ON A GENERAL SOLUTION FOR THE PARAMETERS OF ANY FUNCTION WITH APPLICATION TO THE THEORY OF ORGANIC GROWTH

By Harry Sylvester Will

Part I

I. **The Problem Stated.** A type of problem which continually arises in the ordinary course of statistical analysis is that of determining the numerical values of the parameters of a function used to represent a series of observational data. In mathematical terminology, the problem may be stated as follows:

Given, the observational series $Y_0, Y_1, \cdots Y_{n-1}$.

Assumed, the function $y = f(x, a, b, c, \cdots)$.

To find, the numerical values of the parameters $a, b, c, \cdots$.

If the function $f(x, a, b, c, \cdots)$ is linear in the parameters, the desired solution is easily obtained by familiar methods. In cases where the function is not linear, the standard procedure is to reduce it to the linear form by expansion into Taylor's series, thus:

$$f(x, a, b, c) = f(x, a_0b_0c_0) + f_a(x, a_0b_0c_0)\cdot\Delta a + f_b(x, a_0b_0c_0)\cdot\Delta b + f_c(x, a_0b_0c_0)\cdot\Delta c, \tag{1}$$

where $a = a_0 + \Delta a$, $b = b_0 + \Delta b$, $c = c_0 + \Delta c$.

The use of this method suffers from the excessive labor involved as the number of parameters to be determined increases. In cases where satisfactory values of the first approximations $a_0b_0c_0$ are not obtainable, the solution becomes impossible. The basic difficulty arises from the consideration that the Taylor theorem requires that the increments Δa, Δb, Δc shall be very small quantities.

A method of successive approximation which makes feasible the reduction of gross errors in the corrections will, I take it, be of considerable interest to mathematical statisticians. Let us, therefore, proceed to the development of a technique which accomplishes precisely this result.

II. **The Theta Technique.** Let us begin our development with the following restatement of the technical problem involved:

Given, the observational series $Y_0, Y_1, \cdots Y_{n-1}$.

Assumed, the function $y = f(x, (a_0 + \theta_1\Delta a), (b_0 + \theta_2\Delta b), (c_0 + \theta_3\Delta c))$.

To find, the values of $\theta_1, \theta_2, \theta_3$.

In this set of relations, a_0, b_0, c_0 and Δa, Δb, Δc are known quantities; while θ_1, θ_2 and θ_3 are each assumed not to exceed ± 1 in value. It follows, therefore, that the adjusted values of a, b, and c lie within the bounds $a_0 \pm \Delta a$, $b_0 \pm \Delta b$, $c_0 \pm \Delta c$. We may, then, write the following:

$$\begin{aligned} a_1 &= a_0 - \Delta a; & a_2 &= a_0 + \Delta a. \\ b_1 &= b_0 - \Delta b; & b_2 &= b_0 + \Delta b. \\ c_1 &= c_0 - \Delta c; & c_2 &= c_0 + \Delta c. \end{aligned} \tag{2}$$

The values of θ_1, θ_2 and θ_3 are determined by the following procedure:

First, form the function y from all possible combinations of a_1a_2, b_1b_2, c_1c_2, thus:

$$\begin{aligned} y_{111} &= f(x, a_1b_1c_1). \\ y_{112} &= f(x, a_1b_1c_2). \\ &\cdots\cdots\cdots\cdots \\ y_{222} &= f(x, a_2b_2c_2). \end{aligned} \tag{3}$$

In the case of p parameters, we can evidently form 2^p distinct sets of n values for the function y_{iii}. Since the assigned values of parameters are mere approximations to their true values, each computed set of values for the function y_{iii} will differ from the true values $y = f(x, abc)$.

Second, form the theoretical residuals $y_{iii} - y$, and then compute the corresponding standard errors of estimate σ_{iii}. There will, accordingly, be 2^p values of σ determined, each value being a measure of the error committed in assuming the corresponding approximations to parameters; thus, σ_{111} measures the errors committed in assuming the combination $a_1b_1c_1$; σ_{112} measures the errors committed in assuming $a_1b_1c_2$; $\cdots$; σ_{222} measures the errors committed in assuming $a_2b_2c_2$.

Third, taking the squared reciprocal of σ as a measure of the reliability of a given determination of y_{iii} from the parameters $a_ib_ic_i$, we may form the following comparative tests of the reliability of the 2^p sets of the values of y_{iii}, thus:

$$\begin{aligned} \omega_{111} &= \sigma_{111}^{-2} : (\sigma_{111}^{-2} + \sigma_{112}^{-2} + \cdots + \sigma_{222}^{-2}) = \sigma_{111}^{-2} : \textstyle\sum \sigma_{iii}^{-2}. \\ \omega_{112} &= \sigma_{112}^{-2} : \textstyle\sum \sigma_{iii}^{-2}. \\ &\cdots\cdots\cdots\cdots \\ \omega_{222} &= \sigma_{222}^{-2} : \textstyle\sum \sigma_{iii}^{-2}. \end{aligned} \tag{4}$$

Omega, we shall term the *test constant.* Obviously, $\Sigma\omega_{iii} = 1$.

Fourth, assuming three parameters, let us tabulate the possible subscripts of omega according to the following scheme:

$\omega(a_1)$	$\omega(a_2)$	$\omega(b_1)$	$\omega(b_2)$	$\omega(c_1)$	$\omega(c_2)$
111	211	111	121	111	112
121	221	211	221	211	212
112	212	112	122	121	122
122	222	212	222	221	222

In this table, the subscripts are in the order of abc; so that 111 denotes $\omega(a_1b_1c_1)$; 112 denotes $\omega(a_1b_1c_2)$; etc. Comparing columns $\omega(a_1)$ and $\omega(a_2)$, we observe that the bc subscripts are identical for both; while the a_1 subscripts of the first column are replaced by the a_2 subscripts in the second column. Again, comparing columns $\omega(b_1)$ and $\omega(b_2)$, we see that the ac subscripts are identical for both; while the b_1 subscripts of the one column are replaced by the b_2 subscripts in the other. Finally, comparing columns $\omega(c_1)$ and $\omega(c_2)$, we note that the ab scripts are identical for both; while the c_1 subscripts of the one column are replaced by c_2 subscripts in the other.

Fifth, let us form the column summations $\Sigma\omega(a_1)$, $\Sigma\omega(a_2)$; $\Sigma\omega(b_1)$, $\Sigma\omega(b_2)$; and $\Sigma\omega(c_1)$, $\Sigma\omega(c_2)$. Since the columns $\omega(a_1)$ and $\omega(a_2)$ differ only with respect to the α subscripts, the difference in value between the sums $\Sigma\omega(a_1)$ and $\Sigma\omega(a_2)$ can be due to differences in value between a_1 and a_2 only, and are not at all affected by differences in value between b_1b_2 and c_1c_2. $\Sigma\omega(a_1)$ and $\Sigma\omega(a_2)$ may, therefore, be regarded as the weights of a_1 and a_2 to be used in determining the adjusted value of a; for $\Sigma\omega(a_1) + \Sigma\omega(a_2) = 1$.

We may, then, write the following relations:

$$\begin{aligned} a &= \Sigma\omega(a_1)\cdot a_1 + \Sigma\omega(a_2)\cdot a_2 = \Sigma\omega(a_1)\cdot(a_0 - \Delta a) + \Sigma\omega(a_2)\cdot(a_0 + \Delta a) \\ &= (\Sigma\omega(a_1) + \Sigma\omega(a_2))\cdot a_0 + (\Sigma\omega(a_2) - \Sigma\omega(a_1))\cdot\Delta a = a_0 + \theta(a)\cdot\Delta a. \end{aligned} \tag{5}$$

Since precisely similar reasoning applies to the parameters b_1, b_2 and c_1, c_2, we have the following definitive formulas for computing the values of theta:

$$\begin{aligned} \theta(a) &= \Sigma\omega(a_2) - \Sigma\omega(a_1). \\ \theta(b) &= \Sigma\omega(b_2) - \Sigma\omega(b_1). \\ \theta(c) &= \Sigma\omega(c_2) - \Sigma\omega(c_1). \end{aligned} \tag{6}$$

As the adjusted values of parameters, we have:

$$\begin{aligned} a &= a_0 + \theta(a)\cdot\Delta a. \\ b &= b_0 + \theta(b)\cdot\Delta b. \\ c &= c_0 + \theta(c)\cdot\Delta c. \end{aligned} \tag{7}$$

In this development of the theta technique, we have determined σ_{iii} from the theoretical residuals $y_{iii} - y$. This has served well the purposes of exposition; but, since the true values of the function y are unknown, we must, in practice, compute σ_{iii} from the observational residuals $y_{iii} - Y$. Later in the memoir, it will be shown how the computation of θ may, in numerous cases, be considerably abridged.

Part II

III. The Principle of Malthus. Since a determination of the numerical parameters of a given function by means of the theta technique must, at best, involve a considerable amount of computation, I have chosen for purposes of demonstration a problem which is of much interest in itself. This problem, we shall state in the form of two questions:

First, what is the most appropriate mathematical form of the law of organic growth?

Second, how may the parameters of the indicated function be computed?

Thomas R. Malthus, in his famous essay on *The Principle of Population Growth* assumed that the proportional growth of human populations is properly defined by the differential equation,

$$\frac{1}{p} \cdot \frac{dp}{dt} = b, \tag{8}$$

where p is the population under consideration, t is the measure of time, and b is the stable or geometric rate of growth.

This formula has been destructively criticised on the ground that it fails wholly to give a mathematical description of the manner in which population growth is kept within bounds. So far as any implication of the formula is concerned, populations may grow to infinite magnitudes. An attempt to represent growth by its use must, therefore, result in a succession of discontinuities which are incompatible with the observed facts of organic growth.

IV. **The Symmetric Logistic.** In three memoirs published in 1838, 1845 and 1847, it was suggested by M. Verhulst, Professor of Mathematics in the Ecole Militaire in Brussels, that the rate of population growth might be stated as a function of the population itself. Assuming the limiting value of p to be H, this conception of the growth rate Verhulst expressed by the differential equation,

$$\frac{1}{p} \cdot \frac{dp}{dt} = -\, b(1 - pH^{-1}). \tag{9}$$

Since this equation expresses proportional growth as a linear function of p, it is the simplest relation of its kind that may be conceived. In representing the rate of growth as a quantity which approaches zero as the population approaches its limiting value, it makes, indeed, a significant advance over the Malthusian formula. Nevertheless, the equation is subject to an interesting limitation, the nature of which is made evident by an examination of the integral form of the function, namely:

$$p = H:[1 + e^{a+bt}]. \tag{10}$$

This we shall now prove to be rotationally symmetric with respect to the point of inflection.

Differentiating equation (9) a second time, we have,

$$\begin{aligned} d^2p &= -b\, dp[p(1 - H^{-1}p)]dt \\ &= p[p^{-2}dp^2 - b\, d^2t + bH^{-1}p\, d^2t + bH^{-1}dp\, dt] \\ &= p^{-1}dp^2 + bH^{-1}p\, dp\, dt. \end{aligned}$$

Hence,

$$\frac{d^2p}{dt^2} = b^2p(1 - H^{-1}p)^2 - b^2H^{-1}p^2(1 - H^{-1}p).$$

Setting $\frac{d^2p}{dt^2} = 0$, we get,

$$1 - 2H^{-1}p = 0.$$

Or

$$p = H/2, \tag{11}$$

which gives the value of p at the point of inflection.

Substituting for p from (10), and solving for t, we have,

$$t_i = -a/b, \tag{12}$$

where t_i is the point of inflection of the function p.

Denoting the magnitude of the population at time t_i by p_i, its magnitude at time t_{i+k} by p_{i+k}, and its magnitude at the time t_{i-k} by p_{i-k}, we have,

$$p_i = H:[1 + e^{a+b(-a/b)}] = H/2. \tag{13}$$

$$p_{i+k} = H:[1 + e^{a+b(t+k\Delta t)}] = H:[1 + e^{bk\Delta t}]. \tag{14}$$

$$p_{i-k} = H:[1 + e^{a+b(t-k\Delta t)}] = H:[1 + e^{-bk\Delta t}]. \tag{15}$$

Measuring p in units of H and setting $u = e^{bk\Delta t}$, we may rewrite these last three equations as follows:

$$H^{-1}p_i = 1/2.$$

$$H^{-1}p_{i+k} = 1:[1 + u].$$

$$H^{-1}p_{i-k} = 1:[1 + u^{-1}].$$

On the hypothesis of rotational symmetry, we have, by subtraction,

$$H^{-1}p_{i+k} - 1/2 = 1/2 - H^{-1}p_{i-k}.$$

In proof, we have:

$$\begin{aligned} 1:[1 + u] &= 1 - 1:[1 + u^{-1}] \\ &= u^{-1}:[1 + u^{-1}] \\ &= 1:[u + 1]. \end{aligned}$$

q. e. d.

Part III

V. **Criticisms of the Logistic.** Because of its symmetric form, many critics have called into question the finality of the logistic as a universal representation of population growth. That it applies in particular cases, they contend, is no reason for holding that it must apply in general. Professors Raymond Pearl and Lowell J. Reed of Johns Hopkins University—to whom we are indebted for the rediscovery of the earlier researches of Verhulst—have proposed, as the proper form of the generalized growth curve, the following function:

$$p = H:[1 + e^{a+bt+ct^2+dt^3}]. \tag{16}$$

In their view, this equation is suited not only to representing a single cycle of growth, but two successive cycles as well. This claim, however, must be rejected; for, if true, it would mean that one cycle of growth is predictable from another, a circumstance which is clearly inconsistent with the assumptions laid down by these same investigators.

Moreover, so far as I can learn from their published writings, these authors have never considered the implications of the differential form of the function they propose.

Differentiating (16), we have,

$$\frac{1}{p} \cdot \frac{dp}{dt} = - (b + 2cx + 3dx^2)(1 - H^{-1}p).$$

Here, we find the stable growth constant of Malthus replaced by an expression which is quadratic in t. This means that, for a population which is freed of a restraining limit, proportional growth tends generally toward infinite values. If there are any facts to support such a conception of organic growth, I do not know what they are, and must, perforce, reject the contention that equation (16) is the generalized form of the Verhulst function.

VI. **Fundamental Assumptions.** In order to represent the phenomenon of population growth mathematically, I hold the following assumptions to be necessary:

(a) Under favoring conditions, population may increase at a constant geometric rate.

(b) Under all circumstances, the rate of growth must be a finite and continuous quantity.

(c) The magnitude of a population is always a positive, real number.

(d) The growth of population tends toward restriction within definite bounds.

(e) The growth of population is a function of time.

(f) The basic conditions of growth are free of cataclysmic disturbances.

The first of these assumptions is given in recognition of well known facts concerning organic growth. The second is necessary because, even when the size of a population is freed of definite restriction, the pattern of growth is not necessarily geometric. The third assumption affirms the absurdity of representing a population as a negative or infinite quantity. The fourth merely asserts the indisputable fact that the organism must always grow in a finite environment. The fifth gives place to the concept of growth as the resultant of a complex of causes, no one of which can be isolated as an entirely independent variable. While the final assumption recognizes that major disturbing influences may profoundly affect the course of growth.

VII. **The Skew Logistic.** In accord with our fundamental assumptions, we may form the following differential equations:

$$\frac{1}{p'} \cdot \frac{dp'}{dt} = -\,[b + sm \cdot \cos\,(m(t+q))]\,[1 - H^{-1}p'] \qquad \text{Type } \alpha$$

$$= -\,[b + 2sm^2(t+q) : (1 + m^2(t+q)^2)]\,[1 - H^{-1}p'] \qquad \text{Type } \beta \quad (17)$$

$$= -\,[b + sm^2(t+q) : \sqrt{1 + m^2(t+q)^2}]\,[1 - H^{-1}p'] \qquad \text{Type } \gamma$$

In these equations, $p' = p - L$, and measures p from its lower limit as origin. On separating variables, the following integrations may be performed:

$$-\int [dp' : (p'(1 - H^{-1}p'))] = -\log\,[p' : (1 - H^{-1}p')] = \log\,[(H - p') : (Hp')].$$

Writing $z = m(t+q)$, $dz = mdt$; so that we have:

$$b\int dt + s\int \cos z\,dz = A + bt + s \cdot \sin z.$$

$$b\int dt + 2s\int [z : (1 + z^2)]\,dz = A + bt + s \cdot \log\,(1 + z^2).$$

$$b\int dt + s\int [z : \sqrt{1 + z^2}]\,dz = A + bt + s\sqrt{1 + z^2}.$$

From these integrals, we form the following equations:

$$\log\,[(H - p') : (Hp')] = A + bt + s \cdot \sin\,[m(t+q)].$$

$$\log\,[(H - p') : (Hp')] = A + bt + s \cdot \log\,[1 + m^2(t+q)^2].$$

$$\log\,[(H - p') : (Hp')] = A + bt + s \cdot \sqrt{1 + m^2(t+q)^2}.$$

We have, finally, on taking antilogarithms and making the substitutions $p = p' + L$, $a = A - \log H$:

$$p = L + H : [1 + e^{a + bt + s \cdot \sin(m(t+q))}]. \qquad \text{Type } \alpha$$

$$p = L + H : [1 + e^{a + bt + s \cdot \log(1 + m^2(t+q)^2)}]. \qquad \text{Type } \beta \quad (18)$$

$$p = L + H : [1 + e^{a + bt + s\sqrt{1 + m^2(t+q)^2}}]. \qquad \text{Type } \gamma$$

These equations give the normal forms of the skew logistic.

VIII. **Properties of the Skew Logistic.** We may deduce the properties of the skew logistic by examining both its differential and integral forms. Considering the derivative of Type α, we note that the Malthusian constant b is replaced by a trigonometric function whose amplitude is $b \pm sm$, and whose phase depends on the values of m and q. When $b \pm sm = 0$, the derivative must also equal zero, and a flat point in the curve of p is indicated. When b is absolutely less than sm, the derivative changes sign and the curve of p reverses its direction. Thus, the integral form of Type α modifies the symmetric form of the logistic by a succession of minor cycles in which the rate of growth is alternately accelerated and retarded.

Considering Type β, we find the Malthusian constant replaced by a function whose maximum and minimum values are attained when $t = m^{-1} - q$. Obviously, therefore, this function passes through a single period whose amplitude is $b \pm sm$, and whose phases are b, $b + sm$, b, $b - sm$, b. When $b \pm sm = 0$, a flat point in the curve of p is generated. The effect of skewness on the rate of growth passes through two double phases. Where b and s are of the same sign, these phases are: first, increasing retardation followed by decreasing retardation when $t + q$ is negative; and, second, increasing acceleration followed by decreasing acceleration when $t + q$ is positive. Where b and s are of opposite sign, the corresponding phases are: first, increasing acceleration followed by decreasing acceleration when $t + q$ is negative; and, second, increasing retardation followed by decreasing retardation when $t + q$ is positive. It is to be noted that, when sm is absolutely greater than b, the derivative will change sign twice before the upper limit is reached. Under these circumstances, the function p passes through a double reversal of direction.

Considering Type γ, we find the Malthusian constant of the derivative replaced by a function which is aperiodic and which approaches the limits $b \pm sm$ as t approaches $\pm \infty$. When b and s are of the same sign, skewness passes through the two following phases: first, the phase of decreasing retardation when $t + q$ is negative; and, second, the phase of increasing acceleration when $t + q$ is positive. On the other hand, when b and s are of opposite sign, the corresponding phases are: first, that of decreasing acceleration when $t + q$ is negative; and, second, that of increasing retardation when $t + q$ is positive. When sm is absolutely greater than b, the derivative changes sign, and the function p passes from a continuously increasing phase to a continuously decreasing phase, or *vice versa*.

In general, it may be said of all three types—α, β and γ—that, if the derivative is not restricted to a single change of sign, L denotes a lower asymptote of the function p; while, under the same conditions, H denotes the higher limit approached by the function $p - L$. When H is negative, the effect is to make L an upper, and $L - H$ a lower, asymptote of the curve p.

In the case of Type γ, when the function p makes a single change of sign, either H or L becomes a maximum (or minimum) value instead of an asymptote of the curve. In this event, it will be noted that the factor $1 - H^{-1}p$ appearing in the derivative does not approach zero as a limit with increasing values of t, but rather passes through a minimum and then approaches the limit 1 in either direction.

The parameter s may be positive or negative in sign, and is termed the index of skewness or, briefly, the *skewness* of the function. Obviously, m is always positive, and, since it determines the rate at which skewness develops, is properly termed the *development*. The point in time at which skewness passes from an accelerating to a retarding phase, or *vice versa*, is fixed by the value of q, which is, therefore, termed the *transition*. The parameter b, as has already been stated, is termed the stable growth tendency or, technically, the *stability* of the function. And since the position of the curve p on an arbitrary time scale will vary with the value of a, this parameter I have designated the *location*.

In all three types of the skew logistic, if $e^{\psi(t)}$ is a continuously decreasing function and both H and L are positive, the curve of p may be described as of the *rising hillside form*. In the case of Type γ, if the derivative changes from positive to negative sign, the curve may be described as *mountain formed*. If $e^{\psi(t)}$ increases continuously, the curve is of the *falling hillside* variety, except when the derivative of Type γ changes from negative to positive sign, in which event a *valley form* is generated.

Part IV

IX. **Parameters of the Symmetric Logistic.** The numerical parameters of the symmetric logistic (10) are most easily determined by the method of differences. First, we write,

$$p_i^{-1} = C + e^{A+bt}, \tag{19}$$

where $C = H^{-1}$; $A = a - \log H$; and $i = 0, 1, 2, \cdots n - 1$.

Assuming Δt constant, let us give to t the increment $k\Delta t$, thus:

$$p_{i+k}^{-1} = C + e^{A+b(t+k\Delta t)}. \tag{20}$$

Subtracting (19) from (20), we obtain

$$\Delta_k p_i^{-1} = e^{A+b(t+k\Delta t)} - e^{A+bt} = Be^{A+bt}, \tag{21}$$

where $B = e^{bk\Delta t} - 1$. The quantity $\Delta_k p_i^{-1} = p_{i+k}^{-1} - p_i^{-1}$ is termed a first order difference of rank k.

Giving to t in equation (21) the increment $k\Delta t$, we get

$$\Delta_k p_{i+k}^{-1} = Be^{A+b(t+k\Delta t)}. \tag{22}$$

Dividing (22) by (21), we have,

$$\Delta_k p_{i+k}^{-1} : \Delta_k p_i^{-1} = e^{bk\Delta t}.$$

Taking logarithms, we obtain

$$\Delta_k \log \Delta_k p_i^{-1} = \log \Delta_k p_{i+k}^{-1} - \log \Delta_k p_i^{-1} = bk\Delta t,$$

which defines the parameter b. We can form $n - 2k$ such equations. Hence,

b is uniquely determined by the relation

$$b = [\sum_{i=0}^{i=n-2k-1} \Delta_k \log \Delta_k P_i^{-1}] : [k(n - 2k)\Delta t]$$
$$= [\sum_{i=k}^{i=n-k-1} \log \Delta_k P_i^{-1} - \sum_{i=0}^{i=n-2k-1} \log \Delta_k P_i^{-1}] : [k(n + 2k)\Delta t], \quad (23)$$

where $k = n:3$ to the nearest integer.

Returning to (21), we have the following relation determining the value of A:

$$A = \log [\sum_{i=0}^{i=n-k-1} \Delta_k P_i^{-1}] - \log [B \sum_{i=0}^{i=n-k-1} e^{bt}]$$
$$= \log [\sum_{i=k}^{i=n-1} P_i^{-1} - \sum_{i=0}^{i=n-k-1} P_i^{-1}] - \log [B \sum_{i=0}^{i=n-k-1} e^{bt}], \quad (24)$$

where $k = n:2$ to the nearest integer.

From equation (19), we have

$$C = [\sum_{i=0}^{i=n-1} P_i^{-1} - e^A \sum_{i=0}^{i=n-1} e^{bt}] : n. \quad (25)$$

The values of H and a are, obviously, given by

$$H = C^{-1}. \quad (26)$$

$$a = A + \log H. \quad (27)$$

In the relations defining b, A and C, the values of P must be obtained from the observations. In computing the values of k, the formula is:

$$k = n(r + 1)^{-1},$$

where n is the number of observations, and r denotes the order of reduction involved in the defining relation.

In my first treatment of the subject, I assumed that the value of k for all orders of reduction might be determined from the reduction of highest order involved; but I have since found that I erred in this view. The point is that the function $\psi(p) = k^r(n - rk)$, discussed in the original memoir, must be maximized with respect to k separately for each order of difference involved; or, in other words, the rank constant k must be given a separate determination for each parameter defined if the most accurate results are to be obtained.

X. **Parameters of the Skew Logistic.** I shall now show how the method of differences may be used to abridge the computations involved in applying the theta technique to the determination of the parameters of the skew logistic. In this, as in the preceding section, we assume Δt constant.

Operating on Type γ of equation (18), we write

$$p_i = L + H : [1 + e^{a+bt+s\sqrt{1+m^2(t+q)^2}}]. \quad (28)$$

To begin with, let us write the transformation of ordinate

$$G = \log [H(p - L)^{-1} - 1].$$

Also, let us write

$$F = \sqrt{1 + m^2(t + q)^2}.$$

We may now rewrite equation (28) in the form

$$G_i = a + bt + sF_i. \tag{29}$$

Giving to t the increment $k\Delta t$, we have

$$G_{i+k} = a + b(t + k\Delta t) + sF_{i+k}. \tag{30}$$

Subtracting (29) from (30), we have,

$$\Delta_k G_i = bk\Delta t + s\Delta_k F_i. \tag{31}$$

Again giving to t the increment $k\Delta t$, we obtain

$$\Delta_k G_{i+k} = bk\Delta(t + k\Delta t) + s\Delta_k F_{i+k}. \tag{32}$$

Subtracting (31) from (32), we obtain

$$\Delta_k G_{i+k} - \Delta_k G_i = (bk\Delta t - bk\Delta t) + s(\Delta_k F_{i+k} - \Delta_k F_i),$$

or

$$\Delta_k^2 G_i = s\Delta_k^2 F_i. \tag{33}$$

We can form $n - 2k$ such equations, and may, therefore, form $n - 2k$ approximations to the value of the parameter s, as follows:

$$s_i = [\Delta_k^2 G_i]:[\Delta_k^2 F_i]; \qquad i = 0, 1, \cdots, n - 2k - 1.$$

Taking the mean value of the set s_i as its most probable value, we have,

$$s_0(HL\cdot mq) = \Sigma s_i:(n - 2k); k = n:3 \text{ to the nearest integer} \tag{34}$$

In this determination of s_0, the only parameters directly involved are H, L, m and q, the parameters a and b having been eliminated. By assigning values to H_0, L_0, m_0 and q_0, we may, on setting up the arbitrary corrections ΔH, ΔL, Δm and Δq, write down the following:

$$H_1 = H_0 - \Delta H; \quad H_2 = H_0 + \Delta H; \quad L_1 = L_0 - \Delta L; \quad L_2 = L_0 + \Delta L;$$
$$m_1 = m_0 - \Delta m; \quad m_2 = m_0 + \Delta m; \quad q_1 = q_0 - \Delta q; \quad q_2 = q_0 + \Delta q.$$

Since s_0 is a function of H, L, m and q, we may, by entering the subscripts of the combination $HL\cdot mq$, tabulate the possible determinations of s_0 as follows:

11·11	11·12	11·21	11·22
12·11	12·12	12·21	12·22
21·11	21·12	21·21	21·22
22·11	22·12	22·21	22·22

In this tabulation, the subscripts of parameters are in the order of $HL\cdot mq$; so that 12·21 denotes $s_0(H_1L_2\cdot m_2q_1)$, etc.

From the table, it is seen that we may compute $2^4 = 16$ distinct sets of approximations to $s_0(HL\cdot mq)$. Since the true values of H, L, m and q are unknown, each set of approximations s_i will show a characteristic variation about its mean value, s_0. This variation is most conveniently measured by the mean deviation

$$\epsilon = (s_0 - s_0')2N':N = (s_0'' - s_0)2N'':N, \tag{35}$$

where the second relation serves as a check on the computation by the first; $N = n - 2k$; N' denotes the number of items s_i which are *less* than s_0 in value, and N'', the number of items s_i which are *greater* than s_0 in value; while s_0' denotes the mean of the N' values of s_i which are less than s_0, and s_0'', the mean of the N'' values of s_i which are greater than s_0.

The reliability of a given value of s_0 as a measure of the central tendency of the corresponding set s_i is sufficiently determined by ϵ^{-2}, which serves at the same time to measure the reliability of the combination $HLmq$ figuring in the computation of the given set s_i. We may, therefore, compute the values of the test constant, ω, directly from the values of ϵ^{-2} by means of the relation,

$$\omega(HL\cdot mq) = \epsilon_j^{-2}:[\epsilon_{11\cdot 11}^{-2} + \epsilon_{11\cdot 12}^{-2} + \cdots + \epsilon_{22\cdot 22}^{-2}] = \epsilon_j^{-2}:\Sigma\epsilon^{-2}, \tag{36}$$

where $j = 11\cdot 11, 11\cdot 12, \cdots, 22\cdot 22; \Sigma\omega = 1$.

Since four values of theta are to be determined, we must arrange the sixteen values of omega in four ways, as shown by the following tabulation of subscripts:

$\omega(H_1)$	$\omega(H_2)$	$\omega(L_1)$	$\omega(L_2)$	$\omega(m_1)$	$\omega(m_2)$	$\omega(q_1)$	$\omega(q_2)$
11·11	21·11	11·11	12·11	11·11	11·21	11·11	11·12
11·12	21·12	11·12	12·12	11·12	11·22	11·21	11·22
11·21	21·21	11·21	12·21	12·11	12·21	12·11	12·12
11·22	21·22	11·22	12·22	12·12	12·22	12·21	12·22
12·11	22·11	21·11	22·11	21·11	21·21	21·11	21·12
12·12	22·12	21·12	22·12	21·12	21·22	21·21	21·22
12·21	22·21	21·21	22·21	22·11	22·21	22·11	22·12
12·22	22·22	21·22	22·22	22·12	22·22	22·21	22·22

Knowing the values of omega, we have at once,

$$\theta(H) = \Sigma\omega(H_2) - \Sigma\omega(H_1); \qquad \theta(L) = \Sigma\omega(L_2) - \Sigma\omega(L_1);$$

$$\theta(m) = \Sigma\omega(m_2) - \Sigma\omega(m_1); \qquad \theta(q) = \Sigma\omega(q_2) - \Sigma\omega(q_1). \tag{37}$$

$$H = H_0 + \theta(H)\cdot\Delta H; \qquad L = L_0 + \theta(L)\cdot\Delta L;$$

$$m = m_0 + \theta(m)\cdot\Delta m; \qquad q = q_0 + \theta(q)\cdot\Delta q. \tag{38}$$

The process of adjustment should be repeated until errors in the parameters diminish to negligible proportions.

With H, L, m and q known to a sufficient approximation, we may form anew the functions $G(H, L, m, q)$ and $F(H, L, m, q)$. We can then write $n - 2k$ equations of form (33), viz.:

$$\Delta_k^2 G_i = s\Delta_k^2 F_i.$$

Summing these equations, we have,

$$\Sigma\Delta_k^2 G_i = s\Sigma\Delta_k^2 F_i, \tag{39}$$

where $\Sigma\Delta_k^2 G_i = \sum_{i=2k}^{i=n-1} G_i - 2\sum_{i=k}^{i=n-k-1} G_i + \sum_{i=0}^{i=n-2k-1} G_i$;

$$\Sigma\Delta_k^2 F_i = \sum_{i=2k}^{i=n-1} F_i - 2\sum_{i=k}^{i=n-k-1} F_i + \sum_{i=0}^{i=n-2k-1} F_i;$$

where $k = n:3$ to the nearest integer.

The approximate value of s is now obtained from the relation

$$s = [\sum_{i=0}^{i=n-2k-1} \Delta_k^2 G_i]:[\sum_{i=0}^{i=n-2k-1} \Delta_k^2 F_i]. \tag{40}$$

Returning to equation (31), we solve for $bk\Delta t$, obtaining,

$$bk\Delta t = \Delta_k G_i - s\Delta_k F_i.$$

Since we can form $n - k$ such equations, the approximate value of b is given by the relation

$$b = [\Sigma\Delta_k G_i - s\Sigma\Delta_k F_i]:[k(n - k)\Delta t] \tag{41}$$

$$= [(\sum_{i=k}^{i=n-1} G_i - \sum_{i=0}^{i=n-k-1} G_i) - s(\sum_{i=k}^{i=n-1} F_i - \sum_{i=0}^{i=n-k-1} F_i)]:[k(n - k)\Delta t],$$

where $k = n:2$ to the nearest integer.

From equation (29), we obtain the approximate value of a as follows:

$$a = [\sum_{i=0}^{i=n-1} G_i - b\sum_{i=0}^{i=n-1} t - s\sum_{i=0}^{i=n-1} F_i]:n. \tag{42}$$

Comparing the abridged method of computing the values of theta here outlined with the general procedure of section II, it will be seen that we have been able to reduce the number of values of omega which it is necessary to determine from $2^7 = 128$ to $2^4 = 16$. In cases where L may be assumed to equal zero, the number of values of omega which must be computed is further reduced to $2^3 = 8$.

Part V

XI. Symmetric Parameters for the Population of the United States. I have determined the numerical values of the parameters of both the symmetric and the skew forms of the logistic from the population figures for the United States given by the Bureau of the Census. The only departure in the data from the census figures consists in the interpolation of all items to June 1st as the date of observation. The values of the symmetric parameters are computed from the data of Table I, as follows.

Setting $k = 15 \div 3 = 5$, we have, by equation (23),

$$\sum_0^4 \Delta_5 \log \Delta_5 P_i^{-1} = \sum_5^9 \log \Delta_5 P_i^{-1} - \sum_0^4 \log \Delta_5 P_i^{-1}$$

$$= \bar{9}.71878n - \bar{5}.14555n = -3.42677.$$

TABLE I

Data for the Symmetric Logistic

i	P^{-1}	$\Delta_5 P^{-1}$	$\log \Delta_5 P^{-1}$	10^{bt}
0	0.25582	−0.19724	$\bar{1}.29500_n$	1.00000
1	0.18939	−0.14627	$\bar{1}.16516_n$	0.72934
2	0.13885	−0.10704	$\bar{1}.02955_n$	0.53193
3	0.10431	−0.07838	$\bar{2}.89421_n$	0.38796
4	0.07770	−0.05776	$\bar{2}.76163_n$	0.28295
5	0.05858	−0.04269	$\bar{2}.63033_n$	0.20637
6	0.04312	−0.02996	$\bar{2}.47654_n$	0.15051
7	0.03181	−0.02098	$\bar{2}.32181_n$	0.10978
8	0.02593	−0.01650	$\bar{2}.21748_n$	0.08006
9	0.01994	−0.01182	$\bar{2}.07262_n$	0.05839
10	0.01589			0.04259
11	0.01316			0.03106
12	0.01083			0.02265
13	0.00943			0.01652
14	0.00812			0.01205
Σ	1.00288	−0.70864	$\overline{14}.86433_n$	3.66216

TABLE II(A)

Data for the Skew Logistic

i	G_1	G_2	F_{11}	F_{12}	F_{21}	F_{22}
0	+ 1.67998	+ 1.71132	6.47765	3.35261	9.65194	4.90306
1	+ 1.54968	+ 1.57779	5.68859	2.60000	8.45931	3.73631
2	+ 1.40690	+ 1.43878	4.90306	1.88680	7.26911	2.60000
3	+ 1.27698	+ 1.30927	4.12311	1.28062	6.08276	1.56205
4	+ 1.14130	+ 1.17416	3.35261	1.00000	4.90306	1.00000
5	+ 1.00816	+ 1.04179	2.60000	1.28062	3.73631	1.56205
6	+ 0.85948	+ 0.89428	1.88680	1.88680	2.60000	2.60000
7	+ 0.70540	+ 0.74193	1.28062	2.60000	1.56205	3.73631
8	+ 0.59699	+ 0.63515	1.00000	3.35261	1.00000	4.90306
9	+ 0.44841	+ 0.48956	1.28062	4.12311	1.56205	6.08276
10	+ 0.30840	+ 0.35346	1.88680	4.90306	2.60000	7.26911
11	+ 0.17992	+ 0.22981	2.60000	5.68859	3.73631	8.45931
12	+ 0.02885	+ 0.08647	3.35261	6.47765	4.90306	9.65194
13	− 0.09590	− 0.02968	4.12311	7.26911	6.08276	10.84620
14	− 0.25808	− 0.17670	4.90306	8.06226	7.26911	12.04159
Σ	+10.83647	+11.47739	49.45864	55.76384	71.41783	80.95375

TABLE II(B)

Data for the Skew Logistic

i	$\Delta_5^2 G_1$	$\Delta_5^2 G_2$	$\Delta_5^2 F_{11}$	$\Delta_5^2 F_{12}$	$\Delta_5^2 F_{21}$	$\Delta_5^2 F_{22}$
0	−0.02794	−0.01880	3.16445	5.69443	4.77932	9.04807
1	+0.01064	+0.01904	4.51499	4.51499	6.99562	6.99562
2	+0.02495	+0.04139	5.69443	3.16445	9.04807	4.77932
3	−0.01290	+0.00929	6.24622	1.84451	10.16552	2.60213
4	−0.01360	+0.01834	5.69443	0.81604	9.04807	0.87607
Σ	−0.01885	+0.06926	25.31452	16.03442	40.03660	24.30121

We note that $k(n - 2k)\Delta t = 5(15\text{–}10)1 = 25$; hence,

$$b = -3.42677 \div 25 = -0.1370708.$$

Next, set $k = 15 \div 2 = 7$, to the nearest integer; then, by equation (24), we get

$$\textstyle\sum_0^7 \Delta_7 P_i^{-1} = \sum_7^{14} P_i^{-1} - \sum_0^7 P_i^{-1} = 0.10330 - 0.86777 = -0.76447;$$

$$B = 10^{bk\Delta t} - 1 = 10^{-0.1370708\times 7} - 1 = -0.89022; \qquad \textstyle\sum_0^7 10^{bt} = 3.39884.$$

Hence,

$$A = \log[-0.76447] - \log[-0.89022 \times 3.39884] = \bar{1}.4025324.$$

We have next

$$\textstyle\sum_0^{14} P_i^{-1} = 1.00288; \qquad \sum_0^{14} 10^{bt} = 3.66216; \qquad 10^A = 0.25266.$$

By equation (25), then, we obtain

$$C = [1.00288 - 0.25266 \times 3.66216] \div 15 = 0.0051747.$$

By equation (26), we get

$$H = C^{-1} = 193.25.$$

Finally, by equation (27), we obtain

$$a = A + \log H = \bar{1}.4025324 + 2.2861136 = 1.68865.$$

The point of inflection of the curve is given by

$$t_i = -a{:}b = 1.68865 \div 0.1370708 = 12.319.$$

XII. **Skew Parameters for the Population of the United States.** Assuming $L = 0$, we form

$$H_1 = 198.0 - 7.0 = 191.0; \qquad H_2 = 198.0 + 7.0 = 205.0.$$

$$m_1 = 1.0 - 0.2 = 0.8; \qquad m_2 = 1.0 + 0.2 = 1.2.$$

$$q_1 = -6.0 - 2.0 = -8.0; \qquad q_2 = -6.0 + 2.0 = -4.0.$$

Next, the primary data of Tables II(a) and II(b) are computed. Setting $k = 15 \div 5 = 3$, $n - k$ values of the 2^3 sets of s_i are determined and entered in Table III(a). The values of s_0, ϵ and ω for each set are computed by equations (34), (35) and (36).

In Table III(b), the several values of ω are arranged according to their association: first, with H_1, H_2; second, with m_1, m_2; and, third, with q_1, q_2. The column sums yield the weights $\Sigma\omega$. The values of θ and the adjusted values of parameters are computed by equations (37) and (38):

TABLE III(A)

Data for the Computation of Θ

i	$s(1.11)$	$s(1.12)$	$s(1.21)$	$s(1.22)$	$s(2.11)$	$s(2.12)$	$s(2.21)$	$s(2.22)$
0	−0.00883	−0.00491	−0.00585	−0.00309	−0.00594	−0.00330	−0.00393	−0.00208
1	+0.00236	+0.00236	+0.00152	+0.00152	+0.00422	+0.00422	+0.00272	+0.00272
2	+0.00438	+0.00788	+0.00276	+0.00522	+0.00727	+0.01308	+0.00457	+0.00866
3	−0.00207	−0.00699	−0.00127	−0.00496	+0.00149	+0.00504	+0.00091	+0.00357
4	−0.00239	−0.01667	−0.00150	−0.01552	+0.00322	+0.02247	+0.00203	+0.02093
Σ	−0.00655	−0.01832	−0.00434	−0.01683	+0.01026	+0.04151	+0.00630	+0.03380
s_0	−0.00131	−0.00366	−0.00087	−0.00337	+0.00205	+0.00830	+0.00126	+0.00676
ϵ	+0.00374	+0.00703	+0.00241	+0.00550	+0.00342	+0.00404	+0.00222	+0.00643
ω	+0.10624	+0.03012	+0.25711	+0.04919	+0.12708	+0.09137	+0.00290	+0.03061

TABLE III(B)

Data for the Computation of Θ

	$\omega(h_1)$	$\omega(h_2)$	$\omega(m_1)$	$\omega(m_2)$	$\omega(q_1)$	$\omega(q_2)$
	0.1062	0.1271	0.1062	0.2571	0.1062	0.0301
	0.0301	0.0914	0.0301	0.0492	0.2571	0.0492
	0.2571	0.3029	0.1271	0.3029	0.1271	0.0914
	0.0492	0.0360	0.0914	0.0360	0.3029	0.0360
Σ	0.4426	0.5574	0.3548	0.6452	0.7933	0.2067

TABLE IV(A)

Summary of Adjustments

Parameter	Estimated Value	Δ	Θ	$\Delta\cdot\Theta$	Adjusted Value
H	+198.0	+7.0	+0.1148	+0.8036	+198.80
m	+ 1.0	+0.2	+0.2904	+0.05808	+1.05808
q	− 6.0	+2.0	−0.5866	−1.1732	−7.1732

TABLE IV(B)

Final Transformations

i	$G(Hmq)$	$F(Hmq)$
0	1.69772	7.65559
1	1.56410	6.60800
2	1.42495	5.46440
3	1.29526	4.52752
4	1.15991	3.50336
5	1.02722	2.50753
6	0.87921	1.59408
7	0.72613	1.01784
8	0.61866	1.32865
9	0.47182	2.17626
10	0.33408	3.15374
11	0.20842	4.17075
12	0.06189	5.20418
13	$\bar{1}$.94223	6.24580
14	$\bar{1}$.78913	7.29229
Σ	11.20073	62.54999

Finally, the functions G and F are formed anew from the adjusted values of H, m, q. The adjusted values of s, b and a are computed by equations (40), (41) and (42), as follows:

$$s = [\textstyle\sum_{10}^{14} G_i - 2\sum_5^9 G_i + \sum_0^4 G_i]:[\sum_{10}^{14} F_i - 2\sum_5^9 F_i + \sum_0^4 F_i]$$
$$= [0.33574 - 2 \times 3.72304 + 7.14194] \div [26.06676 - 2 \times 8.62436 + 27.85887]$$
$$= 0.03161 \div 36.67691 = 0.00086185.$$

$$b = [\textstyle\sum_8^{14} G_i - \sum_0^6 G_i - s(\sum_8^{14} F_i - \sum_0^6 F_i)]:[k(n - k)\Delta t]$$
$$= [1.42623 - 9.04837 - 0.00086185(29.57167 - 31.96048)] \div [7(15 - 7)1]$$
$$= [-7.62214 - 0.00086185 \times (-2.38881)] \div [56] = -0.13607.$$

$$a = [\textstyle\sum_0^{14} G_i - b\sum_0^{14} t - s\sum_0^{14} F_i]:n$$
$$= [11.20073 - (-0.13607 \times 105) - 0.00086185 \times 62.54999] \div 15 = 1.69561.$$

In the present case, the values of H_0, m_0 and q_0 were known within definite limits from previous experimentation. The values of the corrections, $\theta \cdot \Delta$, were, on this account, smaller than should ordinarily be expected from a first application of the technique. Always, it is necessary to take Δ sufficiently large to insure $\theta < 1$. As a preliminary step, it is not infrequently advantageous to compute trial values of ϵ by holding constant each two of the parameters H_0, m_0 and q_0 while experimenting roughly with the third.

TABLE V(A)

Ordinates of Fitted Curves

Year	Census Count	Symmetric Ordinates	Percentage Deviations	Skew Ordinates	Percentage Deviations
1790	3.909	3.88	−0.78	3.87	−0.01
1800	5.280	5.28	−0.03	5.27	−0.25
1810	7.202	7.16	−0.52	7.15	−0.73
1820	9.587	9.69	+1.07	9.67	+0.88
1830	12.866	13.04	+1.37	13.02	+1.20
1840	17.069	17.45	+2.22	17.42	+2.09
1850	23.192	23.15	−0.20	23.13	−0.28
1860	31.443	30.38	−3.36	30.37	−3.42
1870	38.558	39.36	+2.09	39.31	+1.95
1880	50.156	50.18	+0.05	50.07	−0.18
1890	62.948	62.61	−0.31	62.60	−0.55
1900	75.995	76.79	+1.05	76.64	+0.86
1910	92.329	91.76	−0.62	91.72	−0.67
1920	106.001	106.96	+0.90	107.16	+1.09
1930	123.068	121.66	−1.14	122.23	−0.69

TABLE V(B)

Extrapolations

Year	Forecast	Sym. O.	Sk. O.
1940	137.20	135.22	136.26
1950	149.29	147.18	148.78
1960	159.88	157.33	159.52
1970	168.71	165.66	168.42
1980	175.83	172.33	175.59
1990	181.46	177.52	181.25
2000	185.82	181.52	185.63
2010	189.14	184.55	188.98
2020	193.11	186.82	192.97
2030	193.54	188.52	193.40
2040	194.94	189.77	194.83
2050	195.98	190.72	195.87
2060	196.75	191.39	196.64
2070	197.31	191.88	197.22
2080	197.73	192.25	197.64
2090	198.03	192.52	197.94
2100	198.25		198.17
2110	198.42		198.34
2120	198.54		198.46
2130	198.63		198.55

Year	Sym. O.	Sk. O.
1780	2.844	2.850
1770	2.083	2.095
1760	1.523	1.539
1750	1.113	1.130
1740	0.813	0.829
1730	0.594	0.608
1720	0.434	0.445
1710	0.316	0.280
1700	0.231	0.238
1690	0.168	0.173
1680	0.123	0.127
1620	0.090	0.092
1610	0.065	0.067
1600	0.048	0.049
1590	0.035	0.036

Part VI

XIII. **General Considerations.** The technique of solution for the numerical values of parameters presented in the foregoing pages is generally applicable to continuous functions of real variables. The abridged procedure may be followed whenever the given function involves a component which is linear in certain of the parameters: for, in such cases, it is always possible to effect a transformation of ordinates which will permit of the elimination of the parameters of the linear component. In any event, the equation of the function may be solved for a single parameter which may then be employed, as in our illustration, as a means of determining the values of the test constant, omega.

XIV. **An Interpretation of Results.** The equations of the symmetric and skew logistic curves as computed for the population of the United States are, written to the natural base, as follows:

$$p = 193.25:[1 + e^{3.88826-0.31562t}].$$

$$p = 198.80:[1 + e^{3.90429-0.31331t+0.0019845\sqrt{1+1.0581^2(t-7.1732)^2}}]$$

The amount of skewness in the second of these equations, as measured by the value of s, is small; but, owing to the fair size of the parameter m, it develops rapidly and affects the form of the curve sensibly. The major effect is to raise the value of the limiting population as given in the first equation by about six millions and to prolong the period of growth by about forty years. The approximate limit of 193 millions in the symmetric form is reached about the year 2090; while the approximate limit of 199 millions of the skew form is not arrived at until about the year 2130.

The positive sign of s makes for a decreasing acceleration of the rate of increase during the earlier phases of growth and for an increasing retardation of this rate during the later phases, the value of q fixing the point of transition in the year 1861. This general epoch has often been cited by sociologists as marking the shift from a dominantly rural-agricultural civilization to a dominantly urban-industrial one. The point at which the change takes place has, to my knowledge, never before been defined mathematically.

Both curves fit the observations excellently, as shown by the percentage deviations of Table V(a). The forecasted growth presented in Table V(b) is based on the skew ordinates, the formula being

$$P_t = p_t(P_{14}/p_{14})^{1/(t-14)}, \tag{43}$$

where P denotes the actual population series, observed or predicted, and p, the skew ordinates. The assumptions of the formula are two: first, that it is the observed population P_{14} which initiates the forecasted series; and, second, that the influence of the correction factor P_{14}/p_{14} diminishes with the time.

The extrapolations of both the skew and symmetric formulas contrast with the results obtained by Doctors Dublin and Lotka, who predict a stationary population of 150 millions by 1970. For the same year, the ordinates of both the skew and symmetric curves exceed this figure, the one by 15.66, and the other by 18.42 millions.

The limit of 150 millions referred to was arrived at by analysis of current tendencies in birth and death rates. The argument is that current birth rates are spuriously high and current death rates spuriously low because of the abnormally high proportion of men and women in the reproductive ages. This circumstance is due, in part, to the influx in the past of immigrants from communities having a high normal birth rate, and, in part, to the high birth rates of preceding generations of parents in this country.

After computation of the necessary corrections has been made, the true rate of natural increase of the white population for the registration area of the United States for the year 1920 is seen to be only about 5.4 per thousand instead of the 10.7 per thousand indicated by the crude rates. For the year 1930, the actual rate of increase is 7.5 per thousand; while the corrected or true rate turns out to be virtually zero. Under the interpretation of the authorities cited, the spurious excess of births over deaths will be entirely dissipated by the year 1970, with the result of the stationary population predicted.

The hazard peculiar to this method of inference arises from two assumptions that are made: first, that the present collection and registration of vital statistics is sufficiently reliable to make precise estimate of the true rate of natural increase possible; second, that the tendencies of fecundity and mortality exhibited by current data are stable.

With respect to the first assumption, the authors have this to say:

"One factor of safety of unknown magnitude remains. There is still some degree of laxity in the registration of births, and the figures of the true rate of natural increase may, on that account, be somewhat larger than recorded above."

The caution of the authors in this statement is in contrast with the uncritical acceptance of their results by those who fail to grasp the implications of technique.

Concerning the second assumption, it may be pointed out that many of the tendencies exhibited by current data must be regarded as statistically reversible. Falling birth rates due to drift of population to cities, to postponement of marriage on the part of professional classes, to the increasing cost of child culture, to the ubanization of rural life and to the restriction of immigration may be definitely altered by reversals in tendency. The flow of population may move into extraurban and subrural districts, where birth rates are more favorable to increase. The cost of child culture may, in part, be socially assumed. Improvement in economic conditions may lessen the drain on the resources of the family. The tendency for rural birth rates to fall may be checked. Immigration may increase with improving economic conditions. Death rates may be further reduced in many age classes and for many causes.

In fine, when we attempt to project into the future the components that

(24)

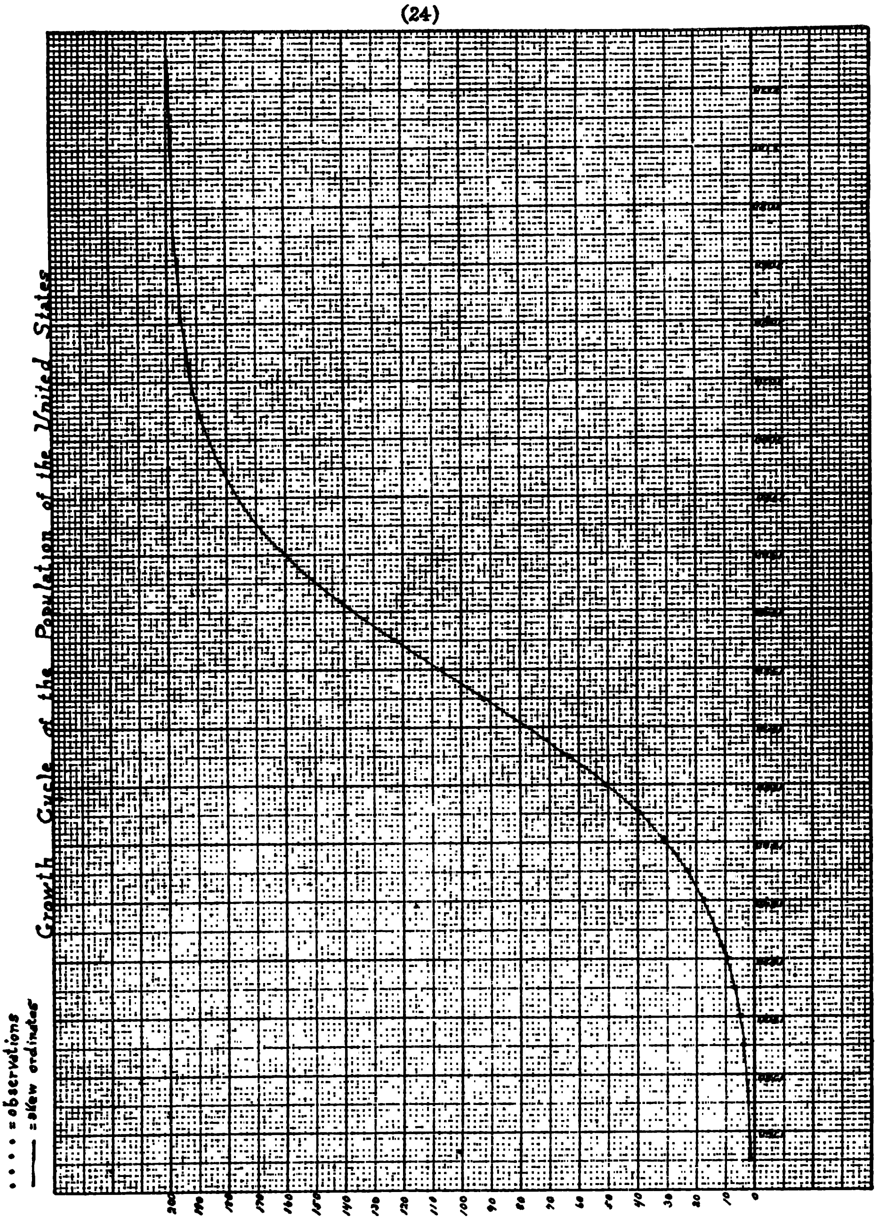
Growth Cycle of the Population of the United States
• • • • = observations
—— = skew ordinates
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

FIG. 1

determine the trend of natural increase, we encounter risks which vastly exceed those involved in the projection of the population series itself. Most of the data from which component trends must be determined cover but a brief period of time; while population data extends back for a century and a half. In this connection, it is not impertinent to inquire the criterion of relevance that will warrant a rejection of the items of the very series we are seeking to forecast.

It is a cardinal principle of logistic theory that the growth of population depends primarily on the continued supply of basic resources, physical and social, and that the dissipation of these resources is registered in the growth rate of the population itself. Any tendency of a population series toward skewness, that is, toward departure from the symmetric type of growth, is more likely to persist if it is systematic in character. The skew forms of the logistic function which we have developed permit us to measure any existing systematic tendency of the data toward skewness, and, therefore, to improve on the symmetric expectation of future growth.

In the case of the United States population, the evidence of skewness, insofar as it bears on the problem of expectation, is adverse to the conclusion that the ultimate limit of growth will be less than the symmetric asymptote. Conceding the light that the analysis of current tendencies may throw on the probable occurrence of future deviations from trend, the best criterion of long-time growth remains the logistic projection.

This statement, to be sure, does not relieve us of the necessity for recognizing the nature of the hazard that inheres in making a prediction from a trend extrapolation. The hazard involved in this type of inference arises from the assumption that the basic conditions of growth are stable, or, in other words, that the values of the parameters of the forecasting formula will remain substantially unchanged with the inclusion of new observations. Time alone can provide the final test of the continued validity of this assumption.

XV. **The Law of Organic Growth.** The law of organic growth in its most general form may be written:

$$p = L + H:[1 + e^{a+bt+s_1u_1+s_2u_2+s_3u_3}], \tag{44}$$

where $u_1 = \sin[m(t + q)]$; $u_2 = \log[1 + m^2(t + q)^2]$; $u_3 = \sqrt{1 + m^2(t + q)^2}$.

For most practical purposes, the evaluation of thirteen parameters is out of the question; hence, the restricted forms α, β, and γ, equation (18), will be the ones most generally employed.

I have made use of the term *law of organic growth* with reference to the logistic forms developed because I believe these functions to be the best means yet devised for the representation of the sequential changes which living organisms regularly manifest as individuals or societies. It states, in a quantitative form, all that is qualitatively implied by the so-called "law of diminishing returns" as this is commonly invoked by economists. The special sense in which I have used the term *law* may be expressed as follows:

A statistical law is a mathematical generalization on the behavior of a system of observations such that the implications of the formula are in accord with the assumptions basic to the phenomenon observed, and such that evaluations of the parameters of the formula determined from random samples are mutually consistent.

A statistical law, then, posits a system of relations manifesting itself in the form of observations which must be subjected to analysis before the true nature of their interrelations can be inferred. It expresses a probable, rather than a certain, inference; but, within the limitations of its claim to precision, it leaves reason no more free to reject its specification of reality than does a law of mechanics. Indeed, the point is still in dispute as to whether any law of science can be more than a statement of probabilities.

In contradistinction, the term *empirical formula* is properly restricted to cover the representation of the single set of observations at hand, and bears no necessary relation to any larger system. A sufficient test of an empirical formula is, therefore, the test of fit.

We may fit an indefinite number of formulas to a population series and obtain satisfactory results so far as agreement is concerned; but, on extrapolating, the same formulas will yield results that are patently absurd. The backward extrapolation for the population of the United States shown in Table V(b) represents the known facts as closely as could be expected when we take into consideration that census enumerations include aboriginal and immigrant populations as well as native born. Certainly, no random empirical formula, selected on the ground of goodness of fit, could be expected to yield as satisfactory a result.

Logistic theory does not, then, profess to guarantee infallibility of prediction. A population is not a mere aggregate of unrelated individuals inhabiting a restricted area, but a unified organization which grows by the utilization of total resources. When the supply of resources is profoundly disturbed or the basis of organizational unity destroyed, then the basis of prediction also is destroyed. And such reasoning is by no means peculiar to the sphere of social organization; for the integrity of any purely mechanical system is likewise conditioned by the assumption that the basis of coherence persists.

At this point, those in whom the speculative disposition is strong may query: if statistical prediction does not yield a certain result, is it, in the final analysis, superior to the ready and far less expensive method of guessing?

In answer, I can only say that, *a posteriori*, we can always, among a sufficiently large batch of guessers, find someone who has guessed well; but how, *a priori*, are we to know the good guesser from the poor? A population series consists of definite magnitudes, and any prediction of its development must result in the selection, out of a vast array of possible magnitudes, that which is most consistent with all the known facts. The gambler may elect to hazard his stake on the result of a random estimate; but the prudent will give heed to the exacting, if laborious, procedure of mathematical analysis.

ADDENDUM

Another solution of the theoretical problem stated in Section I may here be noted.

Given, as before, the function $y = f(x, a, b \cdots)$, we may, by assigning three approximate values to each parameter, compute 3^p sets of values for the function y, thus:

$$y_{11} = f(x, a_1b_1 \cdots); \qquad y_{12} = f(x, a_1b_2 \cdots); \qquad y_{13} = (x, a_1b_3 \cdots); \text{ etc.}$$

From the observations Y, we may compute 3^p sets of the residuals $y - Y$; and from these several sets of residuals, the corresponding standard errors of estimate, σ, may be computed for each set of values of the function y; thus, we have:

$$\sigma_{11} = \phi(Y, x, a_1b_1)$$

$$\sigma_{12} = \phi(Y, x, a_1b_2)$$

$$\sigma_{13} = \phi(Y, x, a_1b_3)$$

Restricting the parameters to a, b, and holding a constant, we observe that the values $\sigma_{1\,1}^2$, $\sigma_{1\,2}^2$, $\sigma_{1\,3}^2$ must vary with the assigned values of the parameter b, and take a minimum value when b takes its true or most probable value. As the errors in the approximation to b increase positively and negatively without limit, the computed values of σ^2 will tend toward the infinite. They may, therefore, be assumed to lie on the arc of a parabola whose equation is a quadratic function of xa_1b; hence, we may form the following equations of representation:

$$\sigma_{1\,1}^2 = k_{11} + l_{11}a_1 + m_{11}a_1^2.$$

$$\sigma_{1\,2}^2 = k_{12} + l_{12}a_1 + m_{12}a_1^2.$$

$$\sigma_{1\,3}^2 = k_{13} + l_{13}a_1 + m_{13}a_1^2.$$

By addition, we have,

$$\sigma_{1\,1}^2 + \sigma_{1\,2}^2 + \sigma_{1\,3}^2 = k_{11} + k_{12} + k_{13} + (l_{11} + l_{12} + l_{13})a_1 + (m_{11} + m_{12} + m_{13})a_1^2.$$

By appropriate variations in subscript, similar equations may be written in a_2 and a_3, thus:

$$\sigma_{2\,1}^2 + \sigma_{2\,2}^2 + \sigma_{2\,3}^2 = k_{21} + k_{22} + k_{23} + (l_{21} + l_{22} + l_{23})a_2 + (m_{21} + m_{22} + m_{23})a_2^2.$$

$$\sigma_{3\,1}^2 + \sigma_{3\,2}^2 + \sigma_{3\,3}^2 = k_{31} + k_{32} + k_{33} + (l_{31} + l_{32} + l_{33})a_3 + (m_{31} + m_{32} + m_{33})a_3^2.$$

These three equations are all of the quadratic form, and may be conveniently written as follows:

$$A_1 = K_1 + L_1a_1 + M_1a_1^2.$$

$$A_2 = K_1 + L_1a_2 + M_1a_2^2.$$

$$A_3 = K_1 + L_1a_3 + M_1a_3^2.$$

By precisely similar reasoning, the following equations in b may be developed:

$$B_1 = K_2 + L_2 b_1 + M_2 b_1^2.$$
$$B_2 = K_2 + L_2 b_2 + M_2 b_2^2.$$
$$B_3 = K_2 + L_2 b_3 + M_2 b_3^2,$$

where

$$B_1 = \sigma_{11}^2 + \sigma_{21}^2 + \sigma_{31}^2; \quad B_2 = \sigma_{12}^2 + \sigma_{22}^2 + \sigma_{32}^2; \quad B_3 = \sigma_{13}^2 + \sigma_{23}^2 + \sigma_{33}^2.$$

Since the values of a_1, a_2, a_3 and b_1, b_2, b_3 are assigned, the two sets of equations may each be simultaneously solved to obtain values for K_1, L_1, M_1 and K_2, L_2, M_2. To obtain the conditions for A = a minimum, B = a minimum, we differentiate with respect to a and b, as follows:

$$D_a(A) = L_1 + 2M_1 a; \qquad D_b(B) = L_2 + 2M_2 b.$$

Setting these two equations equal to zero and solving, we obtain the adjusted values of a and b, thus:

$$a = -L_1{:}2M_1; \qquad b = -L_2{:}2M_2.$$

The extension of this method to the case of p parameters is obvious. Assigning three approximations to each parameter, we hold constant a value of one parameter (say a_1), we form all possible combinations of subscripts for the remaining parameters ($b_1b_2b_3$ with $c_1c_2c_3$ with etc.). This will yield 3^{p-1} values of σ^2, each of which is associated with a_1. Repeating this process, we can form similar sets of values of σ^2 by association with a_2 and a_3. We can then form the sums $A_1 = \sigma(Yxa_1bc \cdots)$; $A_2 = \sigma(Yxa_2bc \cdots)$; $A_3 = \sigma(Yxa_3bc \cdots)$. In all, $3 \times 3^{p-1}$ or 3^p distinct determinations of σ^2 will be required. In like manner, the equations for B_1, B_2, B_3 and C_1, C_2, C_3, etc. are formed. The solutions for the adjusted values $a, b, c, \cdots$ follow directly.

Since the method of solution given in Part I requires the computation of but 2^p values of σ^2, it is evident that the method of this section is the more onerous when considering the determination of a single set of adjusted values of parameters, the excess being of the order $3^p{:}2^p = (1.5)^p$. However, being more precise, the present method will require fewer approximations to arrive at satisfactory values of the parameters sought. In other words, the mathematical advantage of economy lies with the theta technique; while the advantage of precision lies with the quadratic technique.

SELECTED BIBLIOGRAPHY

DUBLIN, L. I. AND LOTKA, A. J. On the True Rate of Natural Increase. J. A. S. A. S. 1925.

The True Rate of Natural Increase of the Population of the United States. Metron, Je. 1930.

HOTELLING, H. Differential Equations Subject to Error and Population Estimates. J. A. S. A. 1927.

HOTELLING, H. AND F. Causes of Birth Rate Fluctuations. J. A. S. A. 1931.
KNIBBS, G. H. Laws of Growth of a Population. J. A. S. A. 1926–27.
LEHFELDT, R. A. The Normal Law of Progress; J. R. S. S. 1916.
LOTKA, A. J. Studies on the Mode of Growth of Material Aggregates. A. J. Sci. 1907.
PEARL, R. AND REED, L. J. On the Rate of Growth of the Population of the United States since 1790 and Its Mathematical Representation. Proc. N. A. Sci. 1920.
On the Mathematical Theory of Population Growth. Metron, 1923.
WILL, H. S. On Fitting Curves to Observational Series by the Method of Differences; Ann. M. S. My. 1930.
WOLFE, A. B. Is there a Biological Law of Human Population Growth? Q. J. Ec. Ag. 1927.
YULE, G. U. The Growth of Population and the Factors which Control It. J. R. S. S. Jn. 1924.

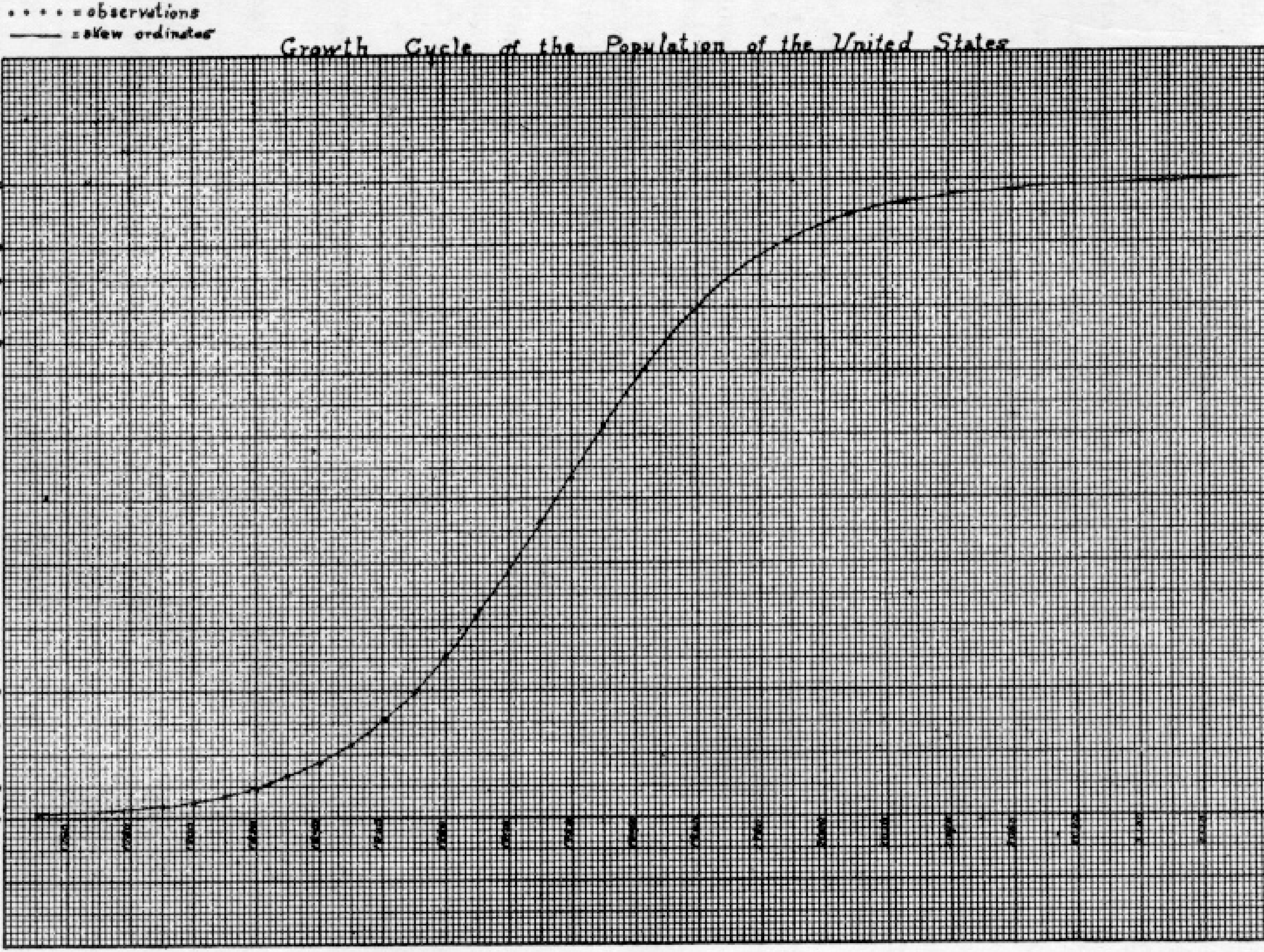
• • • • • = observations
——— = skew ordinates
Growth Cycle of the Population of the United States

FIG. 1