Surrogate models

- Alternative approach to likelihood-free inference is to construct surrogate models for various parts of the system
 - Synthetic likelihood is one of the first surrogate methods
 - At θ approximate $p(x\mid\theta)\approx \text{Normal}(x\mid\hat{\mu}_N(\theta),\hat{\Sigma}_N(\theta)) \text{ with an sample of size }N \text{ drawn from the model}$

Statistical inference for noisy nonlinear ecological dynamic systems Simon N. Wood¹ analysis. Systems with near-chaotic dynamics are little better. Bayesian inference would require that we sample replicate vectors roceases plus demographic and environmental process noise, and in a meaningful way for an 6, as irregular as that shown in Fig. 1b, c are only observable with error. Their sensitivity to history means parameters, will cause drastic changes in the system trajectory. This reproduce the exact course of the observed data in a way that the real of the observable data and the process noise, rendering it useless as the basis for obtaining measures of statistical fit. Because the joint density is the basis for the fit measures used by all conventional feature, which should not contribute to any measure of match or statistical methods2, this is a major theoretical shortcoming. The dynamic models in the chaotic and near-chaotic regimes, other than — judge model fit using statistics that reflect what is dynamically import mantitutive validation that are essential tools in the rest of bio-idea is not new *11. What is new is the ability to assess the consistency of ogical science. Here I show that this impasse can be resolved in a statistics of the model simulations and data without recourse to ad ho. measures of that consistency, but in a way that instead gives access to imple and general manner, using a method that requires only the ability to simulate the observed data on a system from the dynamic — much of the machinery of likelihood-based statistical inference'. The first step in the proposed analysis is therefore to reduce the raw reduced to phase-insensitive summary statistics, quantifying local observed data, y, to a vector of summary statistics, s, designed to cused to obtain the mean and the covariance matrix of the statistics, given model parameters, allowing the construction of a 'synthetic using a straightforward Markov chain Monte Carlo sampler, but one ence. I apply the method to establish the dynamic nature of the uctuations in Nicholson's classic blowfly experiments scaled Ricker map*, describing the time course of a population N_cby $N_{t+1} = rN_te^{-K_t+r_t}$ where the e are independent N(0, of) 'process noise' terms (assumed to be environmental noise here, for illustrative purposes) and r is an nsic growth rate parameter controlling the model dynamics. This odel amply illustrates the collapse of standard statistical methods in the face of chaotic or near-chaotic dynamics. Figure 1a shows data from a realization of equation (1) when log(r) = 3.8, and what is observed are Poisson random deviates, y_0 with mean ϕN_1 (where ϕ a scale parameter), reflecting a reasonably common sampling situation. Suppose that the aim is to make statistical inferences about $\theta^{\top} = (\mathbf{r}, \sigma_c^2, \phi)$ from this data series. Figure 1b, c illustrates the joint robability (density) function, f₀(y, e), of the data vector, y, and noise (log(r) = 3.5, σ = 0.3, $\dot{\rho}$ = 10). b, The log joint probability density ector, e, when the (fixed) noise realization and data from the simulog(f,(y, e)), of data, y, and random process noise terms, e, plotted agains the value of the first process noise deviate, e_p, with the rest of e and y bek lation in Fig. 1a are plugged in Figure 1c shows how log(fs) varies with r, whereas in Fig. 1h r is kept fixed but the first element of the fixed, e, Log(fe(y, e)) plotted against model parameter r, again with e and noise realization, e_1 , is varied. Likelihood-based inference about θ With ematical Sciences, University of Bath, Bath BA2 2AY, UK. 68(1) Macmillan Publishers Limited. All rights reserved

Example

Construct approximate likelihood at an arbitrary parameter value

Set of samples Gaussian empirical estimate at θ $\approx p(\theta \mid x^o)$