A subordinated stochastic process model

Ana Paula Palacios, J. Miguel Marín and Michael P. Wiper

Abstract We introduce a new stochastic model for non-decreasing processes which can be used to include stochastic variability into any deterministic growth function via subordination. This model is useful in many applications such as growth curves (children's height, fish length, diameter of trees, etc) and degradation processes (crack size, wheel degradation, laser light, etc). One advantage of our approach is to be able to easily deal with data that are irregularly spaced in time or different curves that are observed at different moments of time. With the use of simulations and applications, we examine two approaches to Bayesian inference for our model: the first based on a Gibbs sampler and the second based on approximate Bayesian computation.

1 Introduction

Growth processes are usually described using discrete time models where the mean function is deterministic and a stochastic element is introduced via an additive, random noise component. An alternative is to consider continuous time modelling. In the literature some stochastic growth models are proposed using stochastic differential equations to model the variations ([1, 2]). However, the solution of these equations are not monotonically increasing and therefore can fail to model non-decreasing growth processes such as children's height, fish size or crack length. In this work, we introduce a new stochastic model for non-decreasing processes that

Ana Paula Palacios
Plymouth University, School of Computing and Mathematics, e-mail: ana.palacios@plymouth.ac.uk

J. Miguel Marín
Universidad Carlos III de Madrid, Department of Statistics, e-mail: jmmarin@est-econ.uc3m.es

Michael P. Wiper
Universidad Carlos III de Madrid, Department of Statistics, e-mail: m.wiper@est-econ.uc3m.es

overcomes this limitation. This model can be used to include stochastic variability into any deterministic growth function via subordination. The main features of this model are that its paths are non-decreasing everywhere and, in a particular case, the mean function of the process is equal to the deterministic growth function used as time change.

2 The model

The growth model that we proposed is built upon an homogeneous, continuous time Markov process. It is commonly observed that growth in biological processes, and also wear in degradation processes, does not occur continuously. In contrast, growth or damage occur per intervals of time. Even more, the growth velocity can also fluctuate. To represent this discontinuous growth we start our model with a time homogeneous, Markov process $\{U_t : t \geq 0\}$, where the states represent levels of growth rate. Transitions can only occur between neighbours. That is, if at time t the process is in state i, after an exponential amount of time, then it moves to either of the neighbouring states $i \to i+1$ or $i \to i-1$. This allows us to represent possible fluctuations in the growth rate but without abrupt changes. The process U_t is uniquely determined by the generator matrix, $\mathbf{Q}$, and the initial distribution of the process, ν_0. The transition rate matrix $\mathbf{Q}$ is a tri-diagonal matrix with parameters $\alpha > 0$, the instantaneous up-jump rate and $\beta > 0$, the instantaneous down-jump rate. Let $S = \{a + ib; i = 0, \ldots, k\}$ be the state space, where $a \geq 0$ is the minimum state value, b is a jump size and $(k+1)$ is the number of states.

Now we define a continuous state process, $\{V_t : t \geq 0\}$, such that

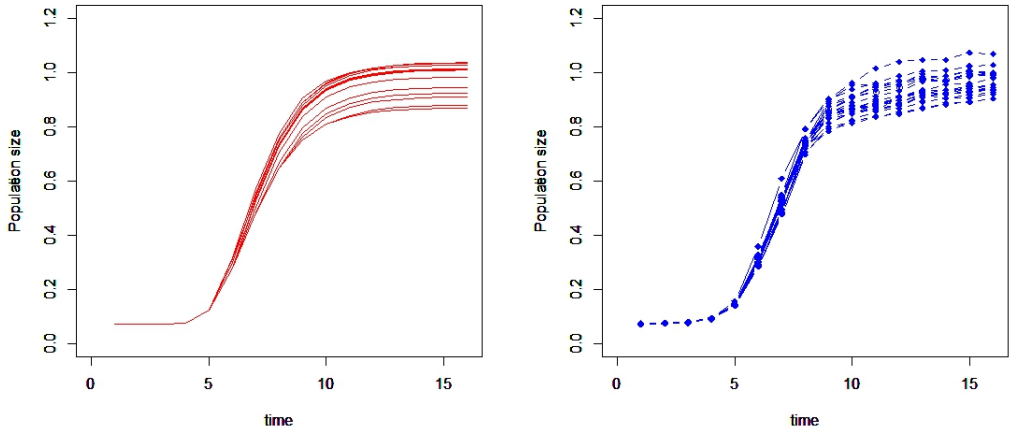
$$V_t = \int_0^t U_s ds. \tag{1}$$

This is a non-decreasing, continuous time process which, being the integral of the growth rate, represents the total growth. Realizations of V_t are the path integrals of a simple stochastic process and their trajectories are piece-wise linear.

Beyond the growth fluctuations described earlier, most of these process are also characterised by growth stages. For example, many growth processes show an S-shape curve, where an initial (almost) steady period is followed by an exponential growth before a decelaration of the growth rate. We introduce these different stages in the model by manipulating the time (see [3] for time change and subordination), e.g. when the growth is slow, we want the time to slow down but when the growth is exponential we want the time to speed up. Thus, the time velocity will be governed by a deterministic non-decreasing function. We define our stochastic growth process, $\{Y_t : t \geq 0\}$ to be the continuous time stochastic process with continuous state space, defined as

$$Y_t = V_{G(t)}, \tag{2}$$

where $G(t)$ is any deterministic non-decreasing function.



(a) Several realizations (b) Replications

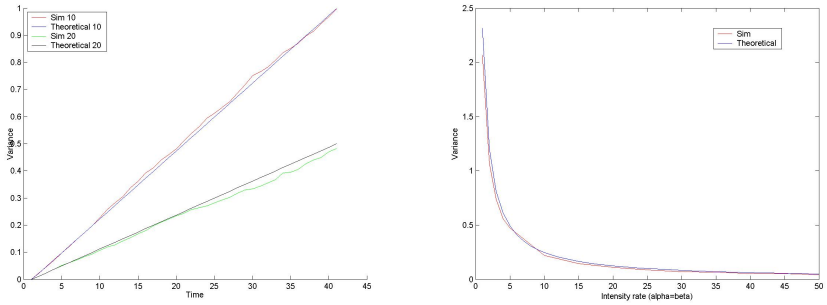
Fig. 1 Simulated realizations and real bacterial growth curves

We show that assuming stationary state of the Markov process, then the mean function of the process Y_t is proportional to the time change function G: $E[Y_t] = \mu G(t)$, where μ is the constant mean of the Markov process in stationary state. As a particular case, if a proportion J of the function G is used as time change and $J = 1/\mu$, then $E[Y_t] = G(t)$. This fact suggests using a standard parametric model as the mean function of the process. For example, for population growth a logistic function could be used; for fish size the von Bertalanffy growth function could be used, etc. Figure 2 shows, on the right, 20 replications of experimental bacterial growth curves and, on the left, 20 realizations of the process Y_t when using the Gompertz function as time change.

The variance of the process V_t is linearly increasing with time and its magnitude depends on the instantaneous intensity rates of the Markov process. The greater the intensity rates, the lower the variance.

3 Bayesian inference

Assume that we observe the population size, for example of bacteria, $y_{t_1} < ... < y_{t_n}$ at a sequence of time points, $0 < t_1 < ... < t_n$ say. Then in the model defined by (2), the likelihood function is analytically unavailable which implies that frequentist approaches are infeasible. But for the case of two states $(0,1)$ and equal up and down transition rates in the Markov process U_t, we can find an explicit expression for the likelihood when conditioning on the initial state and the number of jumps in suc-



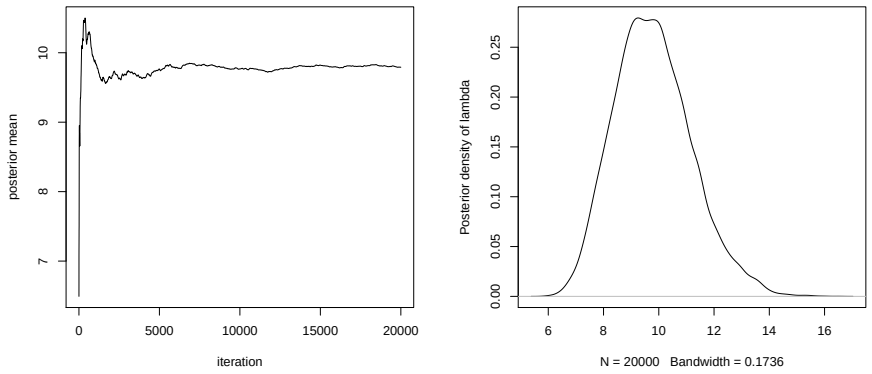
(a) Theoretical (black lines) and simulated variance for $\alpha = \beta = 10$ (red line) and $\alpha = \beta = 20$ line) (green line).

(b) Theoretical (black line) and simulated (red variance for $\alpha = \beta$.

Fig. 2 The variance as a function of time and the intensity rates

cessive time intervals. The difference $\bar{y}_i = y_{t_i} - y_{t_{i-1}}$ is equal to the total time spent in state 1 during the transformed time interval $T_i = G(t_i) - G(t_{i-1})$. Conditioning on the number of jumps, it is possible to show that the distribution of $\bar{y}_i$ is equal to the distribution of order statistics of a Uniform $(0,T)$ where T is the length interval. Then, the likelihood of $\bar{y}_i$, $f(\bar{y}_i|\lambda, s_1, N_i)$, conditional to the intensity rate of jumps $\lambda (= \alpha = \beta)$, the initial state of the Markov process s_1 and the number of jumps N_i in interval i, follows a scaled Beta distribution. This allows for the implementation of a Gibbs sampling algorithm (see [4]). Estimating in a first step the parameters of the time change function, and assuming them known in a second step, the conditional posterior distributions of λ and s_1 can be derived analytically. However, the posterior distribution of N_i does not have a closed form, and a Metropolis-Hasting step is necessary.

To illustrate this approach a data set was simulated for given values of the Gompertz function. Assuming $\lambda = 10$, five replications were generated with 20 observations per curve. The Gibbs sampling was performed to estimate the intensity rate and the results are shown in Figure 3. The cumulative mean has converged and the posterior distribution of λ is centred around the true value. The posterior mean is equal to 9.791 and the median is 9.705. The 95% credible interval is equal to $(7.36, 12.85)$.



(a) Cumulative mean (b) Posterior distribution of the intensity rate

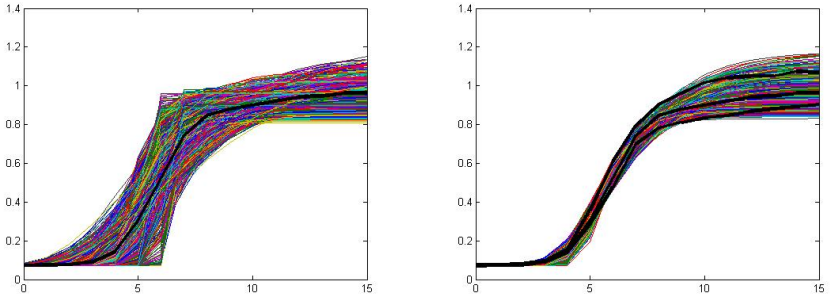
Fig. 3 Gibbs sampling

For the more general case, with multiple states in the Markov chain, free-likelihood methods can be applied. We illustrate this approach with a naive example. We apply the ABC method to the real bacterial data of 2 to estimate the parameters of the time change function and the intensity rate (see [5]) . The simplest rejection-ABC algorithm was implemented, assuming informative prior distributions for the Gompertz parameters. We compared one to one the simulated data with the mean observed data and summed all the distances for each observation in a curve. When computing the distances $|y_i^{sim} - y_i^{obs}|$ we penalised more for departures at earlier times (when less variability between curves is normally observed). Finally, we kept the 1% best of the sampled parameter values, i.e. the ones which minimise the distance. Figure 4 illustrates the results. The posterior mean values for the Gompertz parameter were all biologically reasonable.

4 Conclusions

The aim of our work was to propose a new stochastic model suitable for growth processes and degradation data. Thus, the model developed shows two nice features. First, the growth paths are non-decreasing making the model feasible for a wide variety of phenomena such as crack size, fisheries or human growth. Second, as a particular case of the model, the mean function of the process is equal to the parametric function governing the time change. Another advantage of our approach is to be able to easily deal with data that are irregularly spaced in time or different curves that are observed at different moments of time. Finally, we have shown with the use of simulations and applications, two possible Bayesian approaches to fit the model, Gibbs sampling and approximate Bayesian computation. Results were good in both

cases, however, further work in this direction is needed. For example, it would be very interesting to develop a full Gibbs sampling approach where all parameters are estimated in only one step. Additionally, more efficient ABC algorithms could be applied to allow for non-informative prior distributions.



(a) Whole set of generated curves

(b) The best 1% from the ABC sampler. Thick black lines represent the maximun, the minimun and the mean observed curves.

Fig. 4 ABC algorithm

References

1. Donnet, S., Foulley, J. and Samson, A. "Bayesian Analysis of Growth Curves Using Mixed Models Definied by Stochastic Differential Equations." *Biometrics*, **66**, 733–741 (2010).
2. Rensaw, E. *Stochastic Population processes: Analysis, Approximations, Simulations*. Oxford University Press, Oxford (2011)..
3. Sato KI. *Lévy processes and infinitely divisible distributions*. Cambridge university press (1999).
4. Robert Christian P. and Casella, George *Monte Carlo statistical methods. 2nd Edition*. Springer, New York (2005).
5. Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems." *Journal of the Royal Society Interface* **6.31**, : 187–202 (2009).