Stochastic Competing Species : with application to a bacteria experiment

Reg Kulperger

University of Western Ontario

kulperger@stats.uwo.ca

February, 2008

Joint work with

- Zengjing Chen, Shandong University
  Journal Applied Probability, 2005
- Severien Nkurunziza, University of Windsor

Mathematical modeling :

- relationships between variables (on average)

- (time evolution) dynamics

Stochastic processes and statistical modelling

- relationships between variables and their variability

- (time evolution) dynamics with variability (stochastics)

A simple and classical biological model of two species growth

x_t and y_t population sizes

$$\begin{aligned} \frac{dx_t}{dt} &= x_t(b_{1,1} - b_{1,2}x_t - b_{1,3}y_t) \\ \frac{dy_t}{dt} &= y_t(b_{2,1} - b_{2,2}x_t - b_{2,3}y_t) \end{aligned} \tag{1}$$

Models population average dynamics and growth

Special case : Lotka Volterra model

prey $= x_t$ and predator $= y_t$

$$\begin{aligned} \frac{dx_t}{dt} &= x_t(\mu - \lambda y_t) \\ \frac{dy_t}{dt} &= y_t(-\alpha + \beta x_t) \end{aligned} \tag{2}$$

Long history of studying this model and variations since the 1930's.

Need to incorporate randomness

Method 1 : regression model
Nkurunziza : observe DE + noise
keeps cyclic behaviour
long term stability, never die out

Method II :
DE models average growth rate per individual
make growth rate random
- changes dynamics
- no longer cyclic
- may no longer be stable (ergodic)

Paths of (2) are deterministic

$$F(x,y) = \beta x - \alpha \log(x) + \lambda y - \mu \log(y)$$

It can be shown

$$\frac{dF(x,y)}{dt} = \frac{dx}{dt}\frac{dF(x,y)}{dx} + \frac{dy}{dt}\frac{dF(x,y)}{dy} = 0$$

Solution to (2) lives on a contour $F(x,y) = c$

$$c = F(x_0, y_0)$$

Solution : Deterministic , Cyclic; Figure 1

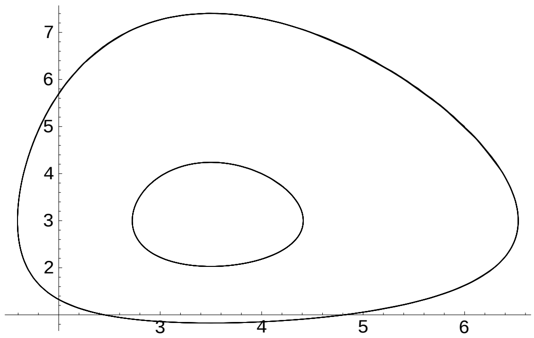

Figure 1: Two Contours of (2) with $(\mu, \lambda, \alpha, \beta)$ = (3,1,7,2) and (x_0, y_0) = $(4.5, 3.5)$ and (6.5,4.5)

Regression extension

$$(\log(X_t), \log(X_t)) = (\log(x_t), \log(y_t)) + \left(\epsilon_t^X, \epsilon_t^Y\right)$$

Observed data (X_t, Y_t)

Noise : iid Gaussian noise or Ornstein Uhlenbeck process (continuous time analogue of autoregressive order 1 process).

System maintains the periodicity with this stationary noise, but the deterministic path is masked by noise.

Random growth rate extension

Linear growth rate per individual is an approximation to population average rate

a natural extension is $\rightarrow$ add noise to growth rates

$$
\begin{aligned}
dx_t &= x_t\left([\mu - \lambda y_t]dt + \sigma_1 dW_t^{(1)}\right) \\
&= x_t(\mu - \lambda y_t)dt + \sigma_1 x_t dW_t^{(1)}
\end{aligned}
$$

$$
\begin{aligned}
dy_t &= y_t\left([-\alpha + \beta x_t]dt + \sigma_2 dW_t^{(2)}\right) \\
&= y_t(-\alpha + \beta x_t)dt + \sigma_2 y_t dW_t^{(2)}
\end{aligned}
$$

Renshaw (1991) - biological models

Mao, Marion and Renshaw (2002) - existence of solution ; when diffusion term is $\sigma x^m dW_t$ and $m = 2$; restriction due to choice of *test function*

can be extended to $m > \frac{3}{2}$ but not for $m = 1$.

Gard and Kannan (1976) - restricted form of ergodicity conditional on solution bounded; that is not true for the noisy rate model

$F_t = F(x_t, y_t)$ (contour function)
apply Itô's Lemma

$$\begin{aligned} dF_t &= \tfrac{1}{2}\left(\alpha\sigma_1^2 + \mu\sigma_2^2\right) dt \\ &\quad + \sigma_1\left(\beta x_t - \alpha\right) dW_t^{(1)} \\ &\quad + \sigma_2\left(\lambda y_t - \mu\right) dW_t^{(2)} \\ &= \tfrac{1}{2}\left(\alpha\sigma_1^2 + \mu\sigma_2^2\right) dt + dM_t \end{aligned} \tag{3}$$

M_t is zero mean martingale

$$\begin{aligned} \mathrm{E}\left(F(x_t, y_t)\right) &= F(x_0, y_0) + \frac{1}{2}\left(\alpha\sigma_1^2 + \mu\sigma_2^2\right) t \\ &\to \infty \text{ as } t \to \infty \end{aligned}$$

Thus $(x_t, y_t) \to \partial R_+^2$ (in mean)

Not ergodic (stable) with ANY NOISE

Competing species

$$\begin{aligned} dx_t &= x_t(\mu - \lambda y_t + \gamma x_t)dt + \sigma_1 x_t dW_t^{(1)} \\ dy_t &= y_t(-\alpha + \beta x_t + \delta y_t)dt + \sigma_2 y_t dW_t^{(2)} \end{aligned} \tag{4}$$

self damping : $\gamma < 0$, $\delta < 0$

Discussions in biological literature : prey-predator ($\gamma < 0$, $\delta < 0$) has draw back of unlimited linear growth

competing species model corrects for this

Lemma 1 *$\epsilon_1' = \frac{\gamma}{\beta}$ and $\epsilon_2' = \frac{\delta}{\lambda}$.*

Rewrite competing species model

$$\begin{aligned} dx_t &= x_t\left[\mu' - \lambda y_t + \epsilon_1'(\beta x_t - \alpha')\right]dt \\ &\quad + \sigma_1 x_t dW_t^{(1)} \\ dy_t &= y_t\left[-\alpha' + \beta x_t + \epsilon_2'(\lambda y_t - \mu')\right]dt \\ &\quad + \sigma_2 y_t dW_t^{(2)} \end{aligned}$$

If either

1. $\gamma > 0$ and $0 < \delta < \frac{\alpha\lambda}{\mu}$ or

2. $\delta < 0$ and $-\frac{\mu\beta}{\alpha} < \gamma < 0$ (biological interest)

then $\mu' > 0$ and $\alpha' > 0$.

Theorem 1 *Consider*

$$\begin{aligned} dx_t &= x_t\left[\mu - \lambda y_t + \epsilon_1(\beta x_t - \alpha)\right] dt \\ &\quad + \sigma_1 x_t dW_t^{(1)} \\ dy_t &= y_t\left[-\alpha + \beta x_t + \epsilon_2(\lambda y_t - \mu)\right] dt \\ &\quad + \sigma_2 y_t dW_t^{(2)} \end{aligned} \tag{5}$$

initial condition $(x, y) \in R_+^2$.

If $\epsilon_2 < 0$, $\epsilon_1 < 0$ *and also*

$$0 < \frac{\sigma_1^2 \alpha + \sigma_2^2 \mu}{2\alpha^2} < -\epsilon_1 = |\epsilon_1|$$

and

$$0 < \frac{\sigma_1^2 \alpha + \sigma_2^2 \mu}{2\mu^2} < -\epsilon_2 = |\epsilon_2|$$

then the solution to the SDE (5) is ergodic.

If these conditions are violated the system can be transient

Gard (1980's - 1990's) studies recurrence and ergodicity in terms of differential operators and STRONG conditions (bounded solution)

Bhattacharya (1978) studies sufficient conditions for ergodicity for diffusions on R^N based on test function $H(r_t)$ and $r_t = ||\mathbf{x}_t||$

Adapt his method for showing ergodicity :

1. choose *distance function* F to play role of Euclidean distance

   $F \geq 0$ and $F \to \infty$ as $(x, y) \to \partial R^2_+$

2. use F to show (x_t, y_t) is recurrent; may need to find function H and use $H(F(X_t))$

3. ergodicity

$$dF_t = A(x_t, y_t)dt + dMt$$

M zero mean martingale; A drift function

- A continuous
- There exists D such that
  (i) $A(x, y) \leq -d_0 < 0$ on D^c and
  (ii)
  $$\sup_{(x,y)\in\bar{D}} A(x, y) \leq a < \infty$$
  (trivial if F and derivatives are continuous)

4. K compact and arbitrary interior to D (may enlarge D)

$$\begin{aligned}\tau_K &= \inf\{t \geq 0 : (x_t, y_t) \in K\} \\ &= \left(\tau_D + \tau_K^{(x_{\tau_D}, y_{\tau_D})}\right) \mathrm{I}(\tau_D < \infty)\end{aligned}$$

Note $\mathrm{I}(\tau_D < \infty) = 1$ almost surely

$\Rightarrow$ ***recurrence***

As part of the recurrence we show $\mathrm{E}^{x_0,y_0}(\tau_D) < \infty$ for all x_0, y_0 (uses continuity of A ; smoothness of F)

Conditional expectation and Markov property

$$\begin{aligned}\mathrm{E}^{x_0,y_0}(\tau_K) &= \mathrm{E}^{x_0,y_0}(\tau_D) + \mathrm{E}^{x_0,y_0}(\mathrm{E}^{x_{\tau_D}, y_{\tau_D}}(\tau_K | \mathcal{F}_{\tau_D})) \\ &\leq \mathrm{E}^{x_0,y_0}(\tau_D) + \sup_{(x,y)\in\partial D} \mathrm{E}^{x,y}(\tau_K)\end{aligned}$$

Proof Theorem 1 is based on a simple geometric property

$$F(x,y) = \beta x - \alpha \log(x) + \lambda y - \mu \log(y) + c$$

$c = \alpha - \alpha \log(\alpha/\beta) + \mu - \mu \log(\mu/\lambda)$

$F(x,y) \geq 0$

Apply Ito's Lemma

$$\begin{aligned} dF_t &= \left[\epsilon_1 \left(\beta x_t - \alpha\right)^2 + \epsilon_2 \left(\lambda y_t - \mu\right)^2 + a\right] dt \\ &\quad + dM_t \\ a &= \tfrac{1}{2}\left(\sigma_1^2 \alpha + \sigma_2^2 \mu\right) \end{aligned} \tag{6}$$

$$dM_t = \sigma_1 \left(\beta x_t - \alpha\right) dW_t^{(1)} + \sigma_2 \left(\lambda y_t - \mu\right) dW_t^{(2)}$$

M_t is zero mean martingale

$$A(x,y) = a - \left[|\epsilon_1| \left(\beta x - \alpha\right)^2 + |\epsilon_2| \left(\lambda y - \mu\right)^2\right]$$

$\epsilon_i < 0$ for the damping condition in the Theorem

$$D = \left\{ (x,y) : \frac{\left(x - \frac{\alpha}{\beta}\right)^2}{\frac{a}{|\epsilon_1|\beta^2}} + \frac{\left(y - \frac{\mu}{\lambda}\right)^2}{\frac{a}{|\epsilon_2|\lambda^2}} \ \le \ 1 \right\} \ .$$

Need $D \subset R_+^2$; hence condition in Theorem has uniform bound on negative drift outside D, forces process back to centre.

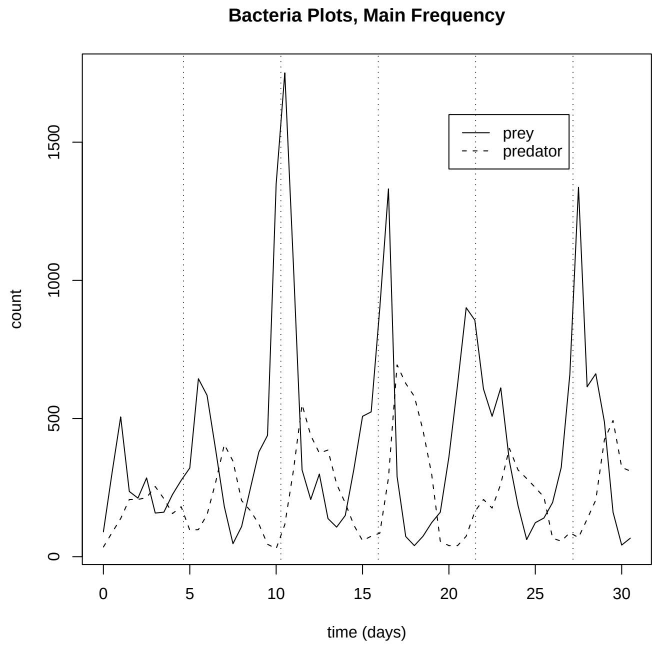

Figure 2: Bacteria Data with "periodic"

Luckinbill (1973) *Ecology*, 56, 1320-1327

Paramecium caudatium (prey)

Didinium nasutum (predator)

62 observations over 30 days

Compute periodograms. From plot find predominant frequency

Figure 2 shows the data with "period"

If model were of regression type process would stay in "phase"

Aside : Annual Seasonal model versus autoregressive model AR(12)

First stays in phase high levels every 12 months

Second gets out of phase after a random amount of time

Data is not periodic as with regression model

Model $(x, y) = $ (prey, predator)

Time t measured in years; day $= 1/365$

$$
\begin{aligned}
d\log(x(t)) &= \left(\mu - \lambda y(t) + \gamma x(t) - \frac{\sigma_1^2}{2}\right) dt + \sigma_1 dW_1(dt) \\
d\log(y(t)) &= \left(-\alpha + \beta x(t) + \delta y(t) - \frac{\sigma_2^2}{2}\right) dt + \sigma_2 dW_2(dt)
\end{aligned}
$$

Estimation

Use continuous time likelihood function

Discretize the estimator to produce estimators

Estimates based on first 52 data points (leave 10 for out of sample study)

μ	λ	γ	σ_1^2
0.10851	0.02813	-0.07361	0.16760
α	β	δ	σ_2^2
0.02870	0.21144	-0.56752	0.11540

The condition for ergodicity fails so this process may be unstable

One step ahead prediction

Conditional distribution of $(\log(x((j+1)\Delta), \log(y((j+1)\Delta))$ given data up to time $j\Delta$

Either solve forward equations or use parametric bootstrap

Use Euler approximation with small time increment size of Δ/m

m= 5 or 10 gives a stable answer, so we use m = 5

Parametric bootsrap with many (1000) replicates to estimate conditional quantiles

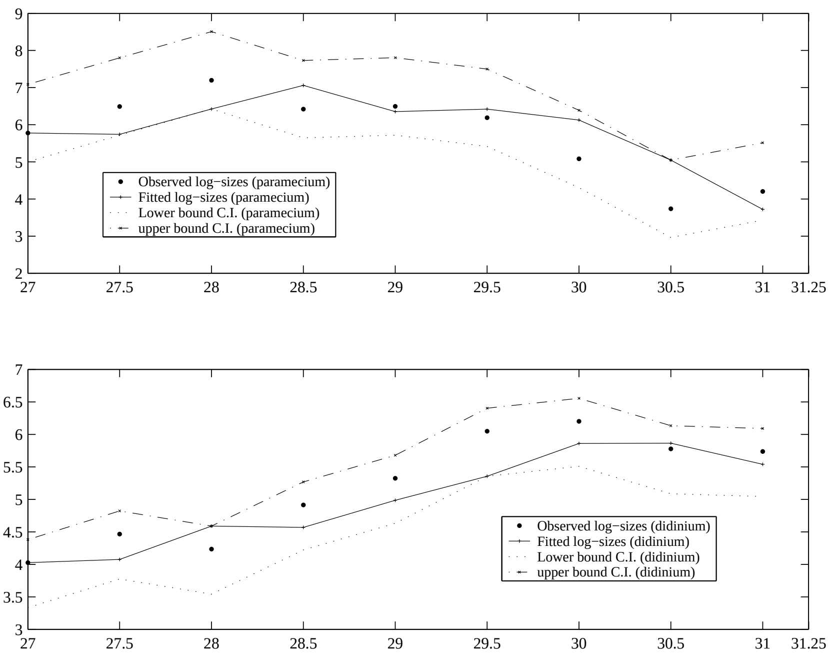

Figure 3: Observed log size, one-step-ahead prediction, and .95 prediction interval (Out-of-sample)

Prediction Intervals in and out of sample based on parametric boostrap

This is an in sample prediction

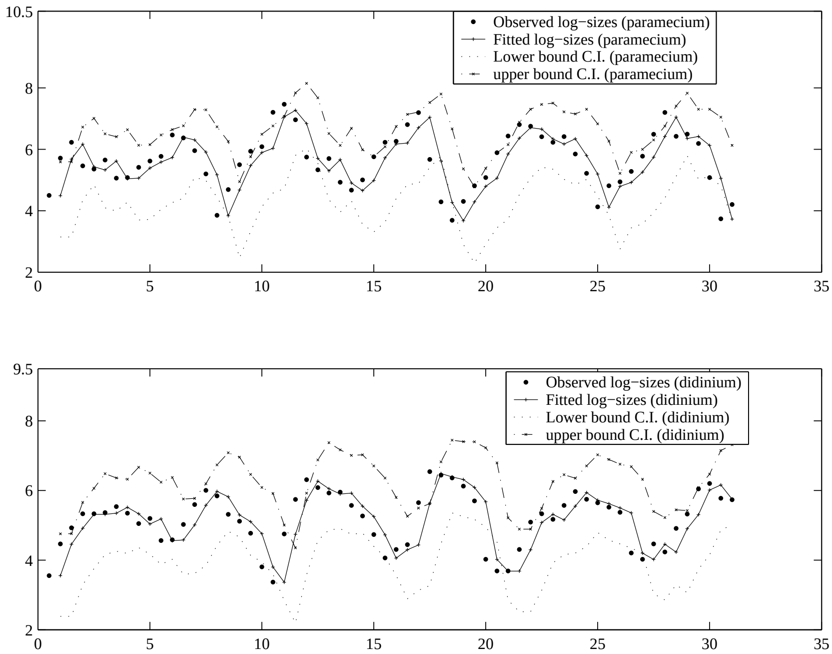

Figure 4: Observed log size, one-step-ahead prediction, and .95 prediction interval (Out-of-sample)