been completed in the time that was available without the data on Virginia pine that was so generously donated by the Forest Service, U.S. Department of Agriculture.

Only my wife, truly knows the frustrations involved with this work. Her help and encouragement have been invaluable.

## LITERATURE CITED

Adams, L. J., and P. A. White. 1961. Analytic Geometry and Calculus. Oxford University Press, New York. 932 p.

British Association for the Advancement of Science Mathematical Tables. 1951. University Press, Cambridge. 1:19-25.

Elderton, W. P. 1953. Frequency Curves and Correlation. University Press, Cambridge. Edition. 15 p.

Hastings, C., Jr. 1955. Approximations for Digital Computers. Princeton University Press, Princeton, New Jersey. 158 p.

Hoel, G. P. 1962. Introduction to Mathematical Statistics. John Wiley and Sons, Inc., New York, London, Sydney. Edition 3. 233-234 p.

Hurst, D. C. 1957. Some Distribution Problems in Malayan Forestry. North Carolina State College Press, Raleigh, North Carolina. D. H. Hill Library, North Carolina State College. 45 p.

Mood, A. M., and F. A. Graybill. 1963. Introduction to the Theory of Statistics. McGraw-Hill, Co., Inc., New York. Edition 2. 443 p.

Pairman, E., M.A. 1954. Tracts for Computers. University Press, Cambridge. 6 p.

Schnur, G. L. April 1937. Yield, Stand, and Volume Tables for Even-Aged Upland Oak Forests. University of Pennsylvania Press. 560:87.

Schumacher, F. X. 1926. Yield, Stand, and Volume Tables for White Fur in the California Pine Region. University of California Printing Office, Berkeley, California. 407:27.

Sokolnikoff, I. S. 1939. Advanced Calculus. McGraw-Hill, Co., Inc., New York. 446 p.

Volume, Yield, and Stand Tables for Second - Growth Southern Pines. September 1929. Office of Forest Experiment Stations, Forest Service, and Cooperating Agencies, Washington, D.C. Miscellaneous Publication 50:202.

The vita has been removed from
the scanned document

# APPENDIX

Pairman (1919) published in Tracts for Computers, a set of tables for the digamma and trigamma for arguments ranging from 0 to 20. The formulas are divergent infinite series expansions that converge slightly before diverging, thus giving an approximation. Eight place accuracy was claimed for the tables, but to obtain eight place accuracy for the smaller values of the argument, it was necessary to use a recursive equation. Investigation of the formulas developed by Pairman showed that the following ranges of the argument and the corresponding formulas give at least eight place accuracy. The formulas for the digamma and trigamma for an argument between 20 and 60 were found in The British Association for the Advancement of Science Mathematical Tables; Volume 1, (1951).

Digamma:

Range of argument 0 to 1

$$\frac{d}{dx} \log x! = \frac{d}{dx} \log (x+3)! - \frac{1}{x+3} - \frac{1}{x+2} - \frac{1}{x+1}$$

where

$$\frac{d}{dx} \log (x+3)! = \log (x+4) - \frac{1}{2}\left(\frac{1}{x+4}\right) + \sum_{r=1}^{\infty} \frac{(-1)^r B_r}{2r(x+4)^{2r}}$$

Range of argument 1 to 2

$$\frac{d}{dx} \log x! = \frac{d}{dx} \log (x+2)! - \frac{1}{x+2} - \frac{1}{x+1}$$

where

$$\frac{d}{dx} \log (x+2)! = \log (x+3) - \frac{1}{2}\left(\frac{1}{x+3}\right) + \sum_{r+1}^{\infty} \frac{(-1)^r B_r}{2r(x+3)^{2r}}$$

Range of argument 2 to 3

$$\frac{d}{dx} \log x! = \frac{d}{dx} \log (x+1)! - \frac{1}{x+1}$$

where

$$\frac{d}{dx} \log (x+1)! = \log (x+2) - \frac{1}{2}\left(\frac{1}{x+2}\right) + \sum_{r=1}^{\infty} \frac{(-1)^r B_r}{2r(x+2)^{2r}}$$

Range of argument 3 to 20

$$\frac{d}{dx} \log x! = \log (x+1) - \frac{1}{2}\left(\frac{1}{x+1}\right) + \sum_{r=1}^{\infty} \frac{(-1)^r B_r}{2r(x+1)^{2r}}$$

Range of argument 20 to 60

$$\frac{d}{dx} \log x! = \frac{1}{2} \log(x(x+1)) + \sum_{r=1}^{\infty} \frac{(-1)^{r-1} B_r}{2r-1} \left(\frac{1}{x^{2r-1}} - \frac{1}{(x+1)^{2r-1}}\right)$$

Trigamma:

Range of argument 0 to 1

$$\frac{d^2}{dx^2}\log x! = \frac{d^2}{dx^2}\log(x+3)! + \frac{1}{(x+3)^2} + \frac{1}{(x+2)^2} + \frac{1}{(x+1)^2}$$

where

$$\frac{d^2}{dx^2}\log(x+3)! = \frac{1}{x+4} + \frac{1}{2}\left(\frac{1}{(x+4)^2}\right) + \sum_{r=1}^{\infty} \frac{(-1)^{r+1}B_r}{(x+4)^{2r+1}}$$

Range of argument 1 to 2

$$\frac{d^2}{dx^2}\log x! = \frac{d^2}{dx^2}\log(x+2)! + \frac{1}{(x+2)^2} + \frac{1}{(x+1)^2}$$

where

$$\frac{d^2}{dx^2}\log(x+2)! = \frac{1}{x+3} + \frac{1}{2}\left(\frac{1}{(x+3)^2}\right) + \sum_{r=1}^{\infty} \frac{(-1)^{r+1}B_r}{(x+3)^{2r+1}}$$

Range of argument 2 to 3

$$\frac{d^2}{dx^2}\log x! = \frac{d^2}{dx^2}\log(x+1)! + \frac{1}{(x+1)^2}$$

where

$$\frac{d^2}{dx^2}\log(x+1)! = \frac{1}{x+2} + \frac{1}{2}\left(\frac{1}{(x+2)^2}\right) + \sum_{r=1}^{\infty} \frac{(-1)^{r+1}B_r}{(x+2)^{2r+1}}$$

Range of 3 to 20

$$\frac{d^2}{dx^2}\log x! = \frac{1}{x+1} + \frac{1}{2}\left(\frac{1}{(x+1)^2}\right) + \sum_{r=1}^{\infty} \frac{(-1)^{r+1}B_r}{(x+1)^{2r+1}}$$

Range of 20 to 60

$$\frac{d^2}{dx^2}\log x! = \frac{1}{2}\left(\frac{1}{x} + \frac{1}{x+1}\right) + \sum_{r=1}^{\infty} (-1)^r B_r \left(\frac{1}{x^{2r}} - \frac{1}{(x+1)^{2r}}\right)$$

The $B_r$'s in the above formulas are Bernoullian numbers, the first twelve of which are:

| | |
|---|---|
| $B_1 = 0.16666667$ | $B_7 = 1.16666667$ |
| $B_2 = 0.03333333$ | $B_8 = 7.0921569$ |
| $B_3 = 0.02380952$ | $B_9 = 54.971178$ |
| $B_4 = 0.03333333$ | $B_{10} = 529.12424$ |
| $B_5 = 0.07575758$ | $B_{11} = 6,192.1232$ |
| $B_6 = 0.25311355$ | $B_{12} = 86,580.253$ |

ABSTRACT

Reported here is an attempt to specify a mathematical model that might supplement and eventually replace the yield table for making growth and yield projections for natural Virginia pine. The model used for these natural stands was similar to the one developed by Clutter and Bennett (1965) for planted slash pine. Although a workable model was not derived, some of the problem associated with the approach have been defined. In addition to defining the problems associated with the approach, a new technique for finding maximum likelihood estimates of $\alpha$ and $\beta$ was developed.