Inverse Problems and Structured-Population Dynamics

November 3, 2003

Thomas Shores

Material is taken from Chapter 19 of text *Structured-Population Models in Marine, Terrestrial and Freshwater Systems*, Shripad Tuljapurkar and Hal Caswell, Ed., Kluwer (1997).

Outline:

1. Forward Problems

2. Inverse Tools

3. Examples

4. Conclusions

We are given a model and its parameters, from which we calculate the solution, which in this discussion is a population function of some sort. Typically, a forward problem is well-posed, i.e., has a unique solution which varies continuously with the parameters of the problem. Three standard examples:

1. Matrix Form of Stage Structured Model (discrete structure, discrete time): Population is grouped according to a structural property such as age or weight into m discrete stages. Population is determined at discrete times t_j, $j = 0, 1, \ldots$ and population of the j-th stage at time t_k is $n_j(t_k)$. It is assumed that individuals within a stage are identical. Vector of all stage populations at time t_k is the m-vector $\mathbf{n}(t_k) = \left[n_j(t_k) \right]_{j=1}^{m}$. We are given a "projection matrix" $A(t_k)$ that may vary with time. We are given initial condition $\mathbf{n}(t_0) = \mathbf{n}_0$ and population dynamics:

$$\mathbf{n}(t_{k+1}) = A(t_k)\mathbf{n}(t_k)$$

Example. (Usher or "standard size-classified matrix" − H. Caswell) In this example the iteration process is stationary, i.e., A is constant of the form

$$
A = \begin{bmatrix}
P_1 & \beta_2 & \beta_3 & \cdots & \beta_m \\
G_2 & P_2 & & & \\
& G_3 & P_3 & & \\
& & \ddots & \ddots & \\
& & & G_m & P_m
\end{bmatrix}
$$

where

β_j: the birthrate in stage j

P_j: the within-stage survival rate of stage j

G_j: between-stage transition rate from stage j to stage $j + 1$

Remark: parameters β_j, P_j, G_j could be functions of time or even $\mathbf{n}$ (e.g., β_j might be a decreasing function of $n_j(t_j)$.

2. Delay-Differential Equation (discrete structure, continuous time):Matrix Form of Stage Structured Model (discrete structure, discrete time):

Population is grouped according to a structural property such as age or weight into m discrete stages. Population is determined at discrete times t_j, $j = 0, 1, \ldots$ and population of the j-th stage at time t_k is $n_j(t_k)$. It is assumed that individuals within a stage are identical, except for a developmental index, like age. Vector of all stage populations at time t is the m-vector $\mathbf{n}(t) = \left[n_j(t) \right]_{j=1}^{m}$. Population dynamics:

$$\frac{dn_j(t)}{dt} = R_j(t) - M_j(t) - \mu_j(t)n_j(t), \, j = 1, \ldots m, \, t > 0$$

where

μ_j: the death rate in stage j

R_j: recruitment rate to stage j

M_j: maturation rate from stage j to stage $j + 1$

Remark: typically, we expect that

$$R_j(t) = M_{j-1}(t)$$

and

$$M_j(t) = R_j(t - \tau_j)e^{-\int_{t-\tau_j}^{t} \mu_j(s)ds}$$

where τ_j is the duration of stage j.

Example. (*Matrix Population Models*, 2nd Ed., Hal Caswell, Sinauer (2001).) A two-stage model for juveniles ($n_J(t)$) and adults ($n_A(t)$).

$$\frac{dn_J}{dt} = R_J(t) - M_J(t) - \mu_J n_J(t)$$
$$\frac{dn_A}{dt} = R_A(t) - \mu_A n_A(t)$$

where

$$R_A(t) = M_J(t)$$
$$M_J(t) = R_J(t - \tau)e^{-\mu_J \tau}.$$

3. PDE Form of Structured Model (continuous structure, continuous time):

[Simple versions! See *An Introduction to Structured Population Dynamics*, J. M. Cushing, CBMS-NSF Series, SIAM (1998) for more complex renditions.] Population is classified according to a continuous structure parameter x such as age, size. Let $n(x,t)$ be the population density (population/structure) at structure value x and time t. A simple conservation law analysis yields the PDE

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x}\left(\gamma(x,t)n(x,t)\right) = -\mu(x,t)n(x,t), \; x \geq 0, \; t > 0$$

with the non-local boundary condition

$$n(0,t) = \int_0^\infty \beta(x,t)n(x,t)\, dt, \; t > 0$$

and initial condition

$$n(x,0) = n_0(x), \; x \geq 0,$$

where

γ: structural growth rate

β: birthrate

Remark: This model is called the Sinko-Streifer model (1967).

Example. The most widely studied special case of this model is the McKendrick-von Foerster age-structured model

$$\frac{\partial n}{\partial t} + \frac{\partial n(x,t)}{\partial a} = -\mu(a,t)n(a,t),\ a \geq 0,\ t > 0$$

with the non-local boundary condition

$$n(0,t) = \int_0^\infty \beta(x,t)n(x,t)\,dt,\ t > 0$$

and initial condition

$$n(a,0) = n_0(a),\ a \geq 0,$$

where

a: age (so that growth rate is clearly 1)

β: birthrate

Inverse Tools

The problem we consider is as follows: Given a sampling of the population of a system at various points (possibly with noise) and possibly a subset of the parameters of the problem, to recover an approximation to the remaining parameters. In other words: given the solution, what is the problem?

Fundamental Difficulties:

1. There may be multiple solution or no solution (Ill-posedness of inverse problem).

2. We will almost certainly be given partial data.

3. We will almost certainly be given faulty data.

All is not lost. We hope that the data plus our own biases about the nature of the will lead to an acceptible approximate answer.

Two basic approaches:

1. Deterministic + fudge factors (e.g., regularization parameters) that make the problem well-posed.

2. Statistical data fitting.

A Conceptual Framework

The forward problem looks like

$$A_P(N) = 0$$

where A_P is some operator depending on the parameters P. The forward problem is to invert this operator to solve for N.

We can express this symbolically as

$$N = F(P)$$

where the dependence of the forward problem on P gets promoted to an argument. But this doesn't fit experiments. If the operator H represents the discretization of the data N to its experimental form $\mathbf{n}$ (e.g., evaluation at certain points) then the appropriate operator for inversion is

$$F(D) = \mathbf{n} = HN.$$

Once we have performed our experiments, we have an experimental approximation $\mathbf{y} \approx \mathbf{n}$.

Steps to take:

1. Discretize the parameter set, if needed. E.g., in estimating a continuous parameter like $\mu(s)$, assume a form for the parameter like a polynomial of a certain degree or a cubic spline of a certain degree.

Remark: The discretization should offer the possibility of converging to the correct μ. This step is *not* necessary if the parameter set is already discrete, e.g., a matrix model or a delayed-differential equation model where the parameters are known to be constants. In all cases we end up with a finite dimensional vector $\mathbf{p}$ that adequately represents P.

2. Smooth the data, if needed.

This provides a mechanism for passing from $\mathbf{n}$ to N. For example, if $N = n(a, t)$ in an age-structured model, and the experimental data provides discrete data $\left\{n(a_j, t_n)\right\}$ for $j = 1, \ldots, m$ and $n = 1, \ldots, s$, we could interpolate the data with splines. This interpolation is our approximation to $n(s, t)$.

3. Express the problem in discretized form

$$f(\mathbf{p}) = \mathbf{n}.$$

4. Plug in experimental data $\mathbf{y}$ in place of $\mathbf{n}$ and solve the system $f(\mathbf{p}) = \mathbf{y}$ as best you can.

Linear system (the easy case): $W\mathbf{p} = \mathbf{y}$

Method of choice is least squares, that is, to minimize the residual

$$\|\mathbf{y} - W\mathbf{p}\|^2 = \mathbf{p}^T W^T W \mathbf{p} - 2\mathbf{p}^T W^T \mathbf{y} + \mathbf{y}^T \mathbf{y}$$

with respect to $\mathbf{p}$, which leads to the normal equations

$$W^T W \mathbf{p} = W^T \mathbf{y}.$$

Remarks:

1. We may still have a problem if we do not have sufficiently many data points. While a solution is guaranteed, it may not be unique.

2. Assuming sufficiently many data points we expect W to be of full column rank, so that $W^T W$ is invertible and $p = \left(W^T W\right)^{-1} W^T \mathbf{y}$.

3. In the case of full column rank, this solution is the best unbiased estimator of the true value of $\mathbf{p}$, provided the error is identically distributed random vector with expected value 0 and finite covariance.

Nonlinear system (the hard case): $f(\mathbf{p}) = \mathbf{y}$

Again, use the least squares idea: minimize the residual

$$F(\mathbf{p}) = \|\mathbf{y} - f(\mathbf{p})\|^2.$$

Model the function f linearly as starting from $\mathbf{p}_0$ and proceeding to minimum at $\mathbf{p} = \mathbf{p}_0 + \alpha$ with

$$f(\mathbf{p}_0 + \alpha) = f(\mathbf{p}_0) + J\alpha$$

where J is the Jacobian of $f(\mathbf{p})$. So we minimize

$$\|\mathbf{y} - f(\mathbf{p}_0) - J\alpha\|^2 = \|\delta - J\alpha\|^2$$

yielding step $\alpha = \left(J^T J\right)^{-1} J^T \delta$. Of course, this is an iterative method yielding a sequence of iterates $\mathbf{p}_0, \mathbf{p}_1, \ldots$ converging to the optimum $\mathbf{p}$.

There are problems with this approach: the sequence of jacobians must be uniformly well conditioned for the method to succeed.

Remedy: (Levenberg-Marquardt) replace the iteration formula by

$$\alpha = \left(J^T J + \lambda I\right)^{-1} J^T \delta.$$

As Wood explains it: if the first term is dominant, we are essentially taking a downhill gradient step because $-2J^T\delta = \nabla F$. If the first term is dominant, we are doing a quadratic approximation to the least squares objective function. "λ is adjusted in the following way: if $F(\mathbf{p} + \alpha) < F(\mathbf{p})$, let $\lambda = \lambda/10$ and $\mathbf{p} = \mathbf{p} + \alpha$, else $\lambda = 10\lambda$ and leave $\mathbf{p}$ unchanged."

Inverse theorist's point of view: we are minimizing

$$\|\delta - J\alpha\|^2 + \lambda \|\mathbf{p}\|^2,$$

so the cycle is: forward solve, linearize, regularize and update. This is the *output least squares* method.

Some Examples

Conclusions