CONFIDENCE BANDS FOR POLYNOMIAL CURVES

By Paul G. Hoel

University of California, Los Angeles[1]

1. Summary. A method is given for constructing confidence bands for polynomial growth-type curves. The method assumes that the mean population size can be expressed as a polynomial in time and that the generalized T function for the mean values of the observations at fixed time points possesses a known and parameter-free distribution. Independence between observations at various times is not assumed. The method yields only a lower bound for the confidence coefficient.

2. Introduction. Consider a random variable y_t that represents some measurable characteristic of an individual from a population, or that represents the size of a population, at time t. The graph of $E(y_t)$ as a function of t, where E denotes expected value, will be called the mean growth curve.

One of the basic problems in studying growth and related phenomena is that of estimating the mean growth curve. In particular, it would be highly desirable to be able to construct a confidence band for the mean growth curve. Then the experimenter would be able to observe the accuracy of his sample curve as an estimate of the mean growth curve.

A method for getting a confidence band for a mean growth curve should be such that, corresponding to a given confidence coefficient C_0, the probability is C_0 that the entire curve will lie inside the band. This means that the probability is C_0 that for all t the mean growth curve ordinate $E(y_t)$ will lie between the corresponding ordinates of the two curves determining the confidence band. In this paper a method is presented for constructing such a confidence band for polynomial curves, but it yields a band which covers with probability $\geqq C_0$ rather than with the exact probability C_0.

Although the method will be described from the point of view of polynomial growth curves, the method is applicable to polynomial curves in general. The language of growth curves is used for its descriptive convenience and to stress the fact that the variables $y_1, \cdots, y_k$ which are involved are dependent variables, and hence that standard regression techniques are not applicable to this problem. The variable t may represent any physical quantity, although it will be described as time in this discussion.

3. Assumptions. It will be assumed that observations are always made at the times $t_1, t_2, \cdots, t_k$ and that $n > k$ independent sets of such observations are made. If y_t represents a measurable characteristic of an individual at time t,

Received 8/25/52, revised 3/9/54.

[1] Work done under the sponsorship of the Office of Naval Research.

this implies that n individuals are observed at the same stages of their growth. If y_t represents the size of a population at time t, this implies that the same initial size population is chosen each time and that the same time pattern for observations is used. It is possible to vary the initial size population, but the statistical interpretation then becomes more complicated. Let $y_1, y_2, \cdots, y_k$ denote the values of y_t at the specified time points for a randomly selected individual, or initial population, and let $\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_k$ denote the sample means of those variables for the n sets of observations.

Two basic assumptions will be made. First, it will be assumed that the mean growth curve is a polynomial of known degree, $k - 1$ or less. Second, it will be assumed that the distribution of the generalized T function for the variables $y_1, y_2, \cdots, y_k$ is known and does not depend upon any unknown parameters.

If one is studying growth problems, the first assumption may seem somewhat unnatural since exponential functions are often encountered in such problems. It is usually possible, however, to approximate such curves quite well over limited time ranges by polynomials of fairly low degree. Furthermore, by choosing a function of t as the independent variable, or by choosing a convenient function of y_t as the basic variable and assuming that its mean curve is a polynomial curve of degree $k - 1$ or less, the range of application of the method is extended considerably. If, however, a function of y_t is used, the interpretation will be in terms of the mean of this function rather than in terms of the mean of the variable. The methods to be presented are actually valid for finding confidence bands for curves expressible in the form $y_t = a_1 g_1(t) + \cdots + a_k g_k(t)$, where the $g_i(t)$ are any given functions of t. For such more general curves, however, the formulas derived in Section 4 for polynomials are not applicable.

The second assumption will be satisfied if the variables $y_1, y_2, \cdots, y_k$ are jointly normally distributed, because then the distribution of T^2 is well known [1]. Even though the variables $y_1, y_2, \cdots, y_k$ are not jointly normally distributed, the second assumption may still be considered to be satisfied if n is fairly large, because it can be shown that under mild restrictions T^2 possesses an asymptotic chi square distribution [2].

4. Derivation. Let the mean growth curve be a polynomial curve of degree $r - 1 \leqq k - 1$ and let the ordinates on this curve for the times $t_1, t_2, \cdots, t_k$ be denoted by $\mu_1, \mu_2, \cdots, \mu_k$. If the first r of the points $(t_1, \mu_1), (t_2, \mu_2), \cdots, (t_k, \mu_k)$ are chosen to determine the curve, its equation can be written in the Lagrange polynomial form

$$y = \sum_{i=1}^{r} \frac{(t - t_1) \cdots (t - t_{i-1})(t - t_{i+1}) \cdots (t - t_r)}{(t_i - t_1) \cdots (t_i - t_{i-1})(t_i - t_{i+1}) \cdots (t_i - t_r)} \mu_i . \tag{1}$$

The coordinates of the remaining $k - r$ points must of course satisfy equation (1).

Now introduce the variables $x_1, x_2, \cdots, x_r$ defined by

$$x_i = \frac{(t - t_1) \cdots (t - t_{i-1})(t - t_{i+1}) \cdots (t - t_r)}{(t_i - t_1) \cdots (t_i - t_{i-1})(t_i - t_{i+1}) \cdots (t_i - t_r)} . \tag{2}$$

Then (1) may be written in the form

$$y = \mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_r x_r . \tag{3}$$

In the coordinate system $(x_1 , \cdots , x_r , y)$, equation (3) represents an r-parameter $(\mu_1 , \cdots , \mu_r)$ family of planes passing through the origin. The method to be presented for constructing a confidence band for (1) is based on finding the envelope of this family, subject to a single restriction on the parameters. This method is a generalization of a similar method used by Hotelling and Working [3] to obtain a confidence band for a line of regression. An extension of their method to more general problems is given in [4].

The restriction that will be placed on the parameters $\mu_1 , \cdots , \mu_r$ is obtained by means of the generalized T function. For the variables $y_1 , y_2 , \cdots , y_k$, Hotelling's generalized T is defined by

$$T^2 = (n - 1) \sum_{i=1}^{k} \sum_{j=1}^{k} s^{ij} (\bar{y}_i - \mu_i)(\bar{y}_j - \mu_j), \tag{4}$$

where $(s^{ij}) = (s_{ij})^{-1}$ and where s_{ij} is the sample covariance,

$$s_{ij} = \frac{1}{n} \sum_{\alpha=1}^{n} (y_{i\alpha} - \bar{y}_i)(y_{j\alpha} - \bar{y}_j).$$

Under the assumptions made in the preceding section, a value of T^2 can be found, which will be denoted by T_0^2, such that

$$P\{T^2 \leqq T_0^2\} = C_0 , \tag{5}$$

where C_0 is a given number satisfying $0 < C_0 < 1$. The number C_0 will be the the lower bound for the confidence coefficient corresponding to the confidence band to be constructed. In terms of the preceding notation, the restriction that will be placed on the parameters of (3) is the restriction

$$T^2 \leqq T_0^2 . \tag{6}$$

From the remark made after (1), it follows that the parameters $\mu_{r+1} , \cdots , \mu_k$ can be expressed as linear combinations of $\mu_1 , \cdots , \mu_r$ and that therefore restriction (6) can be expressed as a restriction on $\mu_1 , \cdots , \mu_r$ only.

Now the technique for finding the envelope of an r-parameter family of surfaces such as (3), subject to a single restriction on those parameters such as $T^2 = T_0^2$, consists in first using the restriction to express (3) as an $(r - 1)$-parameter family of surfaces, and then eliminating those parameters between (3) and the $r - 1$ equations obtained by differentiating (3) with respect to those $r - 1$ parameters. But analytically this technique is equivalent to that employed in finding the maximum and minimum of the function $y = \mu_1 x_1 + \cdots + \mu_r x_r$ for fixed x's when the μ's are subject to restriction (6). The analysis will be carried out from the latter point of view with the aid of matrix algebra.

Let (4) for $T = T_0$ be written in the form

$$\sum_{1}^{k} \sum_{1}^{k} a_{ij} (\mu_i - \bar{y}_i)(\mu_j - \bar{y}_j) = \lambda_0 ,$$

where $a_{ij} = s^{ij}$ and $\lambda_0 = T_0^2/(n-1)$. If the parentheses are removed this equation assumes the form

$$\sum_1^k \sum_1^k a_{ij}\mu_i\mu_j - 2\sum_1^k a_i\mu_i + \lambda_1 = \lambda_0. \tag{7}$$

Let a denote the vector of a_i's, $\bar{y}$ the vector of $\bar{y}_i$'s, and A the matrix of a_{ij}'s. It will be seen that

$$a = A\bar{y}, \tag{8}$$

and that λ_1 is the quadratic form in $\bar{y}_1, \cdots, \bar{y}_r$ given by

$$\lambda_1 = \bar{y}'A\,\bar{y}. \tag{9}$$

Denoting $x_i(t_j)$ by x_{ij}, it follows from (1) that $\mu_j = \mu_1 x_{1j} + \cdots + \mu_r x_{rj}$ for $j = r+1, \cdots, k$. By means of these relations, (7) can be reduced to an equation in the parameters $\mu_1, \cdots, \mu_r$ only. This reduction can be accomplished by means of the transformation

$$\mu = B\nu, \tag{10}$$

where $\nu_i = \mu_i$ for $i = 1, \cdots, r$, and where B is the $k \times k$ matrix

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ x_{1r+1} & & \cdots & x_{rr+1} & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ x_{1k} & & \cdots & x_{rk} & 0 & \cdots & 0 \end{bmatrix}. \tag{11}$$

The reduction of (7) by means of (10) then proceeds as follows:

$$\begin{aligned} \mu'A\mu - 2a'\mu + \lambda_1 &= \lambda_0 \\ \nu'B'AB\nu - 2a'B\nu + \lambda_1 &= \lambda_0 \\ \nu'C\nu - 2c'\nu + \lambda_1 &= \lambda_0 \end{aligned} \tag{12}$$

where

$$C = B'AB \quad \text{and} \quad c = B'a. \tag{13}$$

Since the last $k - r$ columns of B consist of zero elements, the matrix C will contain zero elements in its last $k - r$ rows and columns, and the column vector c will have zeros for its last $k - r$ components. In summation notation, (12) will therefore assume the form

$$\sum_1^r \sum_1^r c_{ij}\mu_i\mu_j - 2\sum_1^r c_i\mu_i + \lambda_1 = \lambda_0, \tag{14}$$

since $\nu_i = \mu_i$ for $i = 1, \cdots, r$.

It will be convenient to express (14) in the form

$$\sum_1^r \sum_1^r c_{ij}(\mu_i - \alpha_i)(\mu_j - \alpha_j) = \lambda_2. \tag{15}$$

Expanding (15) and comparing with (14), it is readily observed that $\underline{c} = \underline{C}\alpha$, where α denotes the vector of α_i's, or that

$$\alpha = \underline{C}^{-1}\underline{c}. \tag{16}$$

Here $\underline{C}$ denotes the $r \times r$ matrix consisting of the first r rows and columns of C, and $\underline{c}$ denotes the r-dimensional column vector consisting of the first r components of c. It is assumed that $\underline{C}^{-1}$ exists. This assumption will be satisfied if (s_{ij}) is nonsingular, and the latter condition will be satisfied with probability one. The comparison of (14) and (15) also shows that $\lambda_2 = \lambda_0 - \lambda_1 + \alpha'\underline{C}\alpha$, which because of (9) and (16) can be written in the form

$$\lambda_2 = \lambda_0 - \bar{y}'A\bar{y} + \underline{c}'\underline{C}^{-1}\underline{c}. \tag{17}$$

Now return to the problem of maximizing the function $y = \mu_1 x_1 + \cdots + \mu_r x_r$ for fixed x's subject to the restriction (6), which, because of the preceding analysis leading to (15), is equivalent to the restriction

$$\sum_1^r \sum_1^r c_{ij}(\mu_i - \alpha_i)(\mu_j - \alpha_j) \leqq \lambda_2. \tag{18}$$

In the parameter space $\mu_1, \cdots, \mu_r$, the function y is the scalar product of the two vectors $(x_1, \cdots, x_r)$ and $(\mu_1, \cdots, \mu_r)$, whereas the restriction (18) states that the terminus of the vector $(\mu_1, \cdots, \mu_r)$ must lie inside or on the ellipsoid whose equation is given by (15). The scalar product is conveniently interpreted here as the length of the x vector multiplied by the projected length of the μ vector projected onto the x vector, together with the proper sign. The vector μ whose terminus must lie inside or on the ellipsoid (15) and which has maximum x-directed projected length is a vector whose terminus is a point on the ellipsoid where the tangent plane to the ellipsoid is perpendicular to the x vector. There will be two points of this type, one yielding a maximum and the other a minimum. The coordinates of these two points can be found as follows.

Let the equation of the ellipsoid (15) be written in the form $F = 0$, where

$$F = \sum_1^r \sum_1^r c_{ij}(\mu_i - \alpha_i)(\mu_j - \alpha_j) - \lambda_2.$$

Direction numbers for the normal to the tangent plane of the ellipsoid are given by the derivatives

$$F_{\mu_i} = 2[c_{i1}(\mu_1 - \alpha_1) + \cdots + c_{ir}(\mu_r - \alpha_r)], \qquad i = 1, \cdots, r.$$

If the tangent plane is to be perpendicular to the x vector, these direction numbers must be proportional to the components of the x vector; hence it is neces-

sary that $F_{\mu_i} = 2\,p_0 x_i$, where $2\,p_0$ is the constant of proportionality. In matrix notation this becomes $\mathcal{C}(\mu - \alpha) = p_0 x$. Hence

$$\mu - \alpha = p_0 \mathcal{C}^{-1} x. \tag{19}$$

The constant p_0 is determined by realizing that $\mu_1, \cdots, \mu_r$ must satisfy equation (15); hence

$$(p_0 \mathcal{C}^{-1} x)' \mathcal{C} (p_0 \mathcal{C}^{-1} x) = \lambda_2 .$$

Solving for p_0 yields

$$p_0 = \pm \sqrt{\lambda_2 / x' \mathcal{C}^{-1} x}. \tag{20}$$

The maximizing vector is the vector given by (19) when the positive value in (20) is selected. The maximum value of the function $y = \mu' x$ is therefore given by

$$y_{\max} = (\alpha + p_0 \mathcal{C}^{-1} x)'\, x = \alpha' x + p_0 x' \mathcal{C}^{-1} x = \alpha' x + \sqrt{\lambda_2}\, \sqrt{x' \mathcal{C}^{-1} x}.$$

Finally, if the values obtained in (16) and (17) are substituted, this will become

$$y_{\max} = \varphi' \mathcal{C}^{-1} x + \sqrt{\lambda_0 - \bar{y}' A \bar{y} + \varphi' \mathcal{C}^{-1} \varphi}\, \sqrt{' x \mathcal{C}^{-1} x}. \tag{21}$$

The minimum value of y is obtained by using the negative value in (20); hence

$$y_{\min} = \varphi' \mathcal{C}^{-1} x - \sqrt{\lambda_0 - \bar{y}' A \bar{y} + \varphi' \mathcal{C}^{-1} \varphi}\, \sqrt{x' \mathcal{C}^{-1} x}. \tag{22}$$

5. Interpretation. Since, by (2), the vector x has components that are polynomials in t, equations (22) and (21) define two curves in the t, y plane such that the curve (1) will lie between these two curves if restriction (6) is satisfied. From (5) the probability is at least C_0 that the mean growth curve (1) will lie between the curves (22) and (21). The latter probability would be exactly C_0 if only parameter points satisfying (6) could yield mean growth curves lying between curves (22) and (21); unfortunately, however, when $r \geqq 3$ there are parameter points not satisfying (6) which yield such curves. As a result, C_0 is only a lower bound for the confidence coefficient corresponding to the confidence band determined by (22) and (21).

The difficulty encountered in the preceding paragraph is best explained by considering a special case. Assume that $k = 4$, $r = 3$, and that the ellipsoid (15) is a sphere of radius $\frac{1}{2}$ with center at the origin. Now it follows from the definition of the x_i in (2) that $\sum_1^3 x_i = 1$; hence the x vector will have its terminus lying in the plane $\sum_1^3 \mu_i = 1$. As it varies, the terminus will describe a curve in this plane that is easily seen to be a parabola with vertex at the point (0, 1, 0) and passing through the points (1, 0, 0) and (0, 0, 1).

As the x vector describes the parabola, its intersection with the sphere will describe a curve on the sphere. This curve will be the locus of the maximizing points on the sphere, because the tangent plane at any point on this curve will be perpendicular to the x vector through that point. The curve on the sphere which is symmetrically opposite this curve will be the locus of the minimizing points. Now as t varies over its range of values, the tangent plane along the

maximizing curve and the tangent plane along the minimizing curve will generate a closed surface. Every point inside this surface will yield a value of the function $y = \mu_1 x_1 + \mu_2 x_2 + \mu_3 x_3$ that lies between $y_{\min}$ and $y_{\max}$ for all values of t.

The closed surface here will resemble, very roughly, the surface formed by two right circular cylinders, of diameters equal to the diameter of the sphere, which intersect at right angles. The actual confidence coefficient here would appear to be not appreciably larger than C_0, since most of the parameter points inside this surface will also lie inside the sphere. For more complicated situations, however, the relationship between C_0 and the actual confidence coefficient is unknown.

A minor difficulty with the method is that for some sample points the relationship between the μ's, given by $\mu_j = \mu_1 x_{1j} + \cdots + \mu_r x_{rj}$ $(j = r + 1, \cdots, k)$, will be inconsistent with restriction (6). Geometrically, this means that the random ellipsoid determined by (6) does not intersect all the planes determined by this relationship. For such sample points, the radical in (22) and (21) will be imaginary. The probability that this event will occur is undoubtedly quite small for most applications. As an illustration that is unrealistic but simple to compute, if n is large, $k = 5$, $r = 4$, and (6) is assumed to be a spherical restriction, it can be shown that the probability is less than .001 that inconsistency will occur when $C_0 = .95$. For larger values of C_0, such as $C_0 = .99$, the probability is extremely small that inconsistency will occur. Since nonintersection can occur only when the parameter point does not satisfy (6), the lower bound C_0 for the confidence coefficient still applies, provided one interprets imaginary confidence bands as bands incapable of covering any growth curve. If one excludes the nonintersection cases, the conditional probability of covering the mean growth curve will be slightly larger than the unconditional probability.

The choice of the first r points to determine the curve (1) was arbitrary. Investigations have not been made on how best to choose the points so that computations become simple, nor on how best to utilize the data. The problem of constructing a confidence band with known confidence coefficient by the method of this paper appears to be very difficult, if it is at all possible.

6. Illustration. The calculations involved in using formulas (21) and (22) will be illustrated by a simple example. Consider the problem of finding a 90 per cent confidence band (lower bound) for a parabola when ten individuals observed at each of four equally spaced time points yield $t_i = 0, 1, 2, 3$; $\bar{y}_i = 5.0$, 5.4, 6.0, 6.9; and

$$(s_{ij}) = \begin{bmatrix} 1 & -.4 & .3 & -.5 \\ -.4 & 1 & -.3 & .4 \\ .3 & -.3 & 1 & -.5 \\ -.5 & .4 & -.5 & 1 \end{bmatrix}.$$

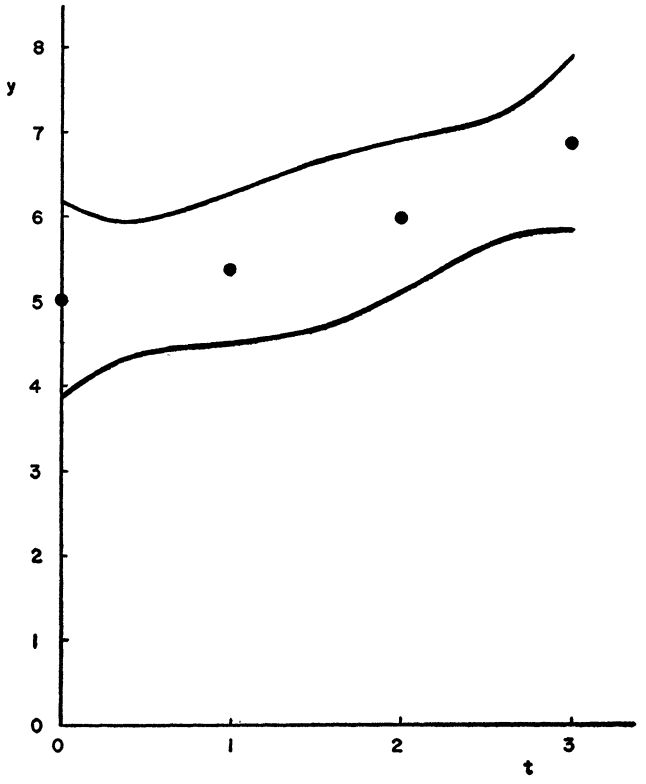

FIG 1.

Here $k = 4$, $r = 3$, $n = 10$, and $C_0 = .90$. Using (2), $x_1 = \frac{1}{2}(t^2 - 3t + 2)$, $x_2 = -t^2 + 2t$, $x_3 = \frac{1}{2}(t^2 - t)$. For $t = 3$ these give $x_{14} = 1$, $x_{24} = -3$, $x_{34} = 3$. These values enable one to write down B in (11). The inverse of (s_{ij}) yields

$$A = \begin{bmatrix} 1.426 & .333 & -.051 & .554 \\ .333 & 1.288 & .149 & -.274 \\ -.051 & .149 & 1.357 & .593 \\ .554 & -.274 & .593 & 1.683 \end{bmatrix}.$$

From (13) and (18) it follows that $c = B'A\,\bar{y}$. First $B'A$ is computed, then $c = B'A\,\bar{y}$ and $C = B'AB$ are computed. Next $\mathcal{C}^{-1}$ is computed, and then $\underline{c}'\mathcal{C}^{-1}$ and $\underline{c}'\mathcal{C}^{-1}\underline{c}$. Finally, $\bar{y}'A\,\bar{y}$ is computed. For this illustration, computations yielded the values

$$\underline{c}' = (28.906, -41.760, 62.167)$$

$$\underline{c}'\mathcal{C}^{-1} = (5.0099, 5.3875, 6.0130)$$

$$\underline{c}'\mathcal{C}^{-1}\underline{c} = 293.644 \qquad \bar{y}'A\,\bar{y} = 293.673.$$

The value of $\lambda_0 = T_0^2/(n-1)$ can be found by means of tables of the F distribution, using the relation $T_0^2 = F_0(n-1)k/(n-k)$, where $\nu_1 = k$ and $\nu_2 = n-k$, or by means of tables of the incomplete beta function, or by numerically solving the proper equation (see [1]). For this illustration λ_0 will be found to have the value $\lambda_0 = 2.121$. With the above computations completed, equations (21) and (22) can be written down in terms of the x's. If the x's are replaced by their expressions in terms of t, (21) and (22) reduce to

$$y = .124\,t^2 + .254\,t + 5.010 \pm \sqrt{(.174\,t^4 - 1.067\,t^3 + 2.109\,t^2 - 1.469\,t + .640)2.092}.$$

The graphs of these two curves, together with the values of the $\bar{y}_i$, are shown in Figure 1.

If the equations of the two curves determining the confidence band are not needed, the graphs can be constructed much faster by using equations (21) and (22) expressed in terms of the x's, rather than in terms of t, and calculating the x's corresponding to convenient t values. When $t = t_j$ and $j = 1, \cdots, r$, it follows from (2) that $x_i(t_j) = \delta_{ij}$ and hence that $x'\underline{C}^{-1}x$ reduces to the element in the jth row and jth column of $\underline{C}^{-1}$. The quantity $\underline{c}'\underline{C}^{-1}x$ then reduces to the jth component of the row vector $\underline{c}'\underline{C}^{-1}$. Although the computations are more difficult for $j > r$, they are still simple.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 407–409.

[2] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), p. 224.

[3] H. Hotelling and H. Working, "Application of the theory of error to the interpretation of trends," *J. Amer. Stat. Assoc., Suppl.*, Vol. 24 (1929), pp. 73–85.

[4] P. G. Hoel, "Confidence regions for linear regression," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 75–81.