Riv. Mat. Univ. Parma (4) 9 (1983), 213-222

G. Di Cola and F. Nicoli (*)

Parameter estimation in age-structured population dynamics (**)

Introduction

One of more complex problems in population dynamics is that of birth-death parameters estimation from demographic data, taking into account the age structure of the population.

In many natural populations the chronological age is not or hardly observable, therefore a substitute of age seems to be necessary. Starting from the Lotka equation we propose a lumped mathematical model which takes into account the chronological or physiological age by means of developmental stages or dimensional classes.

The main dynamical processes of the population are described in term of integral equations.

The parameter estimation is obtained from a constrained minimum problem.

The corresponding inverse problem is solved by means of statistical regularization.

1 - Mathematical model

We assume that the population under consideration is large enough so that is meaningful the concept of a population density function, that is its size can be represented as a continuous function, of age and time, with continuous derivatives.

(*) Indirizzo: Istituto di Matematica, Via Università 12, 43100 Parma, Italy.
(**) Presented of the Meeting « Stochastic Methods in Life Sciences: general aspects and specific models », Trento, september 1980. — Ricevuto: 7-IX-1981.

Moreover we consider only female individuals in the population, assuming that similar considerations can be made for populations comprising both sexes when the sex ratio is known.

Let $n(a, t)$ be the age density function of the population, that is $n(a, t)\,\Delta a$ is the number of individuals of the population whose age lies between a and $a + \Delta a$ at time t. Thus the total population at time t can be expressed as

$$N(t) = \int_0^{+\infty} n(a, t)\, \mathrm{d}a\,. \tag{1}$$

The basic assessment of our population dynamics problem refers to the Von Foerster equation (Von Foerster, 1959)

$$\frac{\partial n(a, t)}{\partial a} + \frac{\partial n(a, t)}{\partial t} = -\,\mu(a, t)\, n(a, t) \qquad (a, t > 0)\,, \tag{2}$$

with initial and boundary conditions

$$n(a, 0) = \varphi(a) \qquad (a > 0)\,, \tag{3}$$

$$n(0, t) = \int_0^{\infty} \lambda(a, t)\, n(a, t)\, \mathrm{d}a \qquad (t > 0)\,, \tag{4}$$

where, $\mu(a, t)$ is the specific death rate, $\lambda(a, t)$ is the specific fecundity rate.

If we assume μ and λ to be time independent functions the solution of the equation (2) is given by

$$n(a, t) = n(0, t - a)\, S(a) \qquad (t > a)\,, \tag{5}$$

where $S(a)$ is the survival function related to the specific mortality rate by the equation

$$\frac{\mathrm{d}S}{\mathrm{d}a} = -\,S(a)\cdot\mu(a)\,, \qquad \text{with} \qquad S(0) = 1\,. \tag{6}$$

The renewal equation (4) for large t becomes

$$b(t) \simeq \int_0^{t} \lambda(a)\, b(t - a)\, S(a)\, \mathrm{d}a\,, \tag{7}$$

where $b(t) = n(0, t)$ is the birth rate.

The equation (7) is known as Lotka equation; if the product $\lambda(a)\,S(a)$ is known, through experimental observations (e.g. by means of life tables), then the birth rate can be estimated by solving eq. (7).

If an exponential growth of population and a stable age structure are assumed, then it is possible to estimate the malthusian parameter r by solving the equation

$$1 = \int_0^t \lambda(a)\,S(a)\,e^{ra}\,da\,,$$

which follows from (7) when $b(t) = e^{rt}$ is assumed.

In general when we deal with populations in which it is not possible to follow the cohorts it is difficult to estimate the survival and fecundity rates, therefore it is of some interest to solve the inverse problem related to the equation (7) starting from experimental observations of the birth rate $b(t)$.

However in many natural populations with continuous recruitment the observation of age structure is practically impossible. When the chronological or physiological age is not observable, the subdivision of the population in age classes or developmental stages seems to be necessary.

In this work we refer to a population in which a subdivision into k developmental stages may be possible, the last stage being the class of the adults.

Let $n_i(a, t)$ and T_i, $i = 1, \ldots, k$ be the specific density and the developmental time of the i-th developmental stage respectively (T_k is the maximum life lenght of the adults).

The following continuity equations hold

$$\frac{\partial n_i(a,t)}{\partial t} + \frac{\partial n_i(a,t)}{\partial a} = -\mu_i(a)\,n_i(a,t) \qquad (0 \leqslant a \leqslant T_i;\ i = 1, \ldots, k)\,, \tag{8}$$

with initial and boundary conditions

$$n_i(a, 0) = n_i^0(a) \qquad (i = 1, \ldots, k)\,, \tag{9$_1$}$$

$$n_1(0, t) = \int_0^\infty \lambda(a)\,n_k(a, t)\,da\,, \tag{9$_2$}$$

$$n_{i+1}(0, t) = n_i(T_i, t) \qquad \big(i = 1, \ldots, (k-1)\big)\,. \tag{9$_3$}$$

If we indicate with $N_i(t)$ the total number of individuals in the i-th stage, that is

$$N_i(t) = \int_0^{T_i} n_i(a, t)\,da \qquad (i = 1, \ldots, k)\,, \tag{10}$$

we can get the following set of integral equations which describe adequately the dynamics of the population for $t > \sum_{i=1}^{k-1} T_i$ and λ as a constant

$$N_i(t) = \int_{\tau_i}^{\tau_{i+1}} f_i(a)\, N_k(t-a)\, \mathrm{d}a \qquad (i = 1, \ldots, k)\,, \tag{11}$$

where

$$\tau_1 = 0\,, \qquad \tau_{k+1} = t\,, \qquad \tau_i = \sum_{j=1}^{i-1} T_j \qquad (i = 2, \ldots, k)\,,$$

$$f_1(a) = S_1(a)\cdot\lambda\,, \qquad f_i(a) = \lambda\cdot S_i(a)\Big(\prod_{j=1}^{i-1} S_j(T_j)\Big)\,,$$

$S_i(a)$ is the survival function of the i-th stage.

Here it has been assumed that the specific fecundity rate is age independent; this restrictive assumption is adequated only for certain populations and for particular environmental factors.

It is clear that if enough observations of populations density are available, the solution of the inverse problem related to the equation (11) allows the estimation of the demographic parameters λ and $S_i(a)$ for $i = 1, \ldots, k$.

2 - Parameter estimation

Our problem is how to solve this kind of equations by means of regularizing functions. The equations (11) are of the type

$$g(x) = [\hat{K}f](x) = \int_{\alpha}^{\beta} K(x-y)\, f(y)\, \mathrm{d}y \qquad x \in [c, d]\,, \tag{12}$$

where α and β are known.

We consider the case that $g(x)$ is not given analytically but a vector $\tilde{g} = (\tilde{g}_1, \ldots, \tilde{g}_m)^T$ of approximate values of $g(x)$ at points $x_1, \ldots, x_m \in [c, d]$.

Owing to measurement errors the quantities $\tilde{g}_j$ found from the measurements are different from their ideal values g_j given by equation (12). This stochastic nature of the quantities observed is an inevitable feature of every actual experiment, and it must appear explicitly in the formulation of the inverse problem. Therefore the algebraized from of the inverse problem related to equation (12) is

$$\sum_{i=1}^{n} K_{ji} f_i \simeq g_j = \tilde{g}_j + \delta_j \qquad (j = 1, \ldots, m)\,, \tag{13}$$

where the f_is form a vector f in the space R^n and are the values of $f(y)$ at certain reference points $y_1, \ldots, y_n$, K_{ji} form a m-by-n matrix K and δ_j form a random vector δ in the space R^m and are introduced owing to the errors of measurements of the quantities g_j.

The unknown quantities f_i determine the quantities g_j; some random process gives us the observed values $\tilde{g}_j$ and the connection between g_j and $\tilde{g}_j$ is δ_j. Thus, if we make the usual assumption that the errors δ_j for different j are indipendent and distributed according to the normal law with mathematical expectation zero, the conditional probability of the vector $\tilde{g} = (\tilde{g}_1, \ldots, \tilde{g}_m)^T$ for a given vector f is known, because

$$P(\tilde{g} = g - \delta/f) = P(\delta) = \det^{\frac{1}{2}}(D^{-2}/2\pi) \exp\left[\|D^{-2}(Kf - \tilde{g})\|^2/2\right]\,, \tag{14}$$

where D is the diagonal matrix $(d_1, \ldots, d_m)$ and d_j is the standard deviation of $\tilde{g}_j$.

The inverse problem related to equation (12) can be formulated in the following way [**5**]: to look for the vector f on condition of the vector g which is known only in probability. The idea can be to find the probability of f under the condition that the measurements have given the results $\tilde{g}$, and therefore assuming the expected value of f over this distribution as estimate of f.

The following observation is important to determine $P(f/\tilde{g})$.

The operator $\hat{K}$ in equation (12) has a very strong « smoothing » action, in the sense that slight perturbation of $g(x)$ might correspond to arbitrarily large perturbations of the solution $f(y)$. Therefore, it is clear that the errors of measurement, though small, cause a solution which may be very different from the true one. Thus, our problem can be regarded as effectively not fully defined and must be completed by a priori information about the solution, so as to get a solution as close as possible to the true one. Thus, we use Bayes' formula to get the conditional probability $P(f/\tilde{g})$

$$P(f/\tilde{g}) = \frac{P(\tilde{g}/f)\,P(f)}{\int_{R^n} P(\tilde{g}/f)\,P(f)\,\mathrm{d}f}\,, \tag{15}$$

because this formula allows the introduction of the a priori information about $f(y)$ by means of the probability density function $P(f)$.

This method to estimate f is particularly suitable for the type of inverse problem related to equation (12) because it takes into account both the stochastic error of observation, and the « smoothing » character of the operator $\hat{K}$ which makes the problem an ill-conditioned one.

If we know that the unknown function $f(y)$ is smooth, we may take, for istance, as degree of smoothness the squared norm of its second derivative

$$J[f(y)] = \int_{y_1}^{y_n} [f''(y)]^2 \, dy \,, \tag{16}$$

and the a priori information is given by fixing a value w which represents the expected value of $J[f(y)]$ over an a priori distribution $P(f)$. Thus, if $(f, \Omega f)$ is the algebraized form of (16), the a priori information is written

$$\int_{R^n} (f, \Omega f) P(f) \, df = w \,. \tag{17}$$

We choose among all functions $P(f)$ satisfying (17) that one which contains the minimum of information about f, i.e. that one which minimizes the functional

$$I(P(f)) = \int_{R^n} P(f) \log (P(f)) \, df \,, \tag{18}$$

which gives a quantitative measure of this information.

Thus the a priori probability $P(f)$ is subject to a problem of conditional extremum, and the corresponding Euler equation is

$$\log (P(f)) + \beta f^T \Omega f/2 + \alpha + 1 = 0 \,, \tag{19}$$

where α and β are Lagrangian parameters.

The function $P(f)$ is found to be [**3**]

$$P(f) = \det^{\frac{1}{2}} (\beta \tilde{\Omega}/2\pi) \exp (- \beta f^T \Omega f/2) \,, \tag{20}$$

where $\tilde{\Omega}$ is a replacement of Ω if Ω is singular, such that $\tilde{\Omega} = \Omega + \varepsilon X$, where X is a positive definite and symmetric matrix and ε is an arbitrarily small positive number.

By virtue of Bayes' formula, the a posteriori conditional probability (15) is then

$$P_\beta(f/g) = \frac{\exp (-(\beta f^T \Omega f + \|D^{-2}(Kf - \tilde{g})\|^2)/2)}{\int_{R^n} \exp (-(\beta f^T \Omega f + \|D^{-2}(Kf - \tilde{g})\|^2)/2) \, df} \,, \tag{21}$$

and the Bayesian estimate f_β of f is found to be the solution of the system

$$(22)\qquad (\frac{\gamma}{2} K^T D^{-2} K + \Omega) f_\beta = \frac{\gamma}{2} K^T D^{-2} \tilde{g},$$

with $\gamma/2 = 1/\beta$.

It has been shown [2] that the solution f_γ of the extremum problem

$$(22)\qquad \sum_{i=1}^{n-1} \int_{y_i}^{y_{i+1}} (f''(t))^2 \, dt + \gamma \| D^{-1}(Kf - g) \|^2 \to \min,$$

where γ is an undetermined Lagrangian parameter, is a natural cubic spline which satisfies the equation

$$(24)\qquad (QT^{-1}Q^T + \frac{\gamma}{2} K^T D^{-2} K) f_\gamma \cdot [\frac{\gamma}{2} K^T D^{-2} \tilde{g},$$

where: $h_i = y_{i+1} - y_i$ $(i = 1, \ldots, (n-1))$; T is a positive definite tridiagonal matrix of order $(n-2)$: $t_{ii} = 2(h_{i-1} + h_i)/3$, $t_{i,i+1} = t_{i+1,i} = h_i/3$; Q is a tridiagonal matrix with n rows and $(n-2)$ columns: $q_{i-1,i} = 1/h_{i-1}$, $q_{ii} = -(1/h_{i-1}) - (1/h_i)$, $q_{i+1,i} = 1/h_i$.

Moreover it is easy to see that the solution of the problem

$$(25)\qquad f^T(QT^{-1}Q^T) f + \gamma \| D^{-1}(\tilde{g} - Kf) \|^2 \to \min$$

is given by

$$(26)\qquad (\frac{\gamma}{2} K^T D^{-2} K + QT^{-1}Q^T) f_\gamma = \frac{\gamma}{2} K^T D^{-2} \tilde{g},$$

so that it is natural to take

$$\sum_{i=1}^{n-1} \int_{v_i}^{v_{i+1}} (f''(t))^2 \, dt \simeq f^T(QT^{-1}Q^T) f.$$

We get a form for Ω, that is $\Omega = QT^{-1}Q^T$, which is very suitable because in this way the solution of (22) is the vector of the values of the cubic natural spline at $y_1, \ldots, y_n$ that may be taken as a natural extension of the Bayesian estimation in all the range $[y_1, y_n]$.

By the Cholesky decomposition of T^{-1}, $T^{-1} = SS^T$, and the positions $H = QS$, $\gamma/2 = p^2$, it is easy to verify that the system (24) is the normal system of the system

$$\left[\frac{pD^{-1}K}{H^T}\right] f_\gamma = \left[\frac{pD^{-1}\tilde{g}}{0}\right]. \tag{27}$$

It is known [1] that every m-by-n real matrix A can be decomposed (singular value decomposition) into the form $A = U\Sigma V^T$, where U is an m-by-m orthogonal matrix, V is an n-by-n orthogonal matrix, and Σ is an m-by-n matrix whose elements are $\sigma_{ij} = 0$ for $i \neq j$ and $\sigma_{ii} = \sigma_i \geqslant 0$.

We may use the singular value decomposition to obtain the Moore-Penrose pseudoinverse A^+ of A, $A^+ = V\Sigma^+ U^T$, where Σ^+ is an n-by-m matrix with $\sigma_i^+ = 1/\sigma_i$ if $\sigma_i > 0$, or $\sigma_i^+ = 0$ if $\sigma_i = 0$.

The solution of the system (24) is expressed as

$$f_\gamma = \left[\frac{pD^{-1}K}{H^T}\right]^+ = \frac{pD^{-1}\tilde{g}}{0}\Big]. \tag{28}$$

3 - Numerical results

We show how the described estimation method can be applied in the study of the dynamics of some populations.

As a first study case we refer to a population of *Ceratitis* which may be represented as having two developmental stages: the first stage of preimmaginals contains eggs, larval and pupae and will be indicated by I, the second stages of adults will be indicated by A.

We try to identify the specific fecundity and survival rate of the population.

To simulate the observations we have choosen

$$S_I(a) = e^{-\mu_I \cdot a}, \qquad S_A(a) = e^{-\mu_A \cdot a},$$

and we have obtained $I(t)$ and $A(t)$ by means of numerical integration of differential equations with $T_I = 14$, $\lambda = 8.5$, $\mu_I = 0.1$ and $\mu_A = 0.02$. We have estimated $f_I(a) = \lambda \cdot S_I(a)$ and $f_A(a) = \lambda \cdot S_I(T_I) S_A(a)$ of the equations (11) in the interval $[0, 14]$ for both preimmaginal and adult stages with age step $\Delta a = 1$. The optimum values for the Lagrangian parameters are

$$\gamma_I = 0.01\,, \quad \gamma_A = 0.0001\,,$$

with the corresponding mean square errors

$$F_I = 2\,, \qquad\qquad F_A = 2.5\,.$$

As a second study case we have considered a population of *Daphnia* represented by means of three developmentel stages: eggs, young and adults.

The simulation of the birth-deat process is been performed as in the previous case with the following parameters: $T_E = 2$, $T_Y = 5$, $\lambda = 5$, $\mu_E = 0.02 = \mu_A$ and $\mu_Y = 0$.

The problems of estimation has been solved in the intervals [0, 2], [0, 5], [0, 7] for eggs, young, and adults respectively, with age step $\Delta a = 0.5$.

The optimum values for the Lagrangian parameters are

$$\gamma_E\ 0.1\,, \qquad \gamma_Y = 0.01\,, \qquad \gamma_A = 0.001\,,$$

with the corresponding mean square errors

$$F_E = 2\,, \qquad F_Y = 1.5\,, \qquad F_A = 3.5\,.$$

Bibliografia

[1] G. Forsythe, M. A. Malcom and C. B. Moler, *Computer methods for mathematical computations*, Prentice Hall (1977), 192-235.

[2] I. Galligani, *Sulla regolarizzazione dei dati sperimentali*, Calcolo **7** (1972), 359-376.

[3] G. Klein, *On spline functions and statistical regularization of ill-posed problems*, J. Comput. Appl. Math. **5** (1979), 259-263.

[4] C. H. Reinsch, *Smoothing by spline functions*, Numer. Math. **10** (1967), 177-183.

[5] V. F. Turchin, V. P. Kozlov and M. S. Malkevich, *The use of mathematical-statistics methods in the solution of incorrectly posed problems*, Uspehi Fiz. Nauk **102** (1971), 345-386 (Soviet Phys. Usp. **13** (1971), 681-840).

[6] H. Von Foerster, (Ed. F. Stoheman jr.), *Some remarks on changing populations, in the kinetics of cellular proliferation*, Grune and Stratton, New York (1959), 382-407.

Summary

We propose a mathematical model in integral form which describes the birth-death process of populations wich may be subdivided in a suitable number of developmental stages.

The demographic parameters of such populations are estimated by solving the inverse problems related to the integral equations by means of statistical regularization method proposed by Turchin which takes into account the stochastic nature of the observations.

The estimated parameters have been obtained as natural spline functions.

Numerical results which refer to zooplanktonic and insect populations are presented.

* * *