Módulo 13. Regulação em reprodutores contínuos: a eq. logística

Objectivos

Suponhamos que se dispõe de observações da densidade populacional (N_1, N_2, N_3, ...) duma população de reprodutores contínuos, na qual se suspeita que ocorrem fenómenos de regulação dependente da densidade (RDD):

1) Como proceder para detectar a existência de RDD a partir desta sucessão de observações ?

2) Admitindo que a população não só exibe RDD mas também que cresce de acordo com o modelo logístico, como é que se podem estimar os valores de *r* e *K* a partir desta sucessão de observações ?

O objectivo desta aula prática é responder a estas perguntas. A detecção (e, mais ainda, a demonstração) de fenómenos de RDD na natureza é, em geral, um tema difícil. Quanto à 2ª pergunta, levanta problemas estatísticos que nunca foram satisfatoriamente resolvidos. Espera-se que, no decorrer dos exercícios, os alunos se apercebam destas dificuldades.

Na aula é também apresentado um método de geração de npa's com distribuição Normal e, mais uma vez, tirar-se-à partido das funções gráficas das folhas de cálculo.

A folha de cálculo com a aula tem o nome "Mod 13 Praticas.xls" e pode ser descarregada do site das aulas.

Métodos resumidos

Vamos assumir que o crescimento da população é do tipo logístico, e que as medições N_1, N_2, N_3, ... apresentam erros (erros de medição e devidos a perturbações ambientais) com distribuição que se pressupõe Normal. Começamos por gerar estas observações, usando valores pré-estabelecidos de *r* e *K*. Em seguida, usamos uma técnica para detectar RDD e depois para estimar *r* e *K*. Os valores estimados de *r* e *K* são comparados com os valores pré-estabelecidos e discutem-se os resultados.

1. Construção da equação logística dos reprodutores contínuos em Excel

1) Abra uma folha de cálculo nova e escreva em A1:A3 os simbolos dos parâmetros da logística, respectivamente: *r*, N_0,*K*. Nas células B1:B3 introduza os respectivos valores numéricos, por exemplo: 1.2, 5, 100.

Em A6 escrever 0 e em A7 introduza a fórmula **=A6+0.2**. Copie e cole esta fórmula em A8:A65 para obter a coluna que tem os instantes de tempo em que N_t será calculado (em intervalos Δt=0.2).

2) Em B6 introduza a fórmula integral da logística: **=(B$3*B$2)/(B$2+(B$3-B$2)*EXP(-B$1*$A6)),** ou seja:

$$N_t = \frac{KN_0}{N_0 + (K - N_0)e^{-rt}}$$

O Excel vai buscar os valores dos parâmetros a B1:B3 e mede o instante de tempo na coluna A. Copie a fórmula e cole-a em B7:B65.

3) Faça o gráfico da logística, usando as colunas A6:A65 e B6:B65. Qual o valor para que a logística tende assintóticamente ? Experimente mudar o valor da célula B1 (mas só esse !) para 0.5, 1.0, 1.5. 1.7, ... e observe as mudanças que o gráfico sofre de cada vez.

2. Construção dum gerador de npa's N(μ, σ) e da logística estocástica

Assumamos agora que a população tem propensão para crescer logísticamente seguindo o nosso gráfico, mas variações ambientais e erros de medição, geram desvios nos valores de N_t observados. Vamos pressupor que, em qualquer t, estes desvios têm distribuição Normal(μ,σ) , definida por:

- média μ, igual ao valor da logística em t,
- desvio-padrão σ, proporcional a μ. Por exemplo: $\sigma = c\mu$, sendo c a constante de proporcionalidade

Antes de prosseguir, convém saber como gerar números pseudo-aleatórios (npa's) com distribuição N(μ,σ) no Excel. A seguinte função desempenha esse papel:

= NORMINV(RAND(); μ; σ)

Por exemplo, se escrevermos em qualquer célula do Excel

= NORMINV(RAND(); 10; 6) <enter>

gera-se um npa com distribuição Normal de média 10 e desvio-padrão 6.

Armados desta informação, prossigamos.

1) Em D3 introduza a constante de proporcionalidade c. Por exemplo, 0.1.
Em D6 introduza a fórmula

= NORMINV(RAND(), B6, D3*B6)

A fórmula gera um npa cuja média é o valor da logística (em B6) e cujo desvio-padrão é a média multiplicada por c.

Copiar e colar a fórmula nas células D7:D65. Obtem-se assim na coluna D uma sucessão de valores de N_t com uma tendência logística subjacente, mas afectados de erros normais - temos portanto uma logística estocástica.

2) Efectuar o gráfico, usando as colunas A6:A65 (para tempo em abcissas), B6:B65 (logística determinística) e D6:D65 (logística estocástica).

Faça variar c ligeiramente (em D3, c=0.05, 0.1, 0.2, 0.3), mantendo o resto constante. Quando c aumenta, a dispersão dos pontos da logística estocástica no gráfico tambem aumenta, espalhando-se em torno da logística determinística (porquê ?). Recorde o significado de c e certifique-se de que compreende o que se está a passar !

Suponha que c=0.2 e que observava esta população a partir de t=5 (células D36:D65), quando não é possível observar a fase ascendente da curva. Acha que reconhecia a existência dum modelo determinístico, subjacente à sequência de observações a partir de t=5 ?

Experimente colocar o cursor numa célula vazia do Excel e clique sucessivamente na tecla <delete>, observando o que se passa no gráfico. A nuvem de pontos vai mudando porque de cada vez que o Excel recalcula (causado pela tecla <delete>) atribui novos valores aos npa's da função RAND() das fórmulas na coluna D.

3. Detecção de regulação dependente da densidade (RDD)

Suponhamos que a sucessão de N_t em D6:D65 eram as observações duma população de que um ecologista dispõe. Como reconhecer a partir delas a existência de RDD (suponha que desconhecia que gerámos esta sucessão usando a equação logística) ?

Se houver RDD, quer o crescimento seja logístico quer não, é de esperar que a contribuição de cada indivíduo para o crescimento da população, em cada instante t, diminua quando N_t aumenta. No caso da logística, a relação entre o crescimento *per capita* (lado esquerdo da equação a seguir) e N, é esta:

$$\frac{1}{N}\frac{dN}{dt} = r - \frac{r}{K}N$$

Para intervalos de tempo pequenos, o lado esquerdo desta equação diferencial pode ser aproximada pela expressão equivalente referente ao intervalo Δt:

$$\frac{N_{t+1} - N_t}{\overline{N}\,\Delta t} = r - \frac{r}{K}N$$

sendo $\overline{N}$ um ponto médio do intervalo (N_t, N_{t+1}) e Δt o intervalo de tempo decorrido (relativamente ao intervalo unitário para o qual *r* está medido). Em geral, para $\overline{N}$, usa-se a média geométrica $N_{méd}=\sqrt{N_t N_{t+1}}$ e, no nosso exercício, Δt=0.2. Observando com cuidado a equação, os estudantes reconhecerão que representa uma recta com declive negativo *(-r/K)* e ordenada na origem, *r*. Temos portanto todos os ingredientes para *i)* detectar RDD e *ii)* estimar *r* e *K*.

Em E6, introduzir a fórmula **=(D7-D6)/((SQRT(D6*D7))*0.2)**, para calcular $(N_{t+1}-N_t)/N_{méd}\Delta t$. Copie e cole em E7:E64 (não ir até E65, pois aí já não existe N_{t+1}).

Faça o gráfico (em Excel, use a opção "scatter XY") de $(N_{t+1}-N_t)/N_{méd}\Delta t$ versus N_t usando as colunas D6:D64 e E6:E64.

Observe a grande dispersão de pontos. Apesar disso, deve ser possível reconhecer a existência duma tendência negativa, indicativa da presença de RDD.

Pode-se ajustar uma recta à nuvem de pontos (faremos isso adiante), concretizando assim uma forma geométrica para a tendência, mas não é legítimo testar se o declive da recta é significativamente diferente de 0 (porquê ?).

A nuvem de pontos que exprime a relação entre $(N_{t+1}-N_t)/ N_{méd}\Delta t$ e N_t foi obtida usando um modelo determinístico afectado de "ruído" Normal com um desvio-padrão relativamente pequeno (σ=0.2μ).

No mundo real, o ecologista pode esperar situações em que o número de observações de N é muito inferior ao utilizado neste exercício, a população não segue a simplicidade da logística e/ou a variabilidade ambiental pode causar erros com um desvio-padrão muito superior a 0.2μ. Tudo isto concorre para que haja grande dificuldade em detectar RDD na natureza.

4. Estimação de r e K.

A forma mais simples de estimar *r* e *K* é usar as funcionalidades da folha de cálculo para ajustar uma recta de regressão à nuvem de pontos do gráfico de $(N_{t+1}-N_t)/ N_{méd}\Delta t$ contra N_t. O ponto em que a recta corta as ordenadas é *r* e o declive da recta é *r/K.* Em Excel, para ajustar uma recta à nuvem, a forma mais simples de proceder é a seguinte:
1) Clique com o botão direito do rato sobre qualquer região da nuvem de pontos. Escolha a opção "Add Trendline".
2) Na caixa de diálogo, em "Trend/Regression Type", escolha Linear. Antes ainda de sair da caixa de diálogo, na parte de baixo do diálogo, marque os quadradinhos correspondentes a "Display equation on chart" e "Display R-squared value on chart", para que o Excel apresente a equação da recta e o coeficiente de determinação. Clique Close para sair do diálogo.

Se tudo correu bem, a recta está agora desenhada sobre a nuvem de pontos, bem como a respectiva equaçâo, de onde se poderá retirar *r* e *K.*

Compare os valores de *r* e *K* com os valores verdadeiros, que estão em B1 e B3.

Para obter várias réplicas deste gráfico e, portanto, de *r* e *K*, volte a colocar o cursor numa célula vazia e vá clicando em <delete>.

Experimente a situação em que não há variabilidade (coloque c=0). Verifique que, como seria de esperar, a recta fornece excelentes estimativas de *r* e *K* (embora subestime *r* ligeiramente). Experimente agora ir aumentando c devagar, e verifique como o aumento de variabilidade deteriora rápidamente as estimativas, em especial de *r*.

O coeficiente de determinação da recta, R^2, é o quadrado do coeficiente de correlação. R^2 deve ser interpretado como "a proporção de variância da nuvem de pontos em torno do seu valor médio que é explicada pela recta". Para os estudantes que se sintam mais à vontade com o coeficiente de correlação, podem calculá-lo, avaliando assim quão bem os pontos têm tendência linear.

Quer com R^2 quer com R, poderão constatar que os erros inerentes a *N* causam uma dispersão muito elevada, não obstante o determinismo do modelo logístico subjacente.

Em condições normais, poder-se-ia testar se o coeficiente de correlação é estatísticamente diferente de zero. Este teste, que é equivalente ao teste do coeficiente de regressão da recta, pressupõe, contudo, que as variáves X e Y da regressão são independentes, coisa que aqui não acontece. De facto, a variável em abcissas, N_t, está contida dentro da variável em ordenadas. Os estudantes devem recordar que, em situações de não-independência das variáveis X e Y, a recta e o coef. correlação podem ser calculados, mas os testes de significância não podem ser feitos, pois a hipótese nula tem pressupostos sobre a distribuição na amostragem de X e Y que não são respeitados em caso de não-independência.

Existem outras formas de investigar a presença de RDD por meio de estatísticas simplistas, como o coeficiente de regressão, mas em todas elas colocam-se problemas idênticos aos descritos, causados pelo habitual ruido aleatório dos dados ecológicos. A análise do problema requer um estudo faseado da dinâmica da população e técnicas estatísticas mais elaboradas. Há um livro, de nível médio-avançado, que é parcialmente dedicado ao assunto:

Royama, T. 1992. *Analytical Population Dynamics*. Chapman and Hall, London

Porém, sou de opinião que antes é recomendavel ter tido uma introdução à análise estatística de séries temporais. Na ausência de tempo ou paciência para ir fazer a cadeira ao dptm de estatística, o melhor livro introdutório é:

Chatfield, C. 2003 (6th ed.) *The Analysis of Time Series. An Introduction*. Chapman and Hall, London.