

THE PROGENY OF AN ENTIRE POPULATION¹

BY ALFRED J. LOTKA

Metropolitan Life Insurance Company

The literature on renewal theory has grown to considerable dimensions, until even admittedly incomplete bibliographies list over 100 titles. But a surprisingly small proportion of these publications exhibits any practical applications to concrete data, and such applications as have been made (e.g. by Wicksell, Hadwiger, Rhodes) are for the most part of restricted scope.

Anyone who has been following the development will, I think, feel that this is unfortunate. It has a double disadvantage. On the one hand the purely theoretical discussions emphasize difficulties which in practice may be relatively unimportant, being inherent either in some of the unrealistic *ad hoc* examples discussed, or in the expressions used to fit smooth curves to the basic data, rather than in these data themselves. On the other hand some real difficulties in application to actual data seem to require further clarification.

Several of the applications that have been made, including some of my own, are restricted to following up the "progeny" of a "population element" comprising only individuals all originating at the same time and therefore all of the same age (in the case of industrial equipment installation all made at one point of time). The analysis set forth in the treatment of this special case is competent also to deal with the practically more important case of the progeny of an initial population of given age distribution, though no example of this has hitherto been published.² Such an example will now be given, and at the same time this will afford an opportunity to clarify some points in the presentation of the more general case.

Let N_t be the total number of females at time t , and $c_t(a)$ the number comprised within the age limits a and $a + da$. Also, let $m_t(a)$ be the age-specific fertility of females of age a , counting daughters only. If α and ω are, respectively the lower and the upper limit of the female reproductive period, and $B(t)$ the annual births of females, then

$$(1) \quad B(t) = \int_{\alpha}^{\omega} N_t c_t(a) m_t(a) da.$$

However, it is not in this perfectly general form that the relation is to be applied. The case to be considered is that in which the "initial" population is throughout its "future" development, subject to constant age-specific fertility

¹ Compare A. J. Lotka, "The progeny of a population element," *Am. Jour. Hygiene*, Vol. 8 (1928), p. 875.

² An example was given by the writer in an oral communication to the Eighth American Scientific Congress, May 1940, the Proceedings of which have not so far been published.

and mortality. If we denote the "initial" time by $t = \omega$ (which we can do since the zero of time is arbitrary), we can then write

$$(2) \quad B(t) = \int_{\alpha}^{\omega} N_t c_t(a) m_{\omega}(a) da, \quad t > \omega.$$

Also, if $p_{t-a}(a)$ is the probability for a female born at time $\tau = t - a$ of surviving to time t , being then a years old, we have

$$(3) \quad B(t - a) p_{t-a}(a) = N_t c_t(a),$$

and, in particular, since in the case under consideration $p_{t-a}(a)$ is constant for $t - a > \omega$, i.e., for individuals born after $t = \omega$

$$(4) \quad B(t - a) p_{\omega}(a) = N_t c_t(a), \quad t > a + \omega.$$

Now, we have been at liberty for the "future" values of $m_t(a)$ and $p_{t-a}(a)$ to make the arbitrary assumption that they retain their values as of $t = \omega$ and $t - a > \omega$, respectively. But for the "past" of the system under consideration we do not have equal liberty, for any assumption we make must be compatible with

- (a) the initial age distribution
- (b) equation (1).

We can, however, within these limitations, assume that (4) still holds for $0 < t < \omega$, thus

$$(5) \quad B(t - a) p_{\omega}(a) = N_t c_t(a), \quad t > 0.$$

Introducing this in (1) we have

$$(6) \quad B(t) = \int_{\alpha}^{\omega} B(t - a) p_{\omega}(a) m_t(a) da, \quad t > 0.$$

But we *cannot* now, further assume that

$$(7) \quad m_t(a) = m_{\omega}(a), \quad t > 0,$$

for, in general, this would make (6) incompatible with (1).

We can, however, split the integral in (6) into two parts, thus

$$(8) \quad B(t) = \int_t^{\omega} B(t - a) p_{\omega}(a) m_t(a) da + \int_{\alpha}^t B(t - a) p_{\omega}(a) m_{\omega}(a) da,$$

with the assumption, *only in the range* $a < t$,

$$(9) \quad m_t(a) = m_{\omega}(a), \quad a < t.$$

Denoting the first integral in (8) by $F(t)$, and contracting $p_{\omega}(a) m_{\omega}(a)$ to $\varphi_{\omega}(a)$, we may write (8) in the form

$$(10) \quad B(t) = F(t) + \int_{\alpha}^t B(t - a) \varphi_{\omega}(a) da,$$

$$(11) \quad = F(t) + \beta(t),$$

with

$$(12) \quad \begin{cases} F(t) = 0 & t > \omega \\ F(t) = B(t) & 0 < t < \omega \end{cases}$$

and

$$(13) \quad B(t) = \int_{\alpha}^{\omega} B(t - a)\varphi_{\omega}(a) da, \quad t > \omega.^3$$

The assumption (9) has a definite physical meaning. The integral in (6) has been so split that the first part, $F(t)$, gives the births of daughters from mothers who themselves were born before $t = 0$, while the second part, $\beta(t)$, gives the births of daughters from mothers born after $t = 0$. Equation (9) therefore expresses the assumption that for mothers born at or after $t = 0$, the age-specific fertilities for ages $a < t$ have the same values $m_{\omega}(a)$, independent of t , as prevail for $t = \omega$. But at time t there are no mothers of age $a > t$, who were born after $t = 0$. Hence the assumption (9) can be quite simply stated to the effect that the age-specific fertilities $m_{\omega}(a)$ apply to all mothers born after time $t = 0$. This assumption cannot, in general be made for mothers born before $t = 0$, because it would not, in general, be compatible with the given initial age distribution and at the same time with assumption (5). Hence in the first integral of (8), denoted by $F(t)$ in (10), we must write $m_t(a)$, not $m_{\omega}(a)$.

Equation (10) is of the form discussed by G. Herglotz,⁴ who writes its solution, for $t > 0$, in the form of an exponential series.

$$(14) \quad B(t) = \sum Q_j e^{r_j t}$$

where the exponents r_j are the roots of the characteristic equation,

$$(15) \quad \Phi(r) = \int_{\alpha}^{\omega} e^{-ra} \varphi_{\omega}(a) da = 1,$$

while the coefficients Q_j are given by

$$(16) \quad Q_j = \frac{\int_0^{\omega} F(t) e^{-r_j t} dt}{\int_{\alpha}^{\omega} a e^{-ra} \varphi_{\omega}(a) da}.$$

There is only one real root of (14), since $\varphi_{\omega}(a) \geq 0$, for all values of a . For complex roots it is convenient to write the corresponding terms of the series (14) in trigonometric form

$$(17) \quad Q e^{rt} = 2U e^{ut} \cos vt - 2V e^{ut} \sin vt,$$

$$(18) \quad = 2\sqrt{(U^2 + V^2)} e^{ut} \cos (vt + \theta),$$

³ Since $\varphi_{\omega}(a) = 0$ for $a > \omega$.

⁴ *Math. Annalen*, Vol. 65 (1908), pp. 87 et seq.

where

$$\begin{aligned} \tan \theta &= V/U, \\ (19) \quad \cos \theta &= \frac{U}{\sqrt{U^2 + V^2}}, \\ \sin \theta &= \frac{V}{\sqrt{U^2 + V^2}}, \end{aligned}$$

and

$$(20) \quad U = \frac{RG + SH}{G^2 + H^2},$$

$$(21) \quad V = \frac{RH - SG}{G^2 + H^2},$$

in which

$$(22) \quad G = \int_{\alpha}^{\omega} a e^{-ua} \cos va \varphi_{\omega}(a) da,$$

$$(23) \quad H = \int_{\alpha}^{\omega} a e^{-ua} \sin va \varphi_{\omega}(a) da,$$

$$(24) \quad R = \int_0^{\omega} e^{-ut} \cos vt F(t) dt,$$

$$(25) \quad S = \int_0^{\omega} e^{-ut} \sin vt F(t) dt.$$

For purposes of numerical application to the problem here considered, we must express the annual births $B(t)$ for $t < \omega$ in terms of the given "initial" age distribution at time ω .

We have, generally

$$(26) \quad B(t - a) = \frac{N_t c_t(a)}{p_{t-a}(a)} = \frac{N_{\omega} c_{\omega}(a + \omega - t)}{p_{\omega}(a + \omega - t)},$$

since individuals of age a at time t , are $a + \omega - t$ years old at time ω .

Introducing the relation (26) in (10) we have

$$(27) \quad B(t) = F(t) + \int_{\alpha}^t \frac{N_{\omega} c_{\omega}(a + \omega - t)}{p_{\omega}(a + \omega - t)} p_{\omega}(a) m_{\omega}(a) da,$$

and

$$(28) \quad F(t) = B(t) - \int_{\alpha}^t \frac{N_{\omega} c_{\omega}(a + \omega - t)}{p_{\omega}(a + \omega - t)} \varphi_{\omega}(a) da,$$

$$(29) \quad = \frac{N_{\omega} c_{\omega}(\omega - t)}{p_{\omega}(\omega - t)} - \int_{\alpha}^t \frac{N_{\omega} c_{\omega}(a + \omega - t)}{p_{\omega}(a + \omega - t)} \varphi_{\omega}(a) da,$$

$$(11a) \quad = B(t) - \beta(t).$$

Note that, in computing the integral $\beta(t)$ for any particular value of t , the argument of the function c_ω runs from $\alpha + \omega - t$ to ω . Thus, for example, if the zero of time is 1865 and $t = \omega$ is at 1920, then, in computing $F(35)$, i.e., the value of F for 1900, the range of the argument of c_ω in the integral will be from $10 + 55 - 35$ to 55 , i.e., from 30 to 55.

Numerical Example. By way of a numerical illustration these principles will now be applied to a concrete case. We shall start with the age distribution of the white female population of the United States as constituted in 1920, for which previous publications furnish some of the required data, including the real root and the first three pairs of complex roots of the characteristic equation.

From this "initial" age distribution in 1920 it is necessary first of all to compute the auxiliary function $F(t)$ for the 55 years prior to 1920. The first term $B(t)$ in the right hand member of (28) is very easily computed for successive values of t from the relation (5a), which simply expresses the fact that persons a years old in the year ω , i.e., 1920, are the survivors of the $B(\omega - a)$ persons born in the year $\omega - a$.

$$(5a) \quad N_\omega c_\omega(a) = B(\omega - a)p_\omega(a).$$

In the diagram Fig. 1, which is drawn in stereographic projection, the age distribution of the (white female) population of the United States in 1920 is represented as plotted in a plane reaching forward at right angles to the plane of the paper. Successive points of $B(t)$ for $0 \leq t \leq \omega$, have been computed "by survivals" according to (5a) and plotted as a curve in the plane of the paper "at the back" of the diagram. The arrows indicate for a selected point, namely age 25 in 1920, the path of the computation according to equation (5a.)

The second term $\beta(t)$ in the expression (11a) for $F(t)$ was computed from the age distribution in 1920, the rates of survival from previous years into 1920,⁵ and the age-specific fertility at each age in the reproductive period, 10 to 55, on the basis of the relation (28). The results, for this second term in the expression for $F(t)$ computed for every fifth calendar year back of 1920 to 1875 and interpolated for intervening years,⁶ were also plotted as a curve in the rear plane of the diagram. The shaded area in the curve for the age distribution in 1920, and the arrows leading from this shaded area to the curve

$$(10, 11) \quad \beta(t) = \int_\alpha^t B(t - a)\varphi_\omega(a) da$$

$$(29, 11a) \quad = \int_\alpha^t \frac{N_\omega c_\omega(a + \omega - t)}{p_\omega(a + \omega - t)} \varphi_\omega(a) da,$$

indicate in this case the path of the computation according to equation (28).

⁵ Using the Foudray life table for white females in 1919-1920. In the first quinquennial age group, the following values were used:

$p(0.5) = .9460$	$p(2.5) = .9135$
$p(1.5) = .9235$	$p(3.5) = .9080$
	$p(4.5) = .9040$

⁶ This term vanishes for $t < 10$, i.e., back of 1875.

From these two curves, taking differences, the curve of $F(t) = B(t) - \beta(t)$ was plotted, as shown.

With the values of $F(t)$ thus obtained, we may proceed, by formulae (14) to (25), to compute values of $B(t)$ for all values of $t > 0$. So far as the period 1865 to 1920, corresponding to $0 < t < \omega$, is concerned, this merely means that we have an analytical expression to fit what is essentially a fundamental datum of the problem. For values of $t > \omega$ the formula gives us a continuation of the function $B(t)$ for all future time so long as the given age-specific fertility and mortality holds.

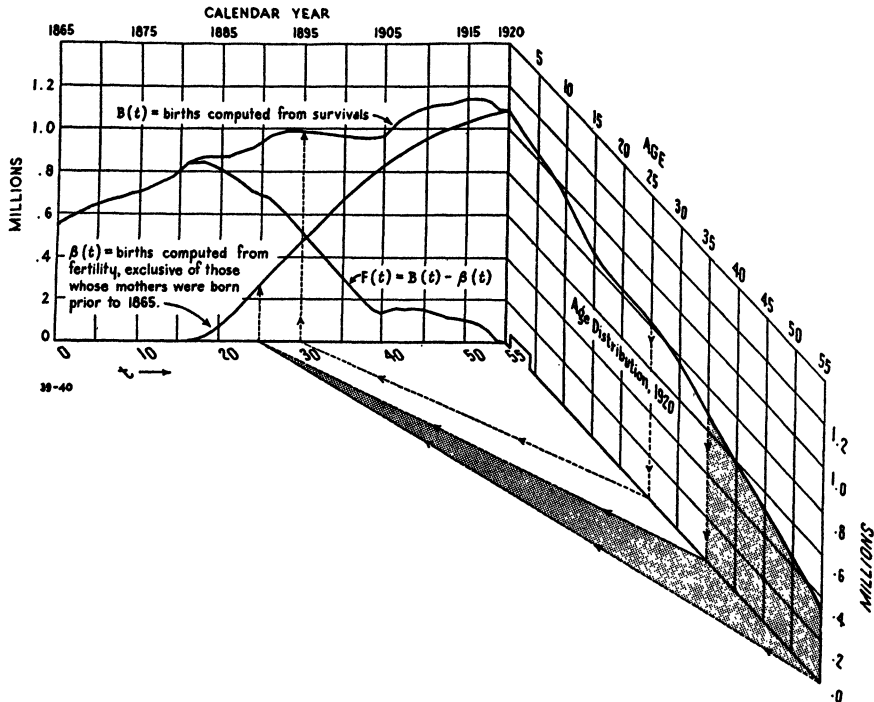


FIG. 1. Graph illustrating computation of auxiliary function $F(t)$ from "initial" age distribution.

The final results of this computation are exhibited in Figs. 2, 3 and 4. Of these, Fig. 2 exhibits the first, second and third oscillatory components for the period from 1890 forward. It will be seen that the waves are heavily damped, so that after a relatively short period the aperiodic component dominates the course of events.

Fig. 3 exhibits, for the years from 1865 to 1920, i.e., for the period $0 < t < \omega$, the aperiodic component (in a dashed line) and, as indicated by small circles, the sum of this component plus the three oscillatory components. It will be seen that from about 1890 forward the points so obtained follow rather closely the value $B(t)$ derived by survivals from the age distribution in 1920.

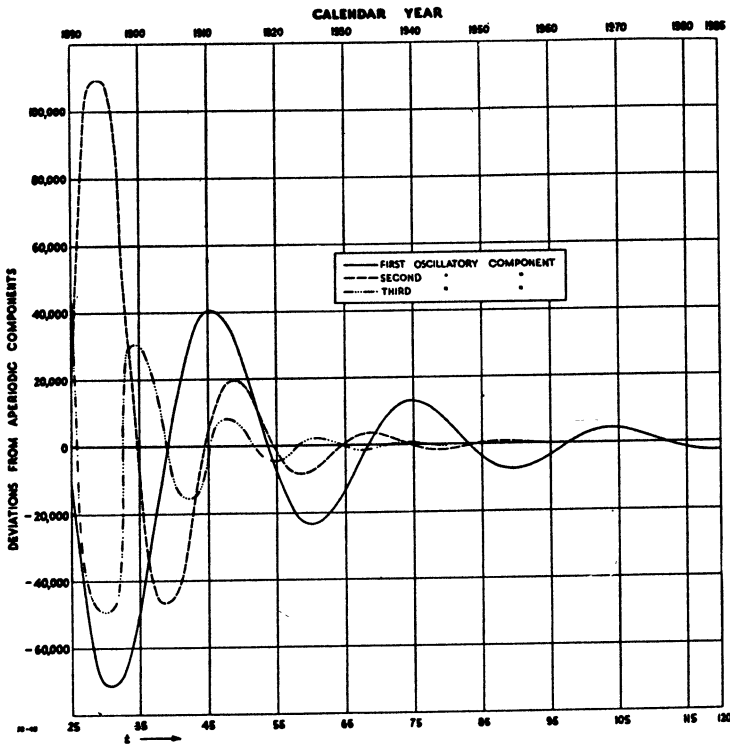


FIG. 2. First three oscillatory components of total annual births

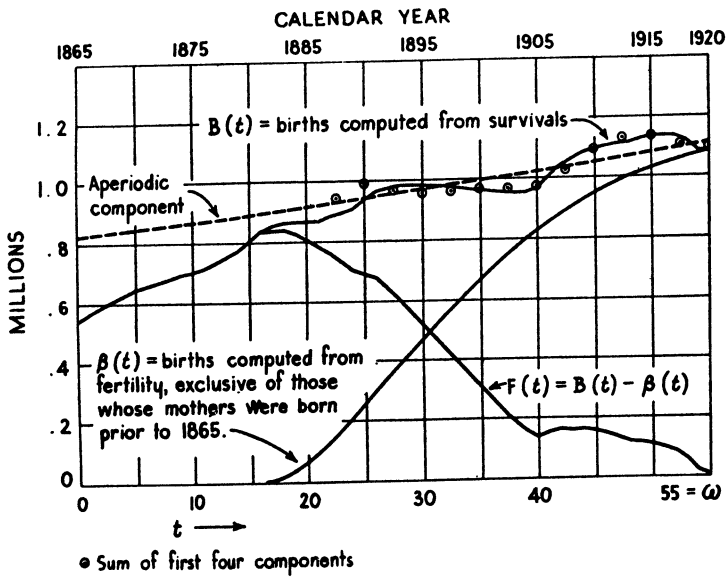


FIG. 3. Graph of functions $B(t)$, $\beta(t)$, and $F(t)$ for $0 < t < \omega$, i.e., for 1865 to 1920, together with aperiodic component; also, summation of aperiodic and first three oscillatory components.

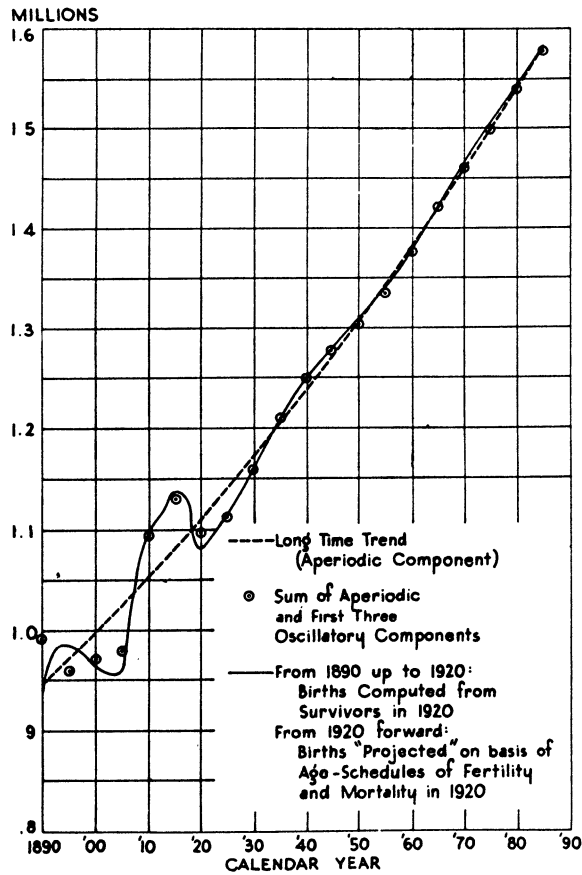


FIG. 4. Sum of aperiodic and three oscillatory terms of series solution compared with results of "step by step" computation of annual births.

TABLE I

Constants of the Series Solution (14) of Integral Equation (10) to Third Oscillatory Component Inclusive $t = 0$ at 1865

Function	Aperiodic Component	Oscillatory Components		
		First	Second	Third
u	$.543 \times 10^{-2}$	$-.386 \times 10^{-1}$	-8.731×10^{-2}	-9.804×10^{-2}
v	0	21.448×10^{-2}	31.542×10^{-2}	48.849×10^{-2}
G	28.226	25.768	51.225	37.008
H	0	14.938	-18.637	17.266
R	23.262×10^6	-17.863×10^5	-37.196×10^6	11.684×10^6
S	0	-31.508×10^5	16.827×10^6	-16.543×10^6
U	82.416×10^4	-10.494×10^4	-74.679×10^4	88.014×10^3
V	0	61.442×10^3	-56.787×10^3	48.808×10^4

Prior to about 1890, four components alone are quite inadequate, and the corresponding points have been omitted from the diagram. The lack of concordance, with such limited components, is inconsequential in this part of the series, since the purpose of this part of the work was merely to compute the auxiliary function $F(t)$, and the fit obtained for $B(t)$ in this range, so far as it goes, is merely a by-product, the main interest being in the course of $B(t)$ for $t > \omega$, i.e., in the years following 1920.

This course is charted in Fig. 4, in which the points obtained by the series solution (14) of (10) are again shown as small circles, while the fully drawn curve is derived from my previous publication "The Progressive Adjustment of Age Distribution to Fecundity."⁷ The annual births in that case were obtained "step by step" by computing age distributions by survivals for successive

TABLE II

United States White Female Population 1920, Observed; Also, the Same Projected Forward for Later Years⁸

Year	Population, thousands	Births, thousands	Birth rate per 1,000 per annum
1920	49,390	1,082	23.32
1930	51,727	1,162	22.46
1940	56,910	1,252	22.00
1950	61,639	1,307	21.20
1960	65,835	1,379	20.95
1970	69,829	1,465	20.98
1975	71,828	1,504	20.94
1980	73,850	1,543	20.89
1985	75,902	1,584	20.87

quinquennial periods, and applying to the reproductive age groups, in each case, the values of the reproductivity $m_{\omega}(a)$.

It will be seen that the points obtained by the solution (14) follow very closely those computed "step by step," although in the computation of the latter an approximation was made, using pivotal values of $p_{\omega}(a)$ for the several quinquennial age groups. A slight error introduced in this way would tend to be cumulative, and perhaps accounts for the fact that towards the end of the period covered (1985), the two sets of values diverge slightly. Even so, in 1985, the divergence is only about .4 percent.

The series solution has, of course, the advantage that it gives directly the result for any particular point of time, whereas the "step by step" method re-

⁷ *Jour. Washington Acad. Sci.*, Vol. 16 (1926), p. 505.

⁸ Calculated step by step from survival ratios and age specific fertilities, both held constant as of 1920 (reproduced for ready reference from *Jour. Wash. Acad. Sci.*, Vol. 16, p. 505).

quires the computation of the annual births for all intervening points in order to obtain the result for the chosen point of time.

Furthermore, the series tells us at once that the course of events is of the nature of a trend proceeding in geometric progression upon which are superposed a series of damped oscillations, of which the fundamental has a wave length equal approximately to the mean length of one generation from mother to daughter, i.e., about 28 years.

Alternative procedure. The procedure set forth in the preceding sections involves not only arbitrary assumptions regarding the values of $p(a)$ and $m(a)$ for "future" time, which are fundamental to the problem under consideration, but involves further incidental assumptions regarding their values prior to the "initial" condition at the instant denoted by $t = \omega$. These incidental assumptions are in a sense superfluous, since the future history of the system is completely determined by the initial age distribution and the assumed "future" values of $p(a)$ and $m(a)$. The additional assumptions were introduced merely for the purpose of translating the initial age distribution into a series of values of $B(t)$ for $0 < t < \omega$, i.e., prior to the given initial age distribution.

In actual fact the age distribution at time $t = \omega$ did not arise in the manner assumed; actually both $p(a)$ and $m(a)$ undoubtedly varied in the period 1865 to 1920, and migration also affected the situation. The quantity $F(t)$ introduced in equation (10) is, in fact, a purely auxiliary function having no direct relation to the biological events at time $t < \omega$.

An alternative procedure which would avoid these conflicts, and introduce assumptions only regarding "future" values of $p(a)$ and $m(a)$, would be to compute $B(t)$ step by step over the period from $B(1920)$ to $B(1920 + \omega) = B(1975)$.

Placing the zero of time $t = 1920$ this would give $B(t)$ for $0 < t < \omega$. For $t > \omega$ we should have, simply

$$B(t) = \int_{\alpha}^{\omega} B(t - a)\varphi_{1920}(a) da, \quad t > \omega,$$

using, in the evaluation of the integral, the values of $B(t - a)$ obtained by the step by step process.

We could here also split the integral into two parts

$$\begin{aligned} B(t) &= \int_t^{\omega} B(t - a)\varphi_{1920}(a) da + \int_{\alpha}^t B(t - a)\varphi_{1920}(a) da \\ &= F(t) + \int_{\alpha}^t B(t - a)\varphi_{1920}(a) da. \end{aligned}$$

But the function $\varphi_{1920}(a)$ is now the same in the two integrals, and there is no occasion, in this case, for distinguishing the two parts of the integral.

If this procedure is adopted, its application to the course of $B(t)$ for $t > \omega$,

i.e. beyond 1975, is of minor interest, for by that time it has practically settled down to the aperiodic (exponential) component, the oscillations being greatly damped down. The major interest in the result of a computation carried out by this procedure would be in the fitting of a series of the form (14) to the function $B(t)$ in the range 1920 to 1975, which, in this setting, figures as a known "arbitrary" function.

Of the two alternative procedures the one carried out in detail in the text and the numerical example is of greater interest, as exhibiting in greater generality the application of the Hertz-Herglotz solution.

BIBLIOGRAPHY

A few more titles may be added to previously published bibliographies (see these Annals, Vol. 10 (1939), p. 22; Vol. 12 (1941), p. 266).

- 1935 W. MÖSCHLER, "Untersuchungen über Eintrittsgewinn und Fehlbetrag einer Versicherungskasse," *Bull. de l'Association des Actuaires Suisses*, October (1935), p. 129.
- 1936 ALI AFZALIPOUR, *Contribution à l'étude de la théorie mathématique de la démographie*, Thèse, 1936.
- 1938 H. HADWIGER, "Ein Konvergenzkriterium für Erneuerungszahlen," *Skandinavisk Aktuarietidskrift*, Vol. 3-4, p. 226.
- H. HESS, *Anwendung der logistischen Funktion in der mathematischen Bevölkerungstheorie*, Inaugural Dissertation der philosophischen Fakultät der Universität Bern.
- W. M. DAWSON and W. EDWARDS DEMING, "On the problem of natural increase," *Growth*, Vol. 2, p. 319.
- 1940 A. W. BROWN, "A note on the use of a Pearson Type III function in renewal theory," *Annals of Math. Stat.*, Vol. 11 (1940), p. 448.
- 1940 E. ZWINGGI, "Entwicklung von Personengesamtheiten—Zusammenfassender Bericht," *Twelfth International Congress of Actuaries*, Vol. 3, p. 263.
- 1940 E. KEINANEN, "Über die Altersverteilung der Bevölkerung," *Twelfth International Congress of Actuaries*, Vol. 3, p. 305.
- 1940 S. S. TOWNSEND, "Some observations on the internal variation in groups of lives assured in industrial assurance and in the general population of England and Wales," *Twelfth International Congress of Actuaries*, Vol. 3, p. 319.
- 1940 R. TARJAN, "Untersuchungen über den Kapitalbedarf des Lebensversicherungsgeschäftes," *Twelfth International Congress of Actuaries*, Vol. 3, p. 335.
- 1940 M. PRESBURGER, "Sur l'étude générale des collectivités de personnes," *Twelfth International Congress of Actuaries*, Vol. 3, p. 353.
- 1940 A. MARET, "Direkte Berechnung der Vorgangsfunktionen einer offenen Gesamtheit," *Twelfth International Congress of Actuaries*, Vol. 3, p. 387.
- 1940 E. ZWINGGI, "Über Zusammenhänge zwischen der technischen Stabilität einer Sozialversicherungskasse und der Entwicklungsformel für den Versichertenbestand," *Twelfth International Congress of Actuaries*, Vol. 3, p. 395.
- 1940 H. HADWIGER and W. WEGMÜLLER, "Entwicklung und Umsichtung von Personengesamtheiten," *Twelfth International Congress of Actuaries*, p. 369.
- 1940 W. DOBBERNACK and G. TIETZ, "Die Entwicklung von Personengesamtheiten vom Standpunkt der Sozialversicherungstechnik," "Zwölfter Internationaler Kongress Der Versicherungs-Mathematiker," Vol. 4, p. 233.
- 1941 R. C. GEARY, "Irish population prospects considered from the viewpoint of reproduction rates," *Statistical and Social Inquiry Society of Ireland*, 1941.

- 1941 H. HADWIGER, "Eine Formel der mathematischen Bevölkerungstheorie," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Vol. 41, p. 67.
- L. FERAUD, "Le renouvellement, quelques problèmes connexes et les equations intégrales de cycle fermé," *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, Vol. 41, p. 81.
- W. FELLER, "On the integral equation of the renewal theory," *Annals of Math. Stat.*, Vol. 12 (1941), p. 243.
- HARRO BERNARDELLI, "Population waves," *Journal of the Burma Research Society*, Vol. 31.