Likelihood-free Inference

What to do when observation is high dimensional?

- Use summary statistics
- Sufficient statistics generally not available
 - Strategies for automatic generation/selection
 - Often bespoke construction via domain/simulator expertise and prior simulations

Synthetic likelihood

• At θ approximate $p(x \mid \theta) \approx \text{Normal}(x \mid \hat{\mu}_N(\theta), \hat{\Sigma}_N(\theta))$ with an sample of

size N drawn from the model

• Naturally combines with e.g. MCMC

Vol 466[26 August 2010]dob10.1038/nature09319 Statistical inference for noisy nonlinear ecological dynamic systems Simon N. Wood1 Chaotic ecological dynamic systems defy conventional statistical intractable and, from Fig. 1b, is numerically intractable as well." analysis. Systems with near-chaotic dynamics are little better. Bayesian inference would require that we sample replicate vectors Such systems are almost invariably driven by endogenous dynamic e and θ from a density proportional to f_θ ; no methods exist to do this processes plus demographic and environmental process noise, and f_θ in a meaningful way for an f_θ as irregular as that shown in Fig. 1b, c. that minute changes in the driving noise realization, or the system — sophical. Naive methods of statistical inference try to make the model parameters, will cause drastic changes in the system trajectory. This reproduce the exact course of the observed data in a way that the real sensitivity is inherited and emplified by the joint probability density of the observable data and the process noise, rendering it useless as system itself would not do if repeated. Although the dynamic processe driving the system are a repeatable feature about which inferences 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 inability to make well-founded statistical inferences about biological to correspond with what is scientifically meaningful, it is necessary to dynamic models in the chaotic and near-chaotic regimes, other than judge model fit using statistics that reflect what is dynamically importon an ad hec basis, leaves dynamic theory without the methods of ant in the data, and to discard the details of local phase. In itself, this quantitative validation that are essential tools in the rest of bio-idea is not new 1.1. What is new is the ability to assess the consistency of ogical science. Here I show that this impasse can be resolved in a simple and general manner, using a method that requires only the measures of that consistency, but in a way that instead gives access to ability to simulate the observed data on a system from the dynamic — much of the machinery of likelihood-based statistical infevence'. model about which inferences are required. The raw data series are The first step in the proposed analysis is therefore to reduce the ray reduced to phase-insensitive summary statistics, quantifying local observed data, y, to a vector of summary statistics, s. designed to dynamic structure and the distribution of observations. Simulation is used to obtain the mean and the covariance matrix of the statistics, given model parameters, allowing the construction of a 'synthetic likelihood' that awesses model fit. This likelihood can be explored using a straightforward Markov chain Monte Carlo sampler, but one ence. I apply the method to establish the dynamic nature of the fluctuations in Nicholson's classic blowfly experiments --The prototypic ecological model with complex dynamics is the scaled Ricker map⁴, describing the time course of a population N_c by where the e_i are independent N(0, σ_i^2) 'process noise' terms (assumed to be environmental noise here, for illustrative purposes) and r is an intrinsic growth rate parameter controlling the model dynamics. This 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_t (where ϕ is a scale parameter), reflecting a reasonably common sampling situation. Suppose that the aim is to make statistical inferences about

 $\theta^- = (r, \sigma_r^2, \phi)$ from this data series. Figure 1b, c illustrates the joint probability (density) function, $f_0(\mathbf{y}, \mathbf{c})$, of the data vector, \mathbf{y} , and noise vector, \mathbf{e} , when the (fixed) noise realization and data from the structure in Fig. 1a are plugged in. Figure 1c shows how $\log(f_0)$ varies with \mathbf{r} , whereas in Fig. 1b ris kept fixed but the first element of the neise realization, e_i , is varied. Elselihood-based inference about θ yhid fixed, θ , by plotted against took θ , and θ point θ , and θ point θ pointed against the value of the first process noise deviate, e_i with the rest of θ and θ below θ plotted against the value of the first process noise deviate, θ , with the rest of θ and θ below θ .

requires that we integrate f_0 over all e, something that is analytically Ricker model and the data given in a $(N_1 - 500)$