

DETERMINAÇÃO DOS PARÂMETROS DE MODELOS BIOMATEMÁTICOS

Marat M. Rafikov

Departamento de Física, Estatística e Matemática

UNIJUI - Ijuí, RS

RESUMO

Sabe-se que existem dois tipos principais da resolução do problema de identificação de modelos dinâmicos. O primeiro deles é quando existe o número suficiente de dados sobre o sistema modelado e há possibilidade de aplicar uma certa hipótese estatística. E o outro é quando não existe uma base séria para aplicar alguma hipótese estatística. No último caso o problema de identificação torna-se problema de ajustamento entre os parâmetros do modelo e os dados observados. Aqui se apresentam dois métodos deste ajustamento. Um deles baseia-se numa propriedade particular do sistema presa-predador, a saber: os valores médios do número de presas e predadores ao longo de um período são os mesmos para todas as soluções. E o outro método reduz o problema de identificação à resolução de um sistema de equações algébricas lineares com número de equações maior que número de variáveis. Os métodos propostos foram aplicados para analisar a incidência de lagartas da soja (*Anticarsia gemmatalis*) e seus inimigos naturais que convivem num sistema presa-predador. Os resultados da identificação mostraram uma boa aproximação dos parâmetros do modelo aos dados observados.

SUMMARY

Two main methods of resolving problems related to the identification of dynamic models are known. The first method can be used when there is sufficient data of the modeled system and when the application of a statistical hypothesis is allowed. The other method can be used when no serious base to apply statistical hypothesis exists. In the latter, the problem of identification becomes a problem of adjusting parameters to the model and data observed. Two algorithms of such adjusting parameters are presented in this paper. One algorithm is based on particular characteristic of the prey-predator system. In this case the average values of prey and predators, during a certain period, are equal for all solutions. The other algorithm reduces the identification problem to the resolution of a linear algebraic equation system, where the number of equations is greater than the number of variables. The proposed algorithms were used to analyze the population of soybean caterpillars (*Anticarsia gemmatalis*) and its natural enemies, living together, in a prey-predator system. The results show that a good approximation between parameters of model and observed data occurred.

1 - INTRODUÇÃO

Os modelos biomatemáticos usualmente são representados por equações diferenciais ou de diferença que modelam o balanço dinâmico de populações ou substâncias em cada momento do tempo. Estas equações são uma representação adequada de processos reais. Usualmente, todos os modelos que descrevem os processos reais, possuem parâmetros desconhecidos. Entre eles figuram coeficientes do modelo, condições iniciais, etc. O pesquisador, se quiser estabelecer as características e regularidades de um sistema biomatemático concreto tem que identificar parâmetros do modelo conforme as informações sobre o processo real. Somente depois de formular corretamente e resolver o problema de

identificação é possível passar com certeza dos modelos matemáticos que imitam os processos reais, ao bom entendimento e controle real dos mesmos processos.

O objetivo e o método de identificação depende essencialmente da informação acessível. Sabe-se que quaisquer dados de observação são imprecisos, ou seja, contêm elementos de incerteza (BARD, 1974). Esta incerteza é consequência de erros de medições, efeitos casuais ou não lineares, influência de processos incógnitos. Se os dados de observações são resultados de medições que revelam regularidades estatísticas, então para determinar os parâmetros de modelos é possível aplicar os métodos estatísticos de avaliação ótima. Neste caso, os valores ótimos de parâmetros maximizam ou minimizam os critérios escolhidos: o critério da verossimilhança máxima, o critério do máximo da probabilidade a posterior, etc. Um aspecto muito importante da aplicação de métodos estatísticos é a presença de um número suficiente de dados sobre o sistema modelado e a possibilidade de aplicar uma certa hipótese estatística.

Em vários problemas da modelagem matemática a situação é a seguinte: os dados observados são as estimativas indiretas ou não existe uma base séria para aplicar alguma hipótese estatística. Neste caso, o problema de identificação torna-se problema de ajustamento entre os parâmetros do modelo e os dados observados, ou seja, é preciso determinar tais valores dos parâmetros do modelo para os quais a diferença entre os valores de variáveis modeladas e observadas seja pequena (BARD, 1974, MARCHUK, 1991). Um critério utilizado usualmente para o ajustamento é a soma dos quadrados de desvios entre os resultados observados e os valores de variáveis do modelo calculada para todos os momentos observados. Para mostrar que o modelo é adequado é necessário realizar a experiência computacional com base no modelo considerado e comparar os resultados obtidos através do modelo com os dados coletados. Este trabalho apresenta uma abordagem que se refere ao segundo enfoque acima referido.

2 - ALGORITMOS DE IDENTIFICAÇÃO DE MODELOS BIOMATEMÁTICOS

Para identificação numérica de sistemas biomatemáticos não existe um método poderoso universal. Por isso existem vários enfoques nesta área, por exemplo, os trabalhos MARCHUK (1991), ZUEV (1988) para modelos matemáticos de imunologia, BACCALÁ, BENNETON (1990) para um caso particular do modelo Lotka-Volterra, HOLMBERG (1982) para modelos do tipo Michaelis - Menten do crescimento de bactérias, RAFIKOV, ARAUJO (1994) para o modelo clássico Lotka-Volterra.

Neste trabalho propomos dois métodos de identificação. O primeiro é uma modificação do algoritmo proposto em RAFIKOV, ARAUJO (1994) que baseia se numa propriedade particular do sistema de duas espécies do tipo presa-predador. O segundo método é uma modificação do algoritmo proposto no livro do MARCHUK (1991) para sistemas imunológicos e serve para vários tipos de sistemas que modelam interações entre muitas espécies. Os elementos novos do algoritmo são seguintes: antes de discretizar o problema tem que ser feita a integração na parte esquerda do sistema; podem ser utilizadas várias maneiras de discretização do sistema de equações diferenciais.

2.1 - Determinação dos parâmetros do modelo clássico Lotka-Volterra

O modelo presa-predador foi estabelecido e analisado independentemente por Lotka e Volterra, por volta de 1925 e tem a forma de sistema de equações diferenciais ordinárias:

$$\begin{aligned}\frac{dx}{dt} &= ax - \alpha xy \\ \frac{dy}{dt} &= -by + \beta xy\end{aligned}\tag{1}$$

onde x é a população de presas e y é a população dos predadores; a, α, b, β são constantes positivas, sendo α a taxa de crescimento das

presas, b a taxa de mortalidade dos predadores e α, β os coeficientes que caracterizam a interação entre as duas espécies.

Os métodos de identificação dos parâmetros do modelo (1) propostos no trabalho (RAFIKOV, ARAUJO, 1994) baseiam-se numa propriedade muito importante deste modelo : os valores médios de x e y ao longo de um período são os mesmos para todas as soluções, embora soluções diferentes tenham períodos distintos (BASSANEZI, FERREIRA, 1988), ou seja,

$$\frac{1}{T} \int_0^T y dt = \frac{a}{\alpha} = \bar{y} \quad \text{e} \quad \frac{1}{T} \int_0^T x dt = \frac{b}{\beta} = \bar{x} \quad (2)$$

Determinando os valores médios de \bar{x}, \bar{y} com base na estatística dada sobre populações de presas e predadores ao longo do

período T , pode-se calcular razões $\frac{a}{\alpha}$ e $\frac{b}{\beta}$. Existem várias opções para

calcular aproximadamente as integrais que existem no lado esquerdo das expressões (2). A idéia principal é a seguinte: fazer a aproximação das funções x e y com base nas informações dadas e depois disso calcular as integrais. No trabalho atual propõe-se realizar a aproximação de funções x e y através de splain aproximação (splain approximation), ou seja, calcular os valores de funções x e y no intervalo entre quaisquer dois dados estatísticos em forma de um polinômio de terceiro grau (cubic interpolation). É importante que para esta aproximação as curvas de aproximação passam pelos pontos da estatística e têm as primeiras e segundas derivadas contínuas nestes pontos. Calculando as integrais é fácil achar os valores de razões da fórmula (2). Para determinar todos os coeficientes do modelo (1) há mais fórmulas:

$$T = \frac{2\pi}{\sqrt{ab}}; \quad A_x = k \frac{b}{\beta}; \quad A_y = k \frac{a}{\alpha} \sqrt{\frac{b}{a}} \quad (3)$$

onde A_x, A_y são as amplitudes de oscilações do número de presas e predadores, respectivamente; k é uma constante arbitrária.

Os valores preliminares de T, A_x, A_y podem ser escolhidos da estatística. Durante a experiência computacional para cada período T podem ser obtidos valores melhores sob ponto de vista de um critério que caracterize a aproximação das trajetórias temporais às estatísticas. Foram minimizados os critérios que calculam o valor médio de módulos de desvios entre os valores de populações de presas e predadores calculados e estatísticos:

$$S_x = \frac{1}{N} \sum_{i=1}^N |x(t_i) - x_i| \quad ; \quad S_y = \frac{1}{N} \sum_{i=1}^N |y(t_i) - y_i| \quad (4)$$

Então foi formulado e resolvido o seguinte problema de identificação: variando os valores de T, A_x, A_y, x_0, y_0 achar os valores mínimos de critérios (4).

Os algoritmos do trabalho de RAFIKOV, ARAUJO (1994) e acima referido baseiam se numa propriedade particular do sistema de duas espécies do tipo presa-predador, e por isso não servem para identificação de sistemas de multiespécies.

2.2 - Determinação dos parâmetros de modelos de n espécies

Para descrever o algoritmo vamos considerar uma generalização do modelo clássico Lotka-Volterra (GROSSBERG, 1978):

$$\dot{x}_i = G_i(x_i) \left(b_i - \sum_{j=1}^n a_{ij} F_j(x_j) \right) \quad (5)$$

onde x_i é a densidade de i -ésima espécie do sistema modelado ($i=1,2,\dots,n$) e as funções G_i, F_j são deriváveis para qualquer ordem e satisfazem as seguintes condições:

$$\begin{aligned}
 f_i(0) &= 0, & i &= 1, 2, \dots, n \\
 \frac{\partial f_i(x_i)}{\partial x_i} &> 0 & \text{para } x_i &\geq 0, \quad i = 1, 2, \dots, n \\
 g_i(0) &= 0, & i &= 1, \dots, n \\
 g_i(x_i) &> 0, & \text{para, } x_i &> 0, \quad i = 1, 2, \dots, n
 \end{aligned}$$

Este tipo de generalização do sistema clássico Lotka-Volterra permite considerar um número maior de problemas que surgem nas aplicações.

Suponhamos que são conhecidos os resultados de $M + 1$ medições de x_i , $i = 1, 2, \dots, n$, ($M > n + 1$) que foram realizadas nos momentos t_0, t_1, \dots, t_M . Para fazer a discretização multiplicaremos cada equação do sistema (5) por $dt / G_i(x_i)$. Integrando cada equação no intervalo $[t_0, t]$, chegamos à forma integral do sistema (5):

$$\int_{x(t_0)}^{x(t)} \frac{dx_i}{G_i(x_i)} = b_i(t - t_0) - \sum_{j=1}^n a_{ij} \int_{t_0}^t F_{ij}(x_j) dt \quad (6)$$

Rescrevendo (6) para cada momento t_k , $k=1, 2, \dots, M$, chegamos ao sistema:

$$b_i(t_k - t_0) - \sum_{j=1}^n J_{ij}(k) a_{ij} = D_i(k) \quad (7)$$

onde

$$\begin{aligned}
 D_i(k) &= \int_{x(t_0)}^{x(t_k)} \frac{dx_i}{G_i(x_i)}; & J_{ij}(k) &= \int_{t_0}^{t_k} F_{ij}(x_j) dt \\
 i &= 1, 2, \dots, n; & k &= 1, 2, \dots, M
 \end{aligned} \quad (8)$$

A função subintegral da integral $D_i(x_i)$ é uma função de x_i conhecida, por isso, o valor de $D_i(x_i)$ pode ser calculado utilizando os métodos conhecidos (por exemplo, método de parábolas). Para calcular valores aproximados das integrais definidas $J_{ij}(k)$ podemos utilizar a mesma idéia que foi utilizada para calcular as integrais (2), a saber, fazer a interpolação com base em medições das variáveis x_i do modelo através de funções $S_i(t)$ de spline-interpolação (ou de outro tipo de interpolação). No trabalho de MARCHUK (1991) estas funções foram chamados *quasisoluções*.

Então, considerando funções $S_i(t)$ como um certo protótipo da solução do modelo que corresponde ao vetor dos parâmetros incógnitos, podemos calcular aproximadamente as integrais $J_{ij}(k)$

$$J_{ij}(k) = \int_{t_0}^{t_k} F_{ij}(x_j) dt \cong \int_0^{t_k} F_{ij}(S_j(t)) dt \quad i=1,2,\dots,n; k=1,2,\dots,M \quad (9)$$

Então para cada i , ($i = 1,2,\dots,n$) obtemos M equações lineares (7) para encontrar $n + 1$ parâmetros b_i, a_{ij} .

Em muitos casos para cálculos numéricos é melhor utilizar as integrais definidas (8) com limites de t_{k-1} a t_k :

$$\bar{D}_i(k) = \int_{x(t_{k-1})}^{x(t_k)} \frac{dx_i}{G_i(x_i)}; \quad \bar{J}_{ij}(k) = \int_{t_{k-1}}^{t_k} F_{ij}(x_j) dt \quad (10)$$

$$i = 1,2,\dots,n; k = 1,2,\dots,M$$

É óbvio que:

$$\bar{J}_{ij}(k) = J_{ij}(k) - J_{ij}(k-1)$$

Neste caso o sistema de equações algébricas lineares para encontrar os coeficientes é:

$$b_i(t_k - t_{k-1}) - \sum_{j=1}^n \bar{J}_{ij}(k) a_{ij} = \bar{D}_i(k) \quad i=1,2,\dots,n; \quad k=1,2,\dots,M \quad (11)$$

Os sistemas (7) e (11) são sistemas de equações algébricas lineares em que o número de equações nM sempre é maior que o número de parâmetros incógnitos $n(n+1)$. Em um modo geral as equações destes sistemas são incompatíveis. As soluções de sistemas (7) e (9) podem ser encontrados aproximadamente pelos mínimos quadrados.

Uma solução pelos mínimos quadrados de um sistema de equações lineares

$$\mathbf{AP} = \mathbf{B} \quad (12)$$

é o vector com a menor norma euclidiana que minimiza

$$\|\mathbf{AP} - \mathbf{B}\|^2.$$

Esse vector é

$$\mathbf{P} = \mathbf{A}^+ \mathbf{B}, \quad (13)$$

onde \mathbf{A}^+ é a pseudoinversa ou inversa generalizada (de Moore-Penrose) da matriz \mathbf{A} .

Quando a matriz \mathbf{A} de dimensão $m \times n$ ($m > n$) é de característica integral, ou seja,

$R(\mathbf{A}) = n$, a pseudoinversa pode ser encontrada através da formula:

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (12)$$

Para nosso caso, os componentes do vector \mathbf{P} são os parâmetros b_i , a_{ij} , $i=1,2,\dots,n$; $j=1,2,\dots,n$; e do vector \mathbf{B} são os valores

$D_i(x_i)$ para o sistema (7) e os valores $\bar{D}_i(x_i)$ para o sistema (11); e a matriz \mathbf{A} consiste de elementos $t_k - t_0$; $J_{ij}(k)$ para o sistema (7) e de $t_k - t_{k-1}$; $\bar{J}_{ij}(k)$ para o sistema (11).

Em muitos casos os parâmetros obtidos através do algoritmo acima referido fornecem soluções que não são próximos aos dados estatísticos. Neste caso os valores obtidos de parâmetros podem servir como a primeira aproximação nos algoritmos de minimização de funções que caracterizam desvios entre os valores modelados e estatísticos. Estes algoritmos não são apresentados no trabalho atual e serão considerados no próximo trabalho.

3. APLICAÇÕES E CONCLUSÕES

Para ilustrar os métodos de identificação foi utilizada a informação estatística do trabalho de RAFIKOV, ARAUJO, (1994) sobre a incidência de lagartas da soja (*Anticarsia gemmatalis*) e seus predadores (aranhas, formigas e outros) que está na Tabela 1.

Tabela 1. Dados Estatísticos da incidência de lagartas da soja e seus predadores

T (tempo em dias)	0	6	22	29	34
X (número de lagartas)	35	16	114	81	32
Y (número de predadores)	15	9	8	22	24

Observação. Para este exemplo simplificado o número total de predadores em cada momento do tempo foi considerado como a soma de todos os tipos de predadores, ou seja, implicitamente foi admitida a hipótese sobre a igualdade das capacidades predatórias de todos os predadores. Para aplicações reais a soma deve ser calculada com coeficientes de peso que caracterizam a capacidade predatória. Um outro método de agregação de predadores é escolher o número de presas atingidas como a medida de populações de predadores. Os algoritmos propostos servem para qualquer método de medição de populações de predadores.

Os dados da Tabela 1 foram amostrados pelo método do pano, na região de Ijuí, durante a safra 92/93.

Da análise da estatística é visto que o período de oscilações fica entre 34 e 41 dias, as amplitudes de oscilações do número de presas e predadores são 55 e 8, respectivamente. Então, os valores preliminares de T , A_x , A_y foram escolhidos dentro destes números. Como resultado de experiências computacionais que foram feitas conforme o primeiro algoritmo acima referido (as fórmulas (2)-(4)), foram obtidos os seguintes valores destes parâmetros: $T = 36$; $A_x = 63$; $A_y = 15$.

Através do segundo algoritmo com base na estatística da Tabela 1 foram testados várias formas concretas do modelo (5). O modelo mais adequado sob o ponto de vista de desvios médios entre pontos estatísticos e valores correspondentes de variáveis do modelo foi o modelo clássico de Lotka-Volterra (1).

A Tabela 2 resume os resultados da identificação dos coeficientes do modelo (1) através de dois métodos.

Tabela 2. Resultados de identificação do modelo Lotka-Volterra

Coeficientes e parâmetros do modelo.	Algoritmo 1	Algoritmo 2
\bar{X}	59,60	58,20
\bar{Y}	11,67	12,04
a	0,143502	0,2016
α	0,012299	0,018
b	0,212274	0,1816
β	0,003562	0,0029

Os gráficos que mostram a variação dos números de lagartas da soja e seus predadores são representados nas Figuras 1 e 2, respectivamente.

A comparação dos métodos mostrou a proximidade dos resultados tanto na identificação de parâmetros como nas curvas que modelam o número de lagartas e seus inimigos naturais.

Os métodos de identificação propostos foram aplicados às várias comunidades de lagartas, pragas da soja e seus inimigos naturais. Em todos os casos foram identificados os coeficientes do modelo e em maioria dos casos a aproximação foi aceitável.

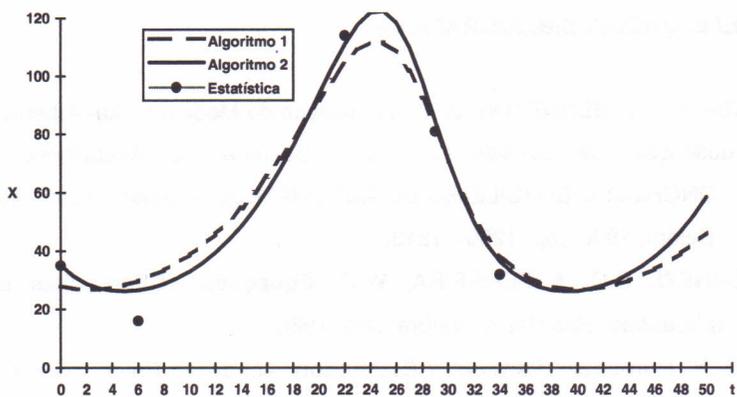


Figura 1. Variação de número de lagartas da soja (*A. gemmatilis*)

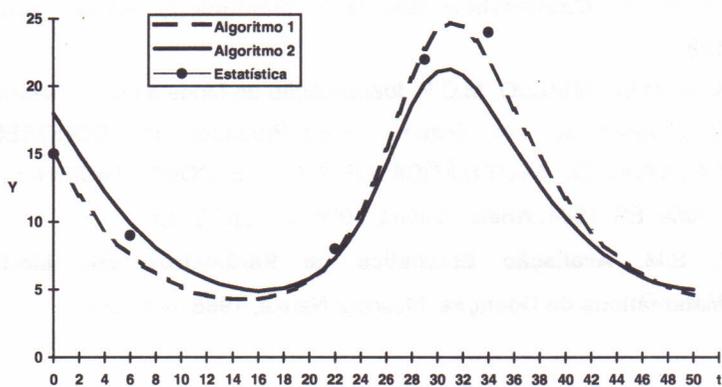


Figura 2. Variação de número de predadores da lagarta da soja (*A. gemmatilis*)

4 - REFERÊNCIAS BIBLIOGRÁFICAS

- BACCALÁ, L.A., BENNETON, J.F. Identificação do Modelo Lotka-Volterra na descrição da sensibilidade de Bactérias a Antibióticos. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 8. Belém, 1990. **Anais** ...Belém, 1990, ,pp. 1209 - 1213.
- BASSANEZI, R.C. & FERREIRA, W.C. **Equações Diferenciais com Aplicações**. São Paulo: Harbra Ltda, 1988.
- BARD, Y. **Non-linear Parameter Estimation**. Academic Press, New York., 1974.
- GROSSBERG, S. Competition, decision and consensus. **J. of Math. Anal. Appl.**, 1978, v. 66, No 2,pp.470-493, 1978.
- HOLMBERG, A. On Practical Identifiability of Microbial Growth Models Incorporating Michaelis - Menten Type Nonlinearities. **Mathematical Biociences**, N 62, pp. 23-43, 1982.
- MARCHUK, G.I. **Mathematical Models in Immunology**. Moscou: Nauka., 1991.
- RAFIKOV M.M., ARAUJO M.C.P. Identificação do Modelo Lotka-Volterra na Modelagem de um Sistema Presa-Predador. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 17. Vitória, ES, 1994. **Anais** . Vitória, 1994, v. I, pp.79 -83.
- ZUEV, S.M. **Avaliação Estatística de Parâmetros dos Modelos Matemáticos de Doenças**. Moscou: Nauka, 1988. (em russo).