Workshop on

Computational Stochastics

January 17–21, 2000

Department of Mathematical Sciences

University of Aarhus

Introduction

Computational stochastics is a new and expanding area of stochastics, dealing with computational methods of analyzing complex mathematical and statistical models. This workshop intended to reveal and discuss the potential strength and impact of this new discipline in a variety of applications, including chemistry, finance, genetics, medical imaging, molecular biology and physics.

The workshop was held at the Department of Mathematical Sciences, University of Aarhus, and was organized by StocLab (Laboratory for Computational Stochastics) and MaPhySto (Centre for Mathematical Physics and Stochastics), University of Aarhus.

In this booklet we have collected brief accounts of the subjects of the talks given during the workshop. Furthermore, at the end of the booklet, the programme and the list of participants of the workshop are included.

We wish to thank all participants — the speakers in particular — for contributing to the Workshop.

Søren Asmussen and and Eva B. Vedel Jensen
Aarhus, March 2000.

Contents

1 Abstracts of Talks **3**

Søren Asmussen 3

Adrian Baddeley 5

Laird Breyer 6

Ole F. Christensen (with J. Møller and R. P. Waagepetersen) 7

Ian Dryden 9

Günter Döge 12

Paul Glasserman (with P. Heidelberger and P. Shahabuddin) 13

Paul Glasserman (with M. Broadie) 15

Niels Væver Hartvig 16

Jotun Hein (with B. Knudsen) 18

Anders Krogh 19

Ole G. Mouritsen 20

Tomáš Mrkvička 22

Søren Feodor Nielsen 23

Olivier Perrin (with S. Iovleff) 25

Fabio Spizzichino 39

Matthew Stephens 43

William Stewart 44

Dietrich Stoyan 45

Cristina Zucca (with M.T. Giraudo and L. Sacerdote) 46

2 Workshop Program **53**

3 List of participants **57**

1 Abstracts of Talks

Søren Asmussen

Lund University

Matrix-analytic algorithms for many–server queues

ABSTRACT: Starting from a survey of the matrix–analytic method in applied probability, we gradually specialize and end up with presenting a new algorithm for computing the waiting distribution in a many–server queue with phase–type service times (in fact, this algorithm is the first complete solution of the waiting time problem in say GI/PH/c queues).

The classical set–up of the matrix–analytic area is bivariate Markov chains (J_n, L_n) where J_n (the phase) has a finite number of values and L_n (the level) has values in $\{0, 1, 2, \dots\}$. The transition matrix P may have one of two forms, GI/M/1 type or M/G/1 type. For example, in the GI/M/1 case

$$P = \begin{pmatrix} B_0 & A_0 & 0 & 0 & \\ B_1 & A_1 & A_0 & 0 & \\ B_2 & A_2 & A_1 & A_0 & \\ \vdots & & & & \ddots \end{pmatrix}$$

where the A_k and B_k are blocks (the block–partitioning corresponds to levels). The focus is on the computation of the stationary distribution, which in similar partitioning can be written in the form $\pi_k = \pi_0 R^k$ for some matrix R, given as solution of the fixpoint problem

$$R = A_0 + RA_1 + R^2 A_2 + \cdots$$

which is usually solved by iteration. See Neuts (1981), Neuts (1989) and Latouche & Ramaswami (1999) for surveys, which include also similar models in continuous time and many examples; a typical application is queues with phase–type services and arrivals governed by a finite Markov process (phase–type distributions are defined as absorbtion time distributions in finite Markov processes).

For continuous–valued processes like waiting times, a parallel theory was developed by Sengupta (1989), who as his main application computed the waiting time distribution in $GI/PH/1$ queues. In this setting, the process (J_t, L_t) is obtained by piecing busy periods together, and the level L_t is the time since arrival of the customer in service, the

phase J_t the same as the phase in which the server is operating. This gives a stationary density $\pi(x)$ (a row vector) at level x of the form $\pi(x) = \pi(0)e^{Tx}$ for some matrix T.

Asmussen & O'Cinneide (1998) pointed out that the many–server queue GI/PH/c (with homogeneous or heterogeneous servers) is included in Sengupta's set–up. The process (J_t, L_t) is now obtained by piecing all–busy periods (i.e., periods where all c servers are working) together, and the level L_t is the time since arrival of the last customer to enter service, the phase J_t the combination of the phases in which the servers are operating. The argument immediately yields an algorithm for computing the matrix T, but the computation of $\pi(0)$ (trivial for $c = 1$) was missing. This requires a careful analysis of non–all–busy periods, carried out by Asmussen & Møller (2000/01), who also treated arrivals governed by a finite Markov process. In connection with Asmussen & Møller (2000/01), MatLab programs have been developed, which are available as shareware.

References

S. Asmussen & C.A. O'Cinneide (1998) Representations for matrix–geometric and matrix–exponential steady–state distributions with applications to many–server queues. *Stochastic Models* **14**, 369–387.

S. Asmussen & J.R. Møller (2000/01) Calculation of the steady–state waiting time distribution in many–server queues. *Queueing Systems* (to appear).

G. Latouche & V. Ramaswami (1999) *Introduction to Matrix Analytic Methods in Stochastic Models*. SIAM.

M.F. Neuts (1981) *Matrix–Geometric Solutions in Stochastic Models*. Johns Hopkins University Press.

M.F. Neuts (1989) *Structured Matrices of the M/G/1 Type and Their Applications*. Marcel Dekker.

B. Sengupta (1989) Markov processes whose steady–state distribution is matrix–exponential with an application to GI/PH/1 queues. *Adv. Appl. Probab.* **14**, 369–387.

Adrian Baddeley

University of Western Australia

Conditional simulation

Abstract: Conditional simulation is a technique for extrapolating spatial data beyond the restricted region or set of sites where the data were observed. For example we may have measured the concentration of a pollutant at a few sample locations, and want to extrapolate this to a continuous spatial map of concentration; or we may have recorded the positions of geological faults inside a mine, and want to predict the likely locations of faults outside the mined region.

In conditional simulation we generate *randomly* the values of the spatial variable/pattern of interest outside the region where data are available, following an assumed stochastic model, in a manner that is faithful to the observed data. In other words, adopting an appropriate stochastic model, we draw a sample from the conditional distribution of the spatial process given the observed data.

In these talks I will explain the concept of conditional simulation and describe some of its pitfalls. While some instances of conditional simulation are simple and elegant, others are complicated by issues such as sampling bias, combinatorial complexity, reducibility, semicontinuity, and unfolding effects. The first talk will introduce the basic ideas, in the case of continuous spatial random processes (random fields). The second talk will address the case of random patterns of geometrical objects (random sets).

The talks include current joint work with Nick Fisher (CSIRO), Marie-Colette van Lieshout (CWI), Henry Cheng (Fisheries WA) and others.

Laird Breyer

Aalborg University

Automatic ways of coupling Markov chains

Abstract: Coupling constructions are central to some of the modern developments in Stochastic Processes. When considering Markov processes, one problem is to define two or more dependent chains with the same marginal distributions in such a way that their sample paths meet in a finite time.

This allows us for example to bound the total variation distance of a Markov chain to its stationary distribution, in an entirely probabilistic way. This has led recently to computable bounds. For another example, in Perfect Simulation couplings are used to produce exact samples from distributions that are otherwise hard to simulate from.

In this talk, I shall describe a family of methods for coupling chains which are applicable whenever the transition probabilities are known, and more importantly do not require the calculation of minorization conditions, which are needed for the well known and widely successful "splitting technique". It is in this sense that the methods presented are "automatic" — they do not require analytic estimates. Various examples from Markov Chain Monte Carlo will be given.

Ole F. Christensen (with J. Møller and R. P. Waagepetersen)

Aalborg University

Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo

ABSTRACT: Conventional geostatistics solves the problem of estimation and prediction for continuous observations (Cressie, 1993). But in many practical applications the available data are binary or counts for which normality cannot be obtained by means of transformation.

For modelling of non-Gaussian data generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993) are extensions of generalized linear models (GLMs) that allows additional components of variability due to unobservable effects which are modeled by the inclusion of random effects in the linear predictor of the generalized linear model. For a spatial GLMM the underlying random effects are modeled by a Gaussian process on $\mathbb{R}^2$. Given the underlying and unobserved Gaussian process the observations at the measured locations are conditionally independent and follows a GLM. Bayesian analysis of such models is studied in Diggle *et al.* (1998), where uniform proper priors are used for the model parameters. We investigate the question of posterior propriety when flat improper priors are used.

Conditional simulation of the unobserved Gaussian field given the observed data is relevant for prediction of a functional of the Gaussian field, and it requires an efficient Markov chain Monte Carlo method. The Hastings-Langevin algorithm (Besag, 1994; Roberts and Tweedie, 1996) turns out to be very useful when the model parameters are considered as fixed. For this algorithm we study the desirable property of geometric ergodicity, which ensures the validity of central limit theorems for a Monte Carlo estimate (see Corollary 2.1 in Roberts and Rosenthal, 1997). The Hastings-Langevin algorithm can also easily be extended with updates of the parameters for a full Bayesian analysis.

We focus on the so-called Poisson-log normal model, where as an example the algorithm is applied to a data set of counts of weed plants on a field. In this example we demonstrate that our Langevin-type algorithm has a much better performance than a Metropolis random walk algorithm.

References

Besag, J. E. (1994). Discussion on the paper by Grenander and Miller. *J. R. Statist. Soc.* B **56**, 591–592.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.* **88**, 9–25.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Revised edition. Wiley, New York.

Diggle, P., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Appl. Statist.* **47**, 299–350.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2**, 13–25.

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2**, 341–363.

Ian Dryden

University of Nottingham

Stochastic deformation

ABSTRACT: Informally, a deformation can be regarded as a function which geometrically transforms an object. Deformations are commonly studied in statistical shape analysis, where the geometrical properties of objects under certain invariances (such as location, rotation and scale) are considered. Deformations can often be decomposed into global and local components, and by global differences we mean large scale trends, such as an overall affine or similarity transformation. Local differences are on a smaller scale, for example highlighting changes in a small part of an object. When comparing the shapes of two objects it is often common to register the two objects together using a global deformation, so that the objects are geometrically 'close' according to some criterion. The local deformations are then used to assess how similar or not the objects are in terms of shape.

For simplicity we shall concentrate on the situation where a set of k corresponding points is available on each object, which is in m real dimensions (usually $m = 2$ or $m = 3$). Such points are often called landmarks and the points correspond between objects in a geometrical or functionally meaningful way. Consider two k landmark configuration matrices in $\mathbb{R}^m$, $T = (t_1, \ldots, t_k)^{\mathrm{T}}$ and $Y = (y_1, \ldots, y_k)^{\mathrm{T}}$ both $k \times m$ matrices, and we wish to deform T into Y, where $t_j, y_j \in \mathbb{R}^m$. A deformation is a mapping from $t \in \mathbb{R}^m$ to $y \in \mathbb{R}^m$ defined by the transformation

$$y = \Phi(t) = (\Phi_1(t), \Phi_2(t), \ldots, \Phi_m(t))^{\mathrm{T}}.$$

Here T is the source and Y is the target. The multivariate function $\Phi(t)$ should have certain desirable properties. In particular, we often want as many of the following properties to hold as possible for the deformation: continuous, smooth, bijective, not prone to gross distortions (e.g. not folding which will be guaranteed if the mapping is bijective), equivariant under certain global transformations of the objects, and an interpolant. If the interpolation property is not satisfied then we call the deformation a smoother. Note that, as we describe it here, the deformation is from the whole space $\mathbb{R}^m$ to $\mathbb{R}^m$, rather than just from a set of landmarks to another or an outline to another. However, there are other notions of deformations and we shall consider an alternative in the second talk, where the definition applies only to the landmarks.

D'Arcy Thompson (1917, *On Growth and Form*, Cambridge) considered deformations from one species to another in order to explain size and shape differences. A regular square grid pattern was drawn on one object and the grid was deformed to lie on the second object, with corresponding biological parts located in the corresponding grid blocks. In Figure 1 we see a famous example, and these grids are known as Cartesian transformation grids. The transformation grids enable a biologist to describe the shape change between the two species, albeit in a rather subjective way. D'Arcy Thompson's (1917)

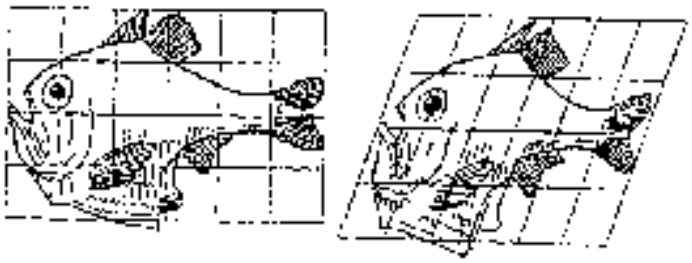

Figure 1: Cartesian transformation grids from one species of fish to another (from D'Arcy Thompson, 1917). The transformation is an affine deformation.

figures were drawn by hand and there have been many attempts since 1917 to recreate these figures more objectively, and we shall briefly discuss these. Perhaps the simplest possible size and shape change between two objects is that of an affine transformation, as seen in Figure 1. In this case the square grid placed on the first fish is deformed uniformly and affinely into a parallelogram grid on the second fish.

Stochastic deformations can be used for a very wide variety of applications. For example, obtaining mean shapes and exploring shape variability; image matching and warping; and object recognition, where templates are deformed to match observed images. Some common methods will be reviewed in the first talk. We shall mainly concentrate on the important $m = 2$ dimensional case, with deformations given by the bivariate function

$$y = \Phi(t) = (\Phi_1(t), \Phi_2(t))^{\mathrm{T}}.$$

Bookstein (1989, IEEE PAMI) has developed a highly successful approach for deformations using a pair of thin-plate splines for the functions $\Phi_1(t)$ and $\Phi_2(t)$. The thin-plate spline is a sensible choice since it minimizes the bending required to take the first form into the second, and it does not suffer from problems with very large bending towards the periphery. The thin-plate spline is related to kriging in spatial statistics, and has the advantage of having a simple analytic solution.

In the second talk we will discuss a different approach to modelling stochastic deformations of outlines, introduced by Grenander and colleagues. The deformation is applied to the edges between landmarks, rather than the whole domain, and is particularly suitable where the objects under study are outlines deformed from an underlying regular object. In some datasets there are no discernible features and yet we are still interested in answering questions about the shape distributions of the outlines of the objects under study.

We shall concentrate on some applications in particle science, describing joint work with John Kent and Catherine Anderson at Leeds. The paper is on the Web at:

`http://www.amsta.leeds.ac.uk/~iand/papers/blobs.ps.gz`

Grenander and Miller (1994, JRSS B) describe a model for representing amorphous 2-

dimensional objects with no obvious landmarks. Each object is represented by a set of vertices/landmarks around its perimeter, and is described by deforming a regular polygon using edge transformations. A multivariate normal distribution with a block circulant covariance matrix is used to model these edge transformations. The talk will describe the statistical properties of this multivariate model and the eigenstructure of the covariance matrix. Various special cases of the model are considered, including articulated models and conditional Markov random field models. We consider maximum likelihood based inference and the model is applied to some datasets to explore shape variability.

Günter Döge

Freiberg University of Mining and Technology

Grand Canonical Simulations of Hard-Disk Systems by Simulated Tempering

Abstract: For the simulation of hard core Gibbs point processes in the two-dimensional space simulated tempering is shown to be an efficient alternative to commonly used Markov chain Monte Carlo algorithms, especially for grand canonical ensembles, i.e., with a variable number of disks The behaviour of the area fraction and various spatial characteristics of the hard core process is studied using simulated samples.

Paul Glasserman (with P. Heidelberger and P. Shahabuddin)

Columbia University

Variance Reduction Techniques for Simulating Value-at-Risk

ABSTRACT: An important concept for quantifying and managing portfolio risk is value-at-risk (VAR). VAR is defined as a quantile of the loss in portfolio value during a holding period of specified duration. If the value of the portfolio at time t is $V(t)$, the holding period is Δt, and the value of the portfolio at time $t + \Delta t$ is $V(t + \Delta t)$, then the loss in portfolio value during the holding period is $L = V(t) - V(t + \Delta t)$. For a given probability p, the VAR, x_p, is defined to be the $(1 - p)$'th quantile of the loss distribution:

$$P\{L > x_p\} = p. \tag{1}$$

Typically, the interval Δt is one day or two weeks and p is close to zero, often $p \approx 0.01$. Monte Carlo simulation is frequently used to estimate the VAR. In such a simulation, changes in the portfolio's "risk factors" (e.g., interest rates, currency exchange rates, stock prices, etc.) during the holding period are generated and the portfolio is re-evaluated using these new values for the risk factors. This is repeated many times so that the loss distribution may be estimated.

The computational cost required to obtain accurate Monte Carlo VAR estimates is often enormous. This is due to two factors. First, the portfolio may consist of a very large number of financial instruments. Furthermore, computing the value of an individual instrument may itself require substantial computational effort. Thus each portfolio evaluation may be costly. Second, a large number of runs (portfolio evaluations) are required in order to obtain accurate estimates of the loss distribution in the region of interest. We focus on this second issue: the development of variance reduction techniques designed to dramatically reduce the number of runs required to achieve accurate estimates of low probabilities. The technique described in this paper combines two general purpose variance reduction techniques: importance sampling and stratified sampling. We provide a rigorous analysis of this approach, and perform extensive experiments on it.

Our approach is to approximate the portfolio loss by a quadratic function of the underlying risk factors and to use this approximation to design variance reduction techniques. Quadratic approximations are widely used without simulation; indeed the second order Taylor series approximation is commonly called the "delta-gamma approximation". While our approach could be combined with other quadratic approximations, many of the first and second derivatives needed for the delta-gamma approximation are routinely computed for other purposes quite apart from the calculation of VAR. A premise of this paper is that these derivatives are thus readily available as inputs to be used in a VAR simulation and do not represent an additional computational burden.

When the change in risk factors has a multivariate normal distribution, as is commonly

assumed (and as we will assume), then the distribution of the delta-gamma approximation can be computed numerically. While this approximation is not always accurate enough to provide precise VAR estimates, we describe how it may be used to guide selection of an importance sampling (IS) change of measure for sampling the changes in risk factors. IS is a particularly appropriate technique for "rare event" simulations, which corresponds to the VAR problem with a small value of p. As the distribution of the quadratic approximation can be computed numerically, it can also be used as either a control variable or for stratified sampling. Numerical results show that while the effectiveness of the control variable decreases as p decreases, the effectiveness of a combination of IS and stratified sampling increases as p decreases. We provide a theoretical analysis showing asymptotic optimality of the method as either the loss threshold or the number or risk factors increases.

Paul Glasserman (with M. Broadie)

Columbia University

Pricing American Options by Simulation

ABSTRACT: Computational methods for pricing derivative securities can be broadly divided into deterministic methods and simulation-based methods. The first type generally involves discretizing time and discretizing the possible levels of the underlying asset prices; the discrete approximation is then solved exactly. Well-known examples of this approach include binomial and trinomial lattices, and finite difference methods. These methods are widely used, particularly in valuing relatively simple derivative securities in relatively simple models.

Deterministic methods can be very fast and effective if the dimension of the state vector representing the underlying model is 1, 2, or perhaps 3. But the time and space requirements of these methods typically grow exponentially in the dimension, rendering these methods inapplicable to high-dimensional problems.

Simulation methods are based on stochastic sampling of paths of the underlying state vector. Their space requirements generally grow linearly in the dimension of the state vector. They typically converge in proportion to the square root of the number of paths generated, a convergence rate independent of the dimension of the problem. This makes simulation-based methods attractive for valuing path-dependent and multi-asset derivatives.

A complication arises, however, with simulation techniques in pricing option contracts with American-style features—i.e., contracts in which the holder can choose the time of exercise. In this case, an optimal exercise boundary has to be determined through some type of dynamic programming procedure. The difficulty arises in combining the forward evolution of simulation with backward induction of dynamic programming. Recently, several methods have been proposed to address this issue.

In this paper we discuss two methods. The first generates random trees of prices and applies a dynamic programming recursion to each tree. High- and low-biased estimators are combined to obtain a conservative but valid confidence interval for the true price. The second method generates a *stochastic mesh.* This method simulates multiple paths in parallel and uses information from all paths to estimate the continuation value (the value of holding an option rather than exercising) at each node along each path. The continuation value at each node is estimated as a discounted weighted average of the option values at the next time step across all paths. The weights are computed from the transition density of the underlying process. The methods are analyzed theoretically and tested on realistic examples.

Niels Væver Hartvig

University of Aarhus

A stochastic geometry model for fMRI data

Abstract: Functional magnetic resonance imaging (fMRI) is a medical imaging technique where fast MR scanners are used to measure changes in blood oxygenation in the brain. It is believed that these oxygenation changes correlate with neural activity in the surrounding tissue, and hence the technique can be used to measure activation in the brain as caused by external stimuli. The data acquired in these experiments consists of a sequence of scans, typically around 100, and the aim of the statistical analysis of the data is to identify regions in the images, where the intensity changes according to the stimulus rhythm.

In a typical analysis of these data the problem is marginalized to a one dimensional time-series problem for each voxel in the scan, see for instance Worsley and Friston (1995) and Lange and Zeger (1997). The spatial structure of the data is included in a second step, when the image of activation estimates is convolved with a smoothing kernel to obtain a non-parametric estimate of the activation. In this approach the focus is on assessing significance of peaks and clusters in the smoothed image, effectively by testing thousands of hypotheses simultaneously.

In the talk I will present an approach to analyzing fMRI data where the focus is shifted towards estimating the location and size of activated areas, rather than testing multiple voxel-wise hypotheses. This is achieved by formulating a more structured spatial model in the spirit of high-level image analysis, see e.g. Baddeley and van Lieshout (1993). More specificly the spatial activation surface is modelled by a collection of Gaussian functions, which to some extent can be thought of as individual centres in the brain. The model is formulated in a Bayesian setting where the centres *a priori* are distributed as a marked point process; here the points are the locations of the centres and the marks describe the shape and height of the centres. The inference in the model is based on simulation techniques, by which we can estimate the posterior mean of functions of interest, such as the mean activation pattern. The model can be formulated either in a spatial setting only or may be embedded in a truly spatio-temporal analysis.

One of the advantages compared to more simple models, is that the uncertainty of the estimated activated pattern can be readily assessed from the posterior distribution. This is largely ignored in common analyses. Secondly we are able to relax the common assumption of stationarity of the temporal activation pattern, which indeed reveals significant features of non-stationarity in the data.

References

Baddeley, A.J. and van Lieshout, M.N.M. (1993) Stochastic geometry models in high-level vision. In K.V. Mardia and G.K. Kanji (eds.), *Statistics and Images*, vol. 1, chap. 11, pp. 231–256, Appl. Statist.

Geyer, C.J. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.*, **21**, 359–373.

Hartvig, N. (1999) A stochastic geometry model for fMRI data. Tech. Rep. 410, University of Aarhus. *Submitted for publication.*

Lange, N. and Zeger, S.L. (1997) Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion). *Appl. Statist.*, **46**, 1–29.

Worsley, K.J. and Friston, K.J. (1995) Analysis of fMRI time-series revisited — again. *Neuroimage*, **2**, 173–181.

Jotun Hein (with B. Knudsen)

University of Aarhus

RNA secondary structure prediction using stochastic context-free grammars and evolutionary history

Abstract: Many computerised methods for RNA secondary structure prediction have been developed. None of these methods, however, employ an evolutionary model, thus they leave out relevant information from the structure determination. This talk introduces a method which incorporates evolutionary history into RNA secondary structure prediction. Furthermore, many methods for structure prediction, from more than one sequence, do not use prior information about structures. The method reported here is based on stochastic context-free grammars (SCFGs) to give a prior distribution of structures.

Anders Krogh

Technical University of Denmark

Applications of hidden Markov models in molecular biology

ABSTRACT: Hidden Markov models (HMMs) are well suited for biological sequences. Applications of HMMs to membrane protein structure prediction and gene finding is presented. Both problems have a grammatical structure which can be described by a regular grammar (to a good approximation). Because of experimental uncertainties, limited datasets, and lack of biological knowledge, it is essential to use a statistical model. Therefore, a stochastic regular grammar, which is the same as a hidden Markov model, is the natural model choice.

In most membrane proteins the transmembrane region is made up of a bundle of alpha helices. Because of the nature of the lipid bilayer, the amino acids in these helices are predominantly hydrophobic, so in the linear sequence of amino acids, a transmembrane helix typically shows up as a stretch of hydrophobic amino acids with a typical length of about 20-25. Additionally, it has been found that there is an abundance of positively charged amino acids on the inside (cytoplasmic side) of the membrane. These are the two most important features used to predict transmembrane helices and their orientation. The advantage of using an HMM is that these and other signals can be combined in one model, and the model can be constrained to only allow sensible structures where for instance inside-helix-inside cannot be predicted because it violates the 'grammar' of membrane proteins.

In many eukaryotic genes the parts coding for protein (exons) are interrupted by long non-coding introns. It is quite difficult to locate the short coding regions. Traditionally coding regions were located by their codon statistics - codons occur with frequencies different from the frequencies in non-coding regions. However, there is not enough information in short exons to rely on a codon statistic. There is a signal of variable strength associated with the junctions between introns and exons (splice sites). By combining these two types of signals (and a few others) it is possible to obtain reasonable gene identification, although it is still not quite satisfactory for higher eukaryotes such as humans. Again the strength of the HMM is that it can deal with all the signals as well as the grammatical constraints in genes.

Ole G. Mouritsen

Technical University of Denmark

The third science — the computer experiment

Abstract: Since the days of Galileo there has been an indissoluble tie within the exact natural sciences between on the one side the *experimental* method of studying the physical universe and on the other side the *theoretical* approach to rationalize and predict observations. Although scholars at different periods have often stressed that exact natural sciences such as chemistry and physics are primarily experimental sciences, experience has shown that experiments without theory and theory without experiments rarely lead to the deepest insights.

With the invention of powerful and fast computers, the tie between experiment and theory has assumed a novel dimension in the form of the *theoretical experiment* or the *computer experiment*. This development has proved so successful that it has been referred to as the *third natural science*. In the computer experiment numerical (mathematical) experiments can be carried out on model systems under fully controlled circumstances by 'teaching' the computer the laws of nature. This allows for new discoveries within the framework of the model under study. For example, numerical experiments can be carried out under extreme or idealized conditions that may not be obtained in conventional laboratory experiments.

Specifically, computer experiments can be used to study the collective behavior of large assemblies of particles (e.g. atoms or molecules) using the rules of statistical mechanics and assuming the physical interactions between the particles. Thereby it is possible to provide a connection between the microscopic description of physical systems and the macroscopic world as we observe it. This allows for investigation of emergent properties of matter and complex pattern-formation processes.

The key mathematical problem in statistical mechanics involves multi-dimensional integration of functions with exponential weight factors (Boltzmann statistics). A powerful approach to simulate the statistical mechanics of many-particle systems is Monte Carlo importance sampling techniques which involves stochastic elements. By these techniques it is possible to extract accurate numerical information in the large-system, long-time limit by appropriate histogram sampling and finite-size scaling techniques. Moreover, it is possible to explore pseudo-non-ergodic situations, where the phase space effectively decomposes into separate subspaces, using non-Boltzmann sampling techniques. This is of particular importance for the investigation of transitions between different states of matter which involve symmetry-breaking.

As an illustration of the use of modern computer-simulation techniques that exploit Monte Carlo methods to solve the statistical mechanical problems of particle systems in and away from thermodynamic equilibrium, a broad selection of examples will be discussed. The examples, which are drawn from the fields of physics, chemistry, and

biology, include magnetic systems, high-temperature superconductors, liquid crystals, liquid mixtures, cholesterol, proteins, cell membranes, and enzymes.

Some references

Efficient Monte Carlo sampling by direct flattening of free energy barriers (G. Besold and O. G. Mouritsen) *Comp. Mat. Sci.* **15**, 311-340 (1999).

Computer simulation of lyotropic liquid crystals as models of biological membranes (O. G. Mouritsen). In *Advances in the Computer Simulations of Liquid Crystals* (P. Pasini and C. Zannoni, eds.) Kluwer Academic Publ. Dordrecht (1999) pp. 139-187.

An off-lattice model for the phase behavior of lipid-cholesterol bilayers (M. Nielsen, L. Miao, J. H. Ipsen, M. J. Zuckermann, and O. G. Mouritsen) Phys. Rev. E **59**, 5790-5803 (1999).

Tomáš Mrkvička

Charles University, Prague

Estimation variances for Poisson processes of compact sets

Abstract: Estimators of intensity functions for various Poisson processes of compact sets are studied. Let Φ be a stationary Poisson process of compact sets in $\mathbb{R}^d$ with intensity $\alpha > 0$ and known primary grain distribution and let $\mathcal{W}$ be a bounded closed family of nonempty compact sets in $\mathbb{R}^d$ and $\mathcal{E}_{\mathcal{W}}$ the set of all estimators (measurable functions of Φ) which depend only on the restriction $\Phi|_{\mathcal{W}}$. It is shown that $\Phi(\mathcal{W})$ (the number of compact sets from $\mathcal{W}$ in the process) is a complete and sufficient statistic for the intensity α. The abstract Rao-Blackwell theorem [1] implies then that $\mathbb{E}\left[e(\Phi)|\Phi(\mathcal{W})\right]$ is the uniformly best unbiased estimator of a parametric function $\tau(\alpha)$ of α among all estimators from $\mathcal{E}_{\mathcal{W}}$ whenever e is an unbiased estimator of $\tau(\alpha)$.

The general theory is then applied to the stationary Poisson segment process, where the length intensity is estimated. The uniformly best unbiased estimator among the estimators using all segments visible in an observed window is found. This estimator is the best one, but it is hardly applicable for its involved form. When considering only the segments which have their reference points within the observed window then the uniformly best unbiased estimator among such estimators is $\Phi(\mathcal{W})$ multiplied by the mean segment length and divided by the window volume. This estimator does not use all information but it is very easily applicable.

Finally, the variance of the last estimator is compared in some particular situations with that of the natural estimator based on summing up the lengths of the visible parts of the segments. It is shown that the estimator based on the number of segments has in many cases lower variance (cf. [2] where a similar problem for Poisson flat processes has been considered).

References

[1] A. J. Baddeley, L. M. Cruz-Orive, *The Rao-Blackwell theorem in stereology and some counterexamples*, Adv. Appl. Prob. **27** (1995), 2–19.

[2] K. Schladitz, *Estimation of the Intensity of Stationary Flat Processes, Dissertation*, Friedrich-Schilrer-Universität, Jena, 1996.

Søren Feodor Nielsen

University of Copenhagen

Simulated EM algorithms: A comparison

ABSTRACT: The EM algorithm is a well-known iterative method for finding the MLE in missing or incomplete data problems. Each iteration updates a current estimate (θ_k) of the unknown parameter by going through two simple steps:

E-step: Calculate the conditional **expectation** of the complete data log likelihood given the observed data using the current estimate θ_k as the "true" parameter. This yields a function $\theta \to Q(\theta|\theta_k)$.

M-step: **Maximize** this function to find the next estimate of the unknown parameter; $\theta_{k+1} = \arg\max_\theta Q(\theta|\theta_k)$.

This leads to a sequence, $(\theta_k)_k$, of estimators converging (under suitable assumptions) to the observed data MLE.

However, in some cases the conditional expectation required in the E-step cannot be calculated and must –for instance– be estimated instead. Thus, the E-step is replaced by an

SimE-step: **Simulate** m values of the complete data given the observed data using the current estimate $\tilde{\theta}_k$ as the "true" value of the unknown parameter and estimate $Q(\theta|\tilde{\theta}_k)$ by the average of the m complete data log-likelihoods.

In the M-step this estimator of $Q(\theta|\tilde{\theta}_k)$ is maximized as a function of θ. This leads to a random sequence, $(\tilde{\theta}_k)_k$, of estimators. We call this algorithm a *simulated EM algorithm*. Two different versions of the simulated EM algorithms have been suggested:

- In each iteration new random numbers are generated, i.e. the simulations in the kth iteration only depend on the previous simulations through $\tilde{\theta}_{k-1}$. The sequence $(\tilde{\theta}_k)_k$ is a Markov chain conditional on the observed data.
- Alternatively the random numbers can be "re-used", i.e. in the kth iteration the simulations are generated using $\tilde{\theta}_k$ but the "same randomness" (the same random numbers or uniforms) as in the previous iteration. The sequence $(\tilde{\theta}_k)_k$ is deterministic conditioned on the observed data and the simulations, i.e. the random numbers generated, in the first iteration. Hence, this version attempts to find the maximum of a random function $\theta \to \widetilde{Q}(\theta|\theta)$ by the method of successive substitutions.

In this talk these two versions are compared and their relative merits are discussed.

References

S. F. Nielsen (2000a) *On simulated EM algorithms* Journal of Econometrics, to appear

S. F. Nielsen (2000b) *The stochastic EM algorithm: Estimation and asymptotic results* Bernoulli, to appear

Estimating a non-stationary spatial structure using simulated annealing

Serge IOVLEFF*and Olivier PERRIN†

April 28, 2000

Abstract

During the past decade, a useful model for non-stationary random fields has been developed. This consists of reducing the random field of interest to isotropy *via* a bijective bi-continuous deformation of the index space. Then the problem consists of estimating this space deformation. We propose to estimate this space deformation using a constrained continuous version of the simulated annealing for a Metropolis dynamic. This method provides a non-parametric estimation of the deformation which has the required property to be bijective; so far, the previous non-parametric methods do not guarantee this property. We illustrate our work with two examples, one concerning a precipitation data set. We also give one idea of how spatial prediction should proceed in the new coordinate space.

Key Words: bijective space deformation; constrained minimisation; correlation function; Delaunay triangulation.

1 Introduction

Assumption of isotropy is clearly violated for many, if not most, spatial environmental phenomena. Factors such as topography, local pollutant emissions, and meteorological influences may cause such assumptions to be violated. This has led to research into modelling a spatially non-stationary second order structure, as reviewed in Guttorp and Sampson (1994). To deal with this non-stationarity, Sampson and Guttorp (1992) have developed a model that consists of reducing the correlation function $r(\mathbf{x}, \mathbf{x}')$ of the spatial phenomenon of interest, modelled by a random field $Z = \{Z(\mathbf{x}), \mathbf{x} \in G \subseteq \mathbb{R}^2\}$, to an isotropic one as follows:

$$r(\mathbf{x}, \mathbf{x}') = \rho_\beta(\|\Phi(\mathbf{x}') - \Phi(\mathbf{x})\|), \tag{1}$$

*Laboratoires SABRES, IUP Vannes-Tohannic, rue Yves Mainguy, 56000 Vannes, France

†Mathematical Statistics, Chalmers University of Technology and Göteborg University, S-412 96 Göteborg, Sweden

where $\|.\|$ denotes the Euclidean norm in $\mathbb{R}^2$, Φ represents a bijective deformation of the geographic coordinate system and ρ_β is an isotropic correlation function with parameter $\beta \in \mathbb{R}^q$, $q \geq 1$. In the sequel, we refer to the geographical coordinate system as the G-space $\subseteq \mathbb{R}^2$, where G stands for geographical, and to the deformed coordinate representation as the D-space $\subseteq \mathbb{R}^2$, where D stands for deformed. One illustration of model (1) is given in Figure 1 (this example is detailed in Paragraph 3.1): (i) represents the positions $\mathbf{x}_i$, $i = 1, 2, \ldots, n$, of n sites in the G-space; (ii) represents their deformations $\Phi(\mathbf{x}_i)$ in the D-space; (iii) represents the inter-site distances in the G-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j) = \rho_\beta(\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_i)\|)$, $1 \leq i, j \leq n$, where $\rho_\beta(u) = \exp(-\beta\|u\|)$ with $\beta = 1$; (iv) represents the inter-site distances $\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_i)\|$ in the D-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j)$.

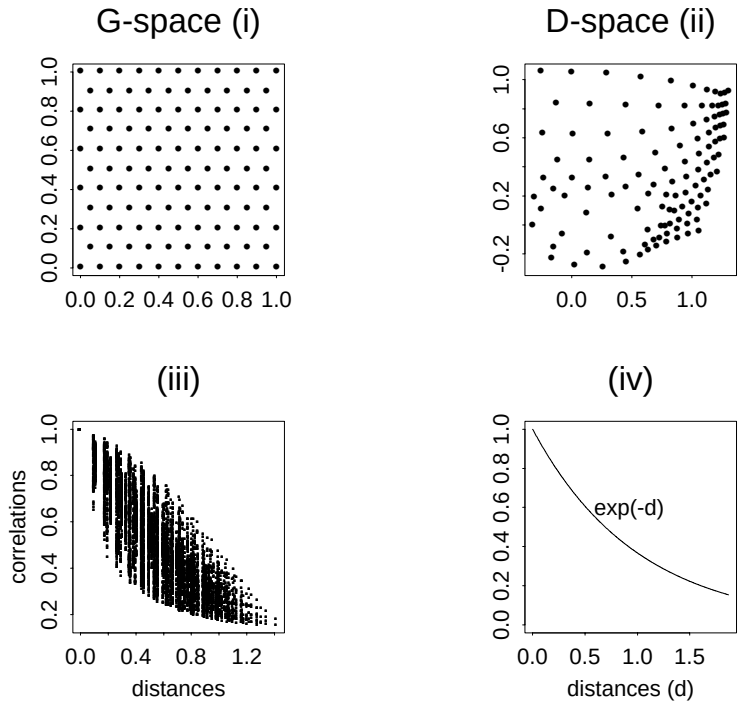

Figure 1: Example of a non-stationary random field using a space deformation.

Unlike the classical geostatistical models, non-stationarity through second order moments is thus taken into account and model (1) gives the opportunity to enlarge the class of models for studying spatial environmental random fields.

When ρ_β is strictly decreasing, Perrin and Meiring (1998) prove the uniqueness of both the deformation Φ and ρ_β up to a homothetic Euclidean motion for Φ and up to a scaling for ρ_β. Perrin and Senoussi (1998) give the general form of the deformation that reduces a non-stationary random field in the way (1), under smoothness assumptions.

Our concern in this paper is the estimation of both the space deformation Φ and the parameter β. This estimation is based on T repetitions (independent and identically distributed observations) of Z at each of n distinct monitoring sites $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in G, which may be irregularly located; G represents the convex hull of these sites. We denote

these records using $Z_t(\mathbf{x}_i)$, $t = 1, 2, \ldots, T$, $i = 1, 2, \ldots, n$.

So far, the deformation Φ has mostly been estimated with the help of a non-parametric approach. This non-parametric estimation of Φ has already been extensively developed (Monestiez *et al.* (1993), Meiring (1995), Meiring *et al.* (1997, 1998)), and applied, for instance, in analyses of acid precipitation (Guttorp *et al.* (1992)), solar radiation (Sampson and Guttorp (1992)) and tropospheric ozone (Sampson *et al.* (1994)). The latter development of the non-parametric approach for estimating Φ consists of modelling Φ using a pair of thin-plate splines and minimising a penalised weighted least squares criterion to estimate the D-space coordinates $\Phi(\mathbf{x}_i)$ of the monitoring sites $\mathbf{x}_i$, $i = 1, 2, \ldots, n$. This treatment is carried out with a Marquardt-type algorithm. Several drawbacks are associated with this method: *(i)* bijection condition is not ensured; *(ii)* the fitting of the model becomes a challenging numerical problem with dimensionality roughly proportional to the number of fixed monitoring sites (Meiring *et al.* (1998)); *(iii)* the objective function to be minimised is non-convex and has a lot of local minima; *(iv)* there is a considerable dependence on the starting values in the minimisation procedure.

Perrin and Monestiez (1998) propose a parametric approach. They model Φ using a composition of a "small" number of bijective elementary radial basis deformations. Then they use a least squares criterion to estimate the parameters; this criterion is minimised in a classical way (Marquardt-type algorithm). Let us specify two disadvantages of this method: *(i)* so far a rational choice of the relevant number of elementary deformations has not been implemented yet; *(ii)* there is a considerable dependence on the starting values in the minimisation procedure.

To avoid all these disadvantages, we propose here to estimate, in a first step, the D-space coordinates $\Phi(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$, by minimising an objective function with a stochastic algorithm, the continuous version of the simulated annealing for a Metropolis dynamic. To ensure a bijective correspondence between the n points in the G-space and their estimated deformations in the D-space, we impose some non-folding constraints in the algorithm. In a second step, to estimate the deformation in the whole G-space, we use a piecewise affine interpolation of the points $\mathbf{x}_i$ in the G-space and the estimations of their deformations $\Phi(\mathbf{x}_i)$ in the D-space, $i = 1, 2, \ldots, n$.

The paper is structured as follows. In Section 2, we address the estimation problem of the deformation together with the parameter β. In Section 3, we illustrate our non-parametric approach with two examples: the first one is purely illustrative and is only concerned with the estimation of the deformation when the isotropic correlation is known; the second one shows how our method can apply to precipitation data from 20 sites in the Languedoc-Roussillon region of France. In Section 4 we give one idea of how spatial prediction should proceed in the new coordinate space and propose a cross validation study to demonstrate the possible improvement in predictions due to the deformation. Further, we apply this cross validation study to our precipitation data set through a comparison of four isotropic correlation models. Finally, Section 5 outlines one extension related to this work.

2 Estimation of the model

2.1 Definition of the objective function

The repetitions of Z allow us to define the sample correlation estimates $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$ for each couple of sites $(\mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i < j \leq n$:

$$\hat{r}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{t=1}^{T} (Z_t(\mathbf{x}_i) - \bar{Z}(\mathbf{x}_i))(Z_t(\mathbf{x}_j) - \bar{Z}(\mathbf{x}_j))/\hat{\sigma}(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_j), \tag{2}$$

where $\bar{Z}(\mathbf{x}_i)$ are the empirical means and $\hat{\sigma}(\mathbf{x}_i)$ are the empirical standard deviations, $i = 1, \ldots, n$. We denote using $\mathbf{y}_i = \Phi(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$, the deformations of the sites in the D-space and we define $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_n$ and $\hat{\beta}$ as the estimations of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ and β which minimise the following objective function:

$$U(\mathbf{y}^\star, \beta^\star) = \sum_{i<j} [\hat{r}(\mathbf{x}_i, \mathbf{x}_j) - \rho_{\beta^\star}(\|\mathbf{y}_j^\star - \mathbf{y}_i^\star\|)]^2, \tag{3}$$

with respect to the parameters $\mathbf{y}^\star = (\mathbf{y}_1^\star, \mathbf{y}_2^\star, \ldots, \mathbf{y}_n^\star)$ and $\beta^\star$, and subject to some non-folding constraints described in Paragraph 2.2.

2.2 Minimising the objective function

The objective function (3) is minimised with respect to $\mathbf{y}^\star$ using a continuous state version of the simulated annealing subject to some non-folding constraints. Simulated annealing is a probabilistic method for finding the global minimum of an objective function that may possess several local minima (Geman and Geman (1984), Kirkpatrick *et al.* (1983), Hajek (1988), Aarts and Korst (1989), Geman (1990) and Azencott (1992) for seminal references). This is motivated by its following advantages:

- it explores the whole objective function's surface and tries to optimise the function while moving both uphill and downhill. Thus, it is largely independent of the starting values, often a critical input in conventional algorithms;
- it can escape from local minima and go on to find the global minimum by the uphill and downhill moves;
- it makes less stringent regularity assumptions regarding the function than do conventional algorithms (it need not even be continuous);
- it is well suited for minimising strongly non-convex functions of several variables ($2n$ variables in our problem) having plenty of local minima;
- it can take intricate constraints into account.

2.2.1 Description of the minimisation algorithm

We describe hereafter the different steps to minimise (3) both with respect to $\mathbf{y}^\star$ and β:

- starting step: set $\mathbf{y}(0) = \{\mathbf{y}_i(0) = \mathbf{x}_i, i = 1, 2, \dots, n\}$. Then to give an initial value to $\beta^\star$, minimise the objective function (3) with respect to $\beta^\star$ ($\mathbf{y}^\star = \mathbf{y}(0)$ is hold fixed) with a Marquardt-type algorithm. Let $\beta(0)$ be the estimation of β at step 0. Finally, take a sequence of "temperatures" $(c_0, c_1, \dots, c_k, \dots)$ decreasing to 0 by step of length n:
$$c_k = \theta^{\lfloor k/n \rfloor} c_0, \quad \theta \in]0, 1[, \ k \in \mathbb{N};$$
- step 0: start from the configuration $\mathbf{y}(0)$ of the sites. The change proposition is: fix $\beta^\star = \beta(0)$ and choose one site (candidate point) j uniformly among the n sites and move it locally and uniformly at a position $\mathbf{y}$ with natural non-folding constraints we precise hereafter. The other sites are hold fixed *i.e.* $\mathbf{y}_i(1) = \mathbf{y}_i(0), \forall i \neq j$. Set $\mathbf{y}(1) = \{\mathbf{y}_i(1), i = 1, 2, \dots, n\}$ where $\mathbf{y}_j(1)$ is chosen as follows:
  - if $\Delta_0 U = U(\mathbf{y}_1(1), \dots, \mathbf{y}, \dots, \mathbf{y}_n(1), \beta^\star) - U(\mathbf{y}(0), \beta^\star) \leq 0$ then take $\mathbf{y}_j(1) = \mathbf{y}$;
  - otherwise sample an uniform law V in $[0, 1]$:
    - if $V \leq \exp(-\Delta_0 U / c_0)$ take $\mathbf{y}_j(1) = \mathbf{y}$;
    - otherwise keep $\mathbf{y}_j(1) = \mathbf{y}_j(0)$.
- step $k > 0$:
  - if $\lfloor k/n \rfloor = \lfloor (k-1)/n \rfloor$ then proceed as in the step 0 by replacing 0 by k and 1 by $k + 1$;
  - otherwise minimise (3) with respect to $\beta^\star$ ($\mathbf{y}^\star = \mathbf{y}(k)$ is hold fixed) with a Marquardt-type algorithm. Let $\beta(k)$ be the estimation of β at step k. Then proceed as in the step 0 by replacing 0 by k and 1 by $k + 1$.
- stopping criterion: if $U(\mathbf{y}(pn), \beta(pn)) - U(\mathbf{y}((p+1)n), \beta((p+1)n)) < 10^{-8}$ for two consecutive values of the integer p we stop the algorithm.

This algorithm is written in C language.

2.2.2 Description of the non-folding constraints

These constraints account for global and local non-foldings.

First, we embed G in a rectangle $\mathcal{R}$. Second, we construct the Delaunay triangulation for the n geographical sites plus the 4 vertices of $\mathcal{R}$ as well as the 4 mid-points of its 4 edges. To compute the Delaunay triangulation we use a program written by Shewchuk (1996).

At step k, for each site i we identify all the triangles for which this site is a vertex. Then consider the polygon P_i composed with the aggregation of these triangles and denote using $A_{i,1}, A_{i,2}, \dots, A_{i,q_i}$ its vertices, where q_i is the number of triangles. We assume

that these vertices are ordered clockwise and denote using $(A^1_{i,l}, A^2_{i,l})$ their coordinates , $l = 1, 2, \ldots, q_i$.

To ensure non-folding constraints in our algorithm when the candidate point j is moved at a position $\mathbf{y}$, in other words to avoid the case where triangles overlap, we impose that this new position is chosen uniformly in a convex set $\mathcal{C}^k_j$ which is the set of points $M_j = (M^1_j, M^2_j)$ such that, for $l = 1, 2, \ldots, q_j$:

$$M^1_j(A^2_{j,l} - A^2_{j,l+1}) + M^2_j(A^1_{j,l+1} - A^1_{j,l}) + A^1_{j,l}A^2_{j,l+1} - A^1_{j,l+1}A^2_{j,l} < 0, \tag{4}$$

with the convention $q_j + 1 = 1$. In other words, this set is the kernel of P_j that is the set of all points $M_j \in P_j$ such that the line segment $[M_j, \mathbf{x}] \subset P_j$ for each $\mathbf{x} \in P_j$. The set $\mathcal{C}^k_j$ corresponds to the marked area of Figure 2. Note that the computation of this region is fairly easy since it consists of a set of linear inequalities like (4).

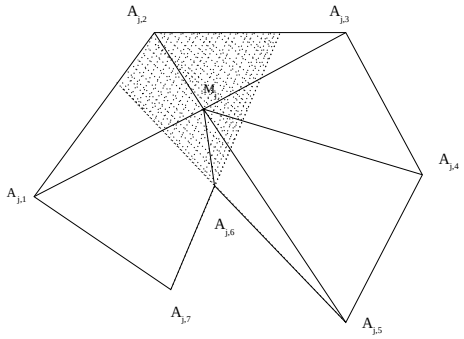

Figure 2: The marked area corresponds to the acceptable move for M_j.

These constraints mean that we impose moves that preserve the topological structure of the Delaunay triangulation the same.

Note that the rectangle $\mathcal{R}$ and its 8 points are hold fixed in the minimisation procedure. As illustrated by the left-hand plot of Figure 4, the embedding of G in $\mathcal{R}$ makes it possible to write the previous constraints (4) in a similar way for all the n sites: indeed, any one of the n site is included in a polygon for which the vertices are also vertices of the Delaunay triangulation for the n sites plus the 8 points of $\mathcal{R}$ represented by ∎. Without this embedding we should distinguish between the boundary points of G and its interior points in the determination of the acceptable move in the algorithm. Therefore, with this embedding it is no longer necessary to distinguish between the boundary points of G and its interior points. Moreover, the virtual links between the sites in G and the 8 points of $\mathcal{R}$ prevent our algorithm from global folding of the triangulation in itself.

2.2.3 Estimating the deformation on the whole G-space

Any point $\mathbf{x}$ in G belongs to a unique triangle for which the vertices are three monitoring sites, say $\mathbf{x}_i$, $\mathbf{x}_j$ and $\mathbf{x}_k$. Thus any point $\mathbf{x}$ is uniquely defined by its barycentric coordinates in the triangle $\{\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k\}$ as follows:

$$\mathbf{x} = a_{\mathbf{x},i}\mathbf{x}_i + a_{\mathbf{x},j}\mathbf{x}_j + a_{\mathbf{x},k}\mathbf{x}_k,$$

where $(a_{\mathbf{x},i}, a_{\mathbf{x},j}, a_{\mathbf{x},k}) \in [0,1]^3$ such that $a_{\mathbf{x},i} + a_{\mathbf{x},j} + a_{\mathbf{x},k} = 1$. We define the estimation $\hat{\Phi}(\mathbf{x})$ at point $\mathbf{x}$ of the deformation Φ as follows:

$$\hat{\Phi}(\mathbf{x}) = a_{\mathbf{x},i}\hat{\mathbf{y}}_i + a_{\mathbf{x},j}\hat{\mathbf{y}}_j + a_{\mathbf{x},k}\hat{\mathbf{y}}_k,$$

where $\hat{\mathbf{y}}_i$, $\hat{\mathbf{y}}_j$ and $\hat{\mathbf{y}}_k$ are estimated with the simulated annealing algorithm. So defined, the estimation $\hat{\Phi}$ holds the following features: continuous; bijective; it is an interpolant, *i.e.* $\hat{\mathbf{y}}_i = \hat{\Phi}(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$.

3 Applications

3.1 Illustrative example

In this example we suppose that the correlation is of the form:

$$r(\mathbf{x}, \mathbf{x}') = \exp(-\|\Phi(\mathbf{x}') - \Phi(\mathbf{x})\|),$$

where Φ is a composition of three bijective functions of the type:

$$\begin{array}{rccl} \mathbf{f}_\alpha : \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ \mathbf{x} & \longmapsto & \mathbf{b} + (\mathbf{x} - \mathbf{b})(1 + a_1 \exp(-a_2 \|\mathbf{x} - \mathbf{b}\|^2)), \end{array}$$

with $\alpha = (a_1, a_2, \mathbf{b})$ a four-dimensional parameter, where $\mathbf{b} \in \mathbb{R}^2$ is the centre of the deformation, $a_1 > 0$ is a range parameter and $a_2 \in]-1, \frac{1}{2}\exp(\frac{3}{2})[$ denotes the intensity of stretching ($a_2 > 0$) or shrinking ($a_2 < 0$) ($a_2 = 0$ corresponds to the identity function). The form of the three functions $\mathbf{f}_\alpha$ we consider as well as the values of the parameters for these functions are given in Figure 3.

We assume that the correlations between the $n = 116$ sites located in the upper-left plot of Figure 1 are known. Our goal is the estimation of the deformation Φ. First we embed these sites in a rectangle and we build the Delaunay triangulation which is represented by the left-hand plot of Figure 4. Then we estimate the positions of the n sites in the D-space with the method described in the previous section (but without the estimation of $\beta = 1$ we suppose known in this example). In the cooling schedule we take $c_0 = 100$ and $\theta = 0.999$, and it takes 730 seconds CPU (on a Sun Ultra 10 model Creator, with a 300 MHz UltraSPARC-II i processor) to get the estimation of the 2×116 coordinates. The right-hand plot of Figure 4 represents both the true deformation and its estimation. The quality of our estimation method can be observed by comparing these two plots. Moreover, in the estimated D-space the isotropic structure is fairly well achieved, as it is shown by Figure 5.

3.2 Application to a precipitation data set

The data consist of monitored precipitation data from $n = 20$ sites in Languedoc-Roussillon in the south of France, for which the geographical configuration is shown in the left-hand plot of Figure 6.

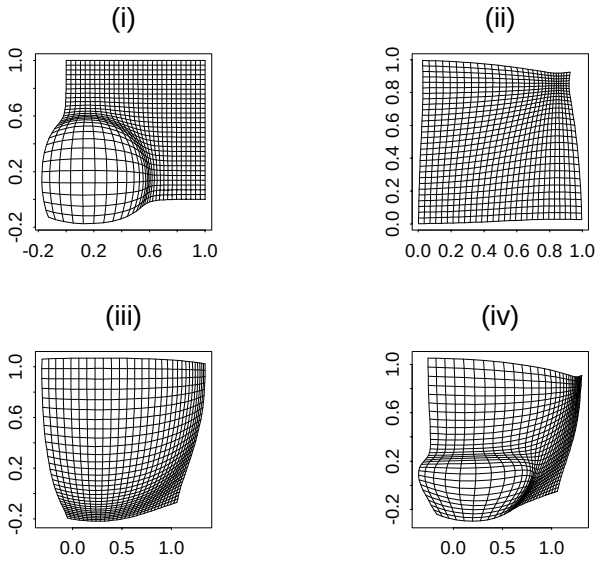

Figure 3: (i) $\mathbf{f}_{\alpha_1}$ with $\alpha_1 = (14, 1.6, 0.15, 0.15)$; (ii) $\mathbf{f}_{\alpha_2}$ with $\alpha_2 = (4, -0.6, 0.85, 0.85)$; (iii) $\mathbf{f}_{\alpha_3}$ with $\alpha_3 = (2, 1.4, 0.25, 0.95)$; (iv) their composition $\Phi = \mathbf{f}_{\alpha_3} \circ \mathbf{f}_{\alpha_2} \circ \mathbf{f}_{\alpha_1}$.

These data are in the form of 10-day aggregates, giving 6 records during November and December each year from 1975 through 1992. These data hold the following features: *(i)* a low frequency of missing values; *(ii)* a similar altitude (between 0 and 200 meters) for all the sites so that an altitude correction is pointless; *(iii)* a homogeneous period in the year so that a seasonal adjustment is not necessary. As pointed out by Meiring (1995), means and variances are positively related, and therefore, sample correlations between observations are calculated on the log scale, after adding 1 to all observations. The sample correlation estimates for a pair of sites is based on all the time points for which both sites have observations. The sample correlations for different pairs of sites are sometimes based on different numbers of observations, so that the sample correlations matrix is not positive definite. However, the model (1) fitted to the sample correlation remains positive definite.

The right-hand plot in Figure 7 represents the geographical inter-site distances *versus* the corresponding correlations $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i < j \leq n$, defined by (2), to which one would fit an isotropic correlation model if we assume isotropy for Z (note this would mean that Φ is the identity function in (1)). According to this plot, we decide to use the exponential model:

$$\rho_\beta(u) = \exp(-\beta u), \quad \beta > 0,$$

so that the objective function (3) is re-written as follows:

$$U(\mathbf{y}_1^\star, \mathbf{y}_2^\star, \ldots, \mathbf{y}_n^\star, \beta^\star) = \sum_{i<j} [\hat{r}(\mathbf{x}_i, \mathbf{x}_j) - \exp\left(-\beta^\star \|\mathbf{y}_j^\star - \mathbf{y}_i^\star\|\right)]^2. \tag{5}$$

To minimise (5) we apply our constrained procedure described in Paragraph 2.2. The Delaunay triangulation for the 20 sites is shown in Figure 6. In the cooling schedule we

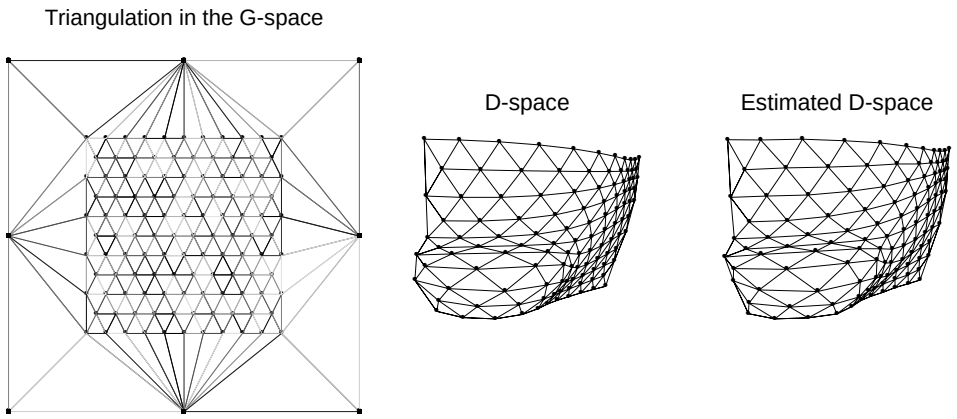

Figure 4: Delaunay triangulation for the $n = 116$ sites plus the 4 vertices of the embedding rectangle as well as the 4 mid-points of its 4 edges. True deformation (D-space) of the Delaunay triangulation restricted to the $n = 116$ sites and its estimation.

take $c_0 = 1000$ and $\theta = 0.99$. The right-hand plot of Figure 6 represents the estimated deformation of the Delaunay triangulation. The upper right-hand plot of Figure 7 shows the fitted exponential correlation as a function of the D-space coordinates. The G-space has been stretched in regions of relatively lower correlations, and shrunk in regions of relatively higher correlations, so that the exponential correlation better models the correlations in the D-space representation (minimum of the objective function is equal to 0.023) than in the G-space system (minimum of the objective function is equal to 0.155). Improvement of this fitting is illustrated by the interquartile intervals: the empirical correlations are less scattered in the D-space than in the G-space.

4 Application to the prediction

We point out one possible application where the space deformation model (1) may be useful. This is the prediction using Kriging. We refer to Cressie (1993) for a presentation of the Kriging method.

In practice, the mean and the variance fields are very seldom known, and must be predicted. As Høst *et al.* (1995) illustrate, separate modelling of the mean, variance and residual fields from monitoring data collected in space and time, may give very valuable information about the standard errors in spatial interpolation. Prediction, of each of the mean, variance and residual fields contributes to the overall spatial interpolation errors. However, estimation of the mean and of the variance is not the topic of this work and we concentrate only on the prediction of the centred and standardised random field Z.

To predict the centred and standardised random process Z at any location $\mathbf{x} \in G$ we

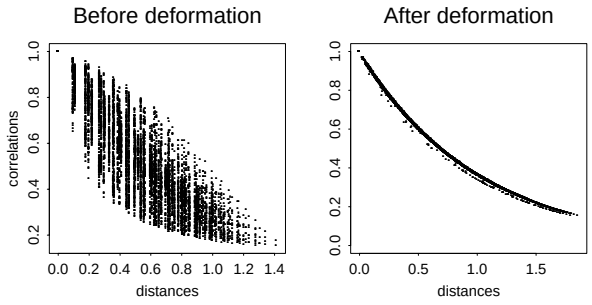

Figure 5: On the left: inter-site distances $\|\mathbf{x}_j - \mathbf{x}_i\|$ in the G-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j)$. On the right: inter-site distances $\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_i)\|$ in the estimated D-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j)$.

use the simple Kriging predictor:

$$\hat{Z}(\mathbf{x}) = \sum_{l=1}^{n} \lambda_l Z(\mathbf{x}_l),$$

with $(\lambda_l) = (\hat{r}_{i,j})^{-1} \left(\rho_{\hat{\beta}}(\|\hat{\Phi}(\mathbf{x}_l) - \hat{\Phi}(\mathbf{x})\|)\right)$ where (λ_l) denote the vector of the Kriging coefficients, $(\hat{r}_{i,j})$ the empirical correlation matrix of the vector $(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_n))$ and $\left(\rho_{\hat{\beta}}(\|\hat{\Phi}(\mathbf{x}_l) - \hat{\Phi}(\mathbf{x})\|)\right)$ the correlation vector between the "known sites" $(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_n))$ and the "unknown site" $Z(\mathbf{x})$.

One way to check whether the deformation may improve the prediction is a cross validation study. More precisely, for each of the T repetitions, we set aside a site and we predict the value of Z at this site with the $n-1$ remaining sites using the simple Kriging predictor described above. We repeat this operation for each of the n sites T times. Then we calculate the mean square error prediction ($MSEP$) that has to be compared with the one we would obtain by using the fitted isotropic correlation function if we assume isotropy for Z. We apply this cross validation study to the precipitation data set for each of the four following isotropic correlation models:

$$\begin{aligned}
\rho_E(u) &= \exp(-\beta_1 u), \text{ exponential model } (E);\\
\rho_{EN}(u) &= \beta_2 \exp(-\beta_1 u), \text{ exponential model with nugget } (EN);\\
\rho_G(u) &= \exp(-\beta_1 u^2), \text{ Gaussian model } (G);\\
\rho_{GN}(u) &= \beta_2 \exp(-\beta_1 u^2), \text{ Gaussian model with nugget } (GN);
\end{aligned}$$

where $\beta_1 > 0$ and $\beta_2 \in]0, 1[$.

The results of the cross validation study are given in the following Table which gives the $MSEP$ before and after the deformation for each of the four previous models:

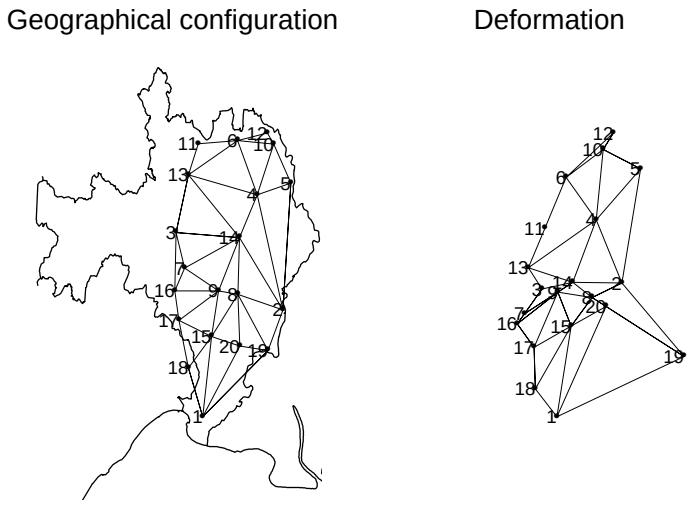

Figure 6: On the left: site locations and the corresponding Delaunay triangulation without the rectangle (the outlines indicate the French department of Gard and the coast). On the right: deformation of the triangulation.

	before deformation	after deformation	% of improvement
E	0.224	0.129	42.4
EN	0.203	0.136	33.1
G	0.271	0.132	51.3
GN	0.180	0.115	36.1

We deduce that the best improvement in terms of prediction is obtained for the Gaussian model: on average the prediction at the monitoring sites is 51.3 % better in the D-space than in the G-space. We also deduce that the best model is the Gaussian model both with and without deformation: nevertheless, with this model the prediction at the monitoring sites is 36.1 % better in the D-space than in the G-space. In conclusion, we can claim that the model with deformation for the correlation of the random field is better than the model without deformation.

5 Discussion

Simulated annealing appears to be a suitable tool for estimating a non-stationary structure. Combined with non-folding constraints, it gives the opportunity to estimate non-parametrically the bijective spatial deformation Φ, in model (1), in such a way that the estimation is bijective; so far the previous method using a pair of thin-plate splines does not guarantee this feature. However we wish to note one research question: so far, the choice of the isotropic correlation model has been made by visual inspection. One future

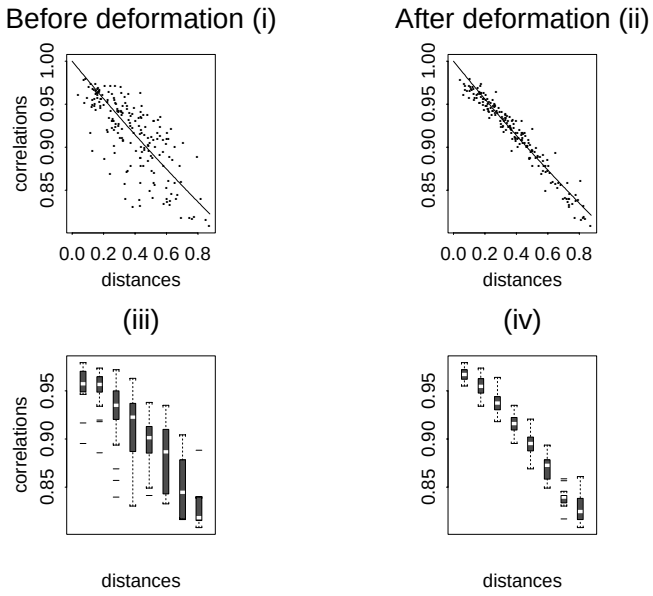

Figure 7: (i) Inter-site distances $\|\mathbf{x}_j - \mathbf{x}_i\|$ in the G-space *versus* empirical correlations $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$ and the fitted exponential correlation model (solid line) (ii) Inter-site distances $\|\hat{\Phi}(\mathbf{x}_j) - \hat{\Phi}(\mathbf{x}_i)\|$ in the estimated D-space *versus* empirical correlations $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$ and the fitted exponential correlation model (solid line) (iii) Box-plots of the empirical correlations as a function of the distances in the G-space (iv) Box-plots of the empirical correlations as a function of the distances in the D-space.

direction would be to establish a rational choice, by using for instance an Akaïke-type criterion. Furthermore, an attempt to use a non-parametric family of correlation models has yet to be developed.

Acknowledgements

Olivier Perrin has been supported by INRA, France; and by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data", Grant #ERB-FMRX-CT960095, while he was visiting the department of Mathematical Statistics at Chalmers University of Technology and Göteborg University in Sweden. The authors are grateful to Dominique Courault from Unité de Bioclimatologie, INRA-Avignon, France, for providing the precipitation data set, and to Jonathan Richard Shewchuk (1996) for providing the program that performs the Delaunay triangulation. We are also grateful to Xavier Guyon for giving us the idea of this work.

References

Aarts, E. and Korst, T.J. (1989), *Simulated annealing and Boltzman machines: stochastic approaches to combinatorial optimization and neural computing*, Wiley.

Azencott, R. (ed) (1992), *Simulated annealing: parallelization techniques*, Wiley.

Cressie, N.A.C. (1993), *Statistics for spatial data*, Revised Edition, Wiley-Interscience.

Geman, D. (1990), *Random fields and inverse problem in imaging*, L.N.M. no. 1427, Springer.

Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, 6, 721-741.

Guttorp, P. and Sampson, P.D. (1994), "Methods for estimating heterogeneous spatial covariance functions with environmental applications," in: Patil, G.P. and Rao, C.R. (eds), *Handbook of Statistics XII: Environmental Statistics*, Elsevier/North Holland, New York, 663-690.

Guttorp, P., Sampson, P.D. and Newman, K. (1992), "Nonparametric estimation of nonstationary spatial covariance structure with application to monitoring network design," in: Walden, A. and Guttorp, P. (eds), *Statistics in Environmental and Earth Sciences*, Edward Arnold, London, 39-51.

Hajek, B. (1988), "Cooling schedules for optimal annealing," *Mathematics of operations research*, 13 (2), 311-329.

Høst, G., Omre, H. and Switzer, P. (1995), "Spatial interpolation errors for monitoring data," *Journal of the American Statistical Association*, 90, 853-861.

Kirkpatrick, S., Gelatt, Jr., C.D., and Vecchi, M.P. (1983), "Optimization by simulated annealing," *Science*, 220, 671-679.

Meiring, W. (1995), "Estimation of heterogeneous space-time covariance," *PhD thesis*, University of Washington, Seattle.

Meiring, W., Guttorp, P. and Sampson, P.D. (1998), "Computational issues in fitting spatial deformation models for heterogeneous spatial correlation," *Computing Science and Statistics*, 29 (1), (Scott, D.W. (ed)), 409-417.

Meiring, W., Monestiez, P., Sampson, P.D. and Guttorp, P. (1997), "Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data," in: Baafi, E.Y. and Schofield, N. (eds), *Geostatistics Wollongong '96*, 1, Kluwer Academic Publishers, 162-173.

Monestiez, P., Sampson, P.D. and Guttorp, P. (1993), "Modelling of heterogeneous spatial correlation structure by spatial deformation," *Cahiers de Geostatistique*, Fascicule 3, Compte Rendu des Journées de Géostatistique, 25-26 May 1993, Fontainebleau. Published by the École Nationale Supérieure des Mines de Paris.

Perrin, O. and Meiring, W. (1998), "Identifiability for non-stationary spatial structure," to appear in *Journal of Applied Probability*, 36 (4).

Perrin, O. and Monestiez, P. (1998), "Modelling of non-stationary spatial structure using parametric radial basis deformations," to appear in: Soares, A., Gómez-Hernandez, J. and Froidevaux, R. (eds), *geoENV98*, Kluwer Academic Publishers.

Perrin, O. and Senoussi, R. (1998), "Reducing non-stationary random fields to stationarity or isotropy using a space deformation," submitted for publication.

Sampson, P. and Guttorp, P.D. (1992), "Nonparametric estimation of nonstationary spatial covariance structure," *Journal of the Americal Statistical Association*, 87, 108-119.
Sampson, P.D., Guttorp, P. and Meiring, W. (1994), "Spatio-temporal analysis of regional ozone data for operational evaluation of an air quality model," *Proceedings of the Section on Statistics and the Environment*, American Statistical Association.
Shewchuk, J. R. (1996), Triangle: engineering a $2D$ quality mesh generator and Delaunay triangulator, *First Workshop on Applied Computational Geometry* (Philadelphia, Pennsylvania), 124-133, ACM. Available online from http://www.cs.cmu.edu/ quake/triangle.research.html.

Fabio Spizzichino

"La Sapienza", Rome

Exchangeable heterogeneous populations and computation of probability distributions for vectors of "occupation numbers"

ABSTRACT: We consider situations of heterogeneity of the following type: let $\mathcal{P}$ be a population formed with n individuals $U_1, ..., U_n$; to U_i $(i = 1, ..., n)$ we attach an observable random variable T_i and an unobservable random variable Z_i. The latter is an endogenous variable which describes the "type" of the individual U_i and let $\mathcal{Z}$ denote the space of values of the $Z_i's$. $T_1, ..., T_n$ are conditionally independent, given $(\mathbf{Z} = \mathbf{z})$, and, more in particular, the conditional distribution of T_i, given $(\mathbf{Z} = \mathbf{z})$, only depends on z_i; for a given, dominated, family $\{G_z\}_{z\in\mathcal{Z}}$, such a conditional distribution coincides with G_{z_i}, i.e. we have

$$P\{T_1 \leq t_1, ..., T_n \leq t_n | Z_1 = z_1, ..., Z_n = z_n\} = \prod_{i=1}^{n} G_{z_i}(t_i).$$

We consider in details the discrete, exchangeable, case when $\mathcal{Z} \equiv \{1, 2, ..., D\}$ and $(Z_1, ..., Z_n)$ has a joint exchangeable law $\mathcal{L}_n$. It is easy to see that $(T_1, ..., T_n)$ is exchangeable as well. Such a model for $(T_1, ..., T_n)$ is then described by the joint survival function

$$P\{T_1 \leq t_1, ..., T_n \leq t_n\} = \sum_{\mathbf{z}} P\{Z_1 = z_1, ..., Z_n = z_n\} \prod_{i=1}^{n} G_{z_i}(t_i)$$

will be denoted by the symbol $\mathcal{H}\left(n; \mathcal{Z}; \{G_z\}_{z\in\mathcal{Z}}; \mathcal{L}_n\right)$

It is interesting to notice that the $m-$dimensional $(m < n)$ marginal distribution of $\mathcal{H}\left(n; \mathcal{Z}; \{G_z\}_{z\in\mathcal{Z}}; \mathcal{L}_n\right)$ is $\mathcal{H}\left(m; \mathcal{Z}; \{G_z\}_{z\in\mathcal{Z}}; \mathcal{L}_m\right)$ where $\mathcal{L}_m$ denotes the $m-$dimensional marginal of $\mathcal{L}_n$. Furthermore, due to conditional independence, we have the following:

Proposition 1. *Let E be an event in the $\sigma-$algebra generated by $T_1, ..., T_n$; then the conditional distribution of $T_1, ..., T_n$, given E is of the form $\mathcal{H}\left(n; \mathcal{Z}; \{G_z|E\}_{z\in\mathcal{Z}}; \mathcal{L}_n|E\right)$.*

Due to possible complexity in computations of the distributions of the type $\mathcal{L}_m$ $(1 \leq m \leq n)$, and $\mathcal{L}_n|E$ (from which one would obtain corresponding distributions for $T_1, ..., T_n$) one can be rather interested in their simulation.

To this purpose, when n is big compared with D, it can be convenient to consider the vector of "occupation numbers" associated to $(Z_1, ..., Z_n)$, which are the random variables defined by:

$$\Lambda_j = \sum_{i=1}^{n} \mathbf{1}_{\{Z_i=j\}}, j = 1, ..., D.$$

We shall also write $\Lambda_j = \phi_j(\mathbf{Z})$ and $\Lambda = \phi(\mathbf{Z})$ with $\phi_j(\mathbf{z}) = \sum_{i=1}^n \mathbf{1}_{\{z_i=j\}}$ and denote by $\nabla_{n,D}$ the space of possible values of Λ : the elements of $\nabla_{n,D}$ are the $D-$dimensional vectors with non-negative integer elements, having sum equal to n.

Different MCMC methods may be natural for the simulation of distributions on $\nabla_{n,D}$ and the following considerations can be of interest, concerning the simulation of distributions for $(Z_1, ..., Z_n)$ and then for $(T_1, ..., T_n)$. First we notice that there is a one-to-one correspondence between exchangeable distributions over $\mathcal{Z}^n$ (i.e.for $(Z_1, ..., Z_n)$) and distributions over $\nabla_{n,D}$; more precisely it is

$$P\{\mathbf{Z} = \mathbf{z}\} = \frac{\prod_{j=1}^D \phi_j(\mathbf{z})}{n!} P\{\Lambda = \phi(\mathbf{z})\}.$$

Since Λ takes values in the finite set $\nabla_{n,D}$, it admits joint moments of any finite order. For $\mathbf{h} \equiv (h_1, ..., h_D)$, $h_j = 0, 1, ...$, let us denote by $\mu(\mathbf{h})$ the joint moment $\mu(\mathbf{h}) = \mathbb{E}\left(\prod_{j=1}^D (\Lambda_j)^{h_j}\right)$.

The joint distribution $\mathcal{L}_n$ of $\mathbf{Z}$ determines $\mu(\mathbf{h})$, for all $\mathbf{h}$. Viceversa, from the knowledge of the joint moments $\mu(\mathbf{h})$ for all $\mathbf{h}$, we can recover $\mathcal{L}_n$ and its marginal distributions. More precisely, we have (see [Gerardi, Spizzichino, Torti (2000a)])

Proposition 2. *For $m \leq n$ the distribution $\mathcal{L}_m$ of $\mathbf{Z}$ is determined by the set $\{\mu(\mathbf{h})\}_{\sum_j h_j \leq m}$ of joint moments of Λ of order not larger than m.*

Corollary 3. *The joint distribution of $(T_1, ..., T_m)$ is determined by $\{G_z\}_{z\in\mathcal{Z}}$ and $\{\mu(\mathbf{h})\}_{\sum_{z\in\mathcal{Z}} h_z \leq m}$.*

Remark 4. We consider the case when the space $\mathcal{Z}$ of possible values for the endogenous variables $Z_1, ..., Z_n$ is finite; however in some applications it is not necessarily endowed with an intrinsic complete ordering: for instance in the case when there is a double classification for individuals $U_1, ..., U_n$, Z_i is a pair $Z_i \equiv \left(Z_i^{(1)}, Z_i^{(2)}\right)$ and $\mathcal{Z} \equiv \mathcal{Z}^{(1)} \times \mathcal{Z}^{(2)}$ is the space of values of Z_i , with $\mathcal{Z}^{(1)} \equiv \{1, 2, ..., R\}, \mathcal{Z}^{(2)} \equiv \{1, 2, ..., S\}$, $D = R \times S$, say. In this case, the set of occupation numbers can be looked at as the $R \times S$ random matrix $(\Lambda_{r,s})$, where obviously

$$\Lambda_{r,s} = \sum_{i=1}^n \mathbf{1}_{\{Z_i=(r,s)\}}, r = 1, ..., R, s = 1, ..., S; \qquad \sum_{r=1}^R \sum_{s=1}^S \Lambda_{r,s} = n.$$

Our interest for models of the type $\mathcal{H}(n; \mathcal{Z}; \{G_z\}_{z\in\mathcal{Z}}; \mathcal{L}_n)$ in particular arised from applications in the fields of reliability and survival analysis, where the variables $T_1, ..., T_n$ are non-negative (*lifetimes* of individuals $U_1, ..., U_n$). In such fields it is natural to consider conditional survival probabilities for *residual lifetimes* given an observed history of *failure and survivals*, of the type

$$P\{T_{h+1} > t + \tau_{h+1}, ..., T_n > t + \tau_n | H_t\} \tag{1}$$

with

$$H_t \equiv \{T_1 = t_1, ..., T_h = t_h, T_{h+1} > t, ..., T_n > t\}, 0 < t_1 < t_2 < ... < t_h < t; \qquad (2)$$

notice that rearrangment of indexes is allowed, in view of exchangeability, and the history in (2) can be seen as the history observed in the time-interval $[0 \leq s \leq t]$ for the process $N(s)$ which counts, for any s, the total number of items failed up to time s:

$$N(s) = \sum_{j=1}^{n} 1_{\{Tj \leq s\}} \qquad (3)$$

Denote by $\Lambda(t)$ the vector of occupation numbers for the subpopulation $\mathcal{P}^{n-N(t)}$ formed by the only individuals which survived at time t, i.e. let, for

$$N_j(t) = \sum_{i=1}^{n} \mathbf{1}_{\{T_i \leq t\}} \mathbf{1}_{\{Z_i = j\}}, \qquad (4)$$

$$\Lambda_j(t) = \Lambda_j - N_j(t), j = 1, ..., D, \qquad (5)$$

$\Lambda(t) = \{\Lambda_j(t)\}_{j \in \mathcal{Z}}$ is then a pure jump process such that $\sum_{j=1}^{d} \Lambda_j(t) = n - N(t)$ and it is a non-homogeneous Markov process (in that the intensities of $N_1(t)..., N_D(t)$ at time t depend only on t and $\Lambda(t)$); its jump times are $T_{(1)}, ..., T_{(n)}$, which are the order statistics of $T_1, ...T_n$.

In view of Propositions 1 and 2, the problem of deriving the conditional distribution in (1) substantially reduces to that of computing the conditional law of $\Lambda(t)$ given the history $\{N(s), s \leq t\}$. Formally the problem is one to find the conditional law of a Markov process given the point process which counts its jumps. This is then a problem in the theory of stochastic filtering. This theory provides the tools to prove the formula below (see [Gerardi, Spizzichino, Torti (2000b)]): for $\mathbf{x} \in \nabla_{n-N(t),D}$

$$P\left(\Lambda(t) = \mathbf{x} \mid H_t\right) \propto$$

$$\propto \overline{F}_t(\mathbf{x}) \sum_{k_1,...,k_{N(t)}} \prod_{j=1}^{N(t)} g_{k_j}(T_{(j)}) \left[x_{k_j} + \sum_{i=1}^{N(t)} \delta_{k_j k_i}\right] P\left(\Lambda = \mathbf{x} + \sum_{i=1}^{N(t)} e^{k_i}\right) \qquad (6)$$

where

$$\overline{F}_t(\mathbf{x}) \equiv \prod_{j=1}^{D} \left[\overline{G}_j(t)\right]^{x_j}; \mathbf{e}^{(z)} \equiv (\delta_{1,z}, ..., \delta_{D,z}), z = 1, ..., D \qquad (7)$$

Complexity involved in the formula shows the need of using suitable simulation techniques. In this respect, we highligh that the distribution to be simulated is one on the

space $\nabla_{n-N(t),D}$ and we can consider a Metropolis algorithm for simulating $P(\Lambda(t) = \mathbf{x} \mid H_t)$, starting with a symmetric, irreducible chain on $\nabla_{n-N(t),D}$ having transition probabilities of the type

$$q_{\mathbf{x},\mathbf{y}} = \begin{cases} 0 & \text{if } \sum_{j=1}^{D} |x_j - y_j| > 1 \\ k(\mathbf{x},\mathbf{y}) & \text{if } (\mathbf{x},\mathbf{y}) \in \Sigma_{j_1,j_2} \text{ for some } 1 \leq j_1 \neq j_2 \leq D \end{cases}$$

where

$$\Sigma_{j_1,j_2} \equiv \{(x,y) | x_{j_1} > 0, y_{j_2} > 0; x_{j_1} = y_{j_1} + 1, x_{j_2} = y_{j_2} - 1; x_j = y_j, j \neq j_1, j_2$$

(only transitions to adjacent states are allowed) and then computing ratios of the type $\frac{P(\Lambda(t)=\mathbf{y}|H_t)}{P(\Lambda(t)=\mathbf{x}|H_t)}$. In view of (6) and (7), we have, for $(\mathbf{x},\mathbf{y}) \in \Sigma_{j_1,j_2}$, $\frac{\overline{F}_t(\mathbf{y})}{\overline{F}_t(\mathbf{x})} = \frac{\overline{G}_{j_2}(t)}{\overline{G}_{j_2}(t)}$ and

$$\frac{P(\Lambda(t) = \mathbf{y} \mid H_t)}{P(\Lambda(t) = \mathbf{x} \mid H_t)} = \frac{\overline{G}_{j_2}(t)}{\overline{G}_{j_2}(t)} \times$$

$$\frac{\sum_{k_1,\ldots,k_{N(t)}} \prod_{j=1}^{N(t)} g_{k_j}(t_{(j)}) \left[y_{k_j} + \sum_{i=1}^{N(t)} \delta_{k_j k_i} \right] P\left(\Lambda = \mathbf{y} + \sum_{i=1}^{N(t)} e^{k_i} \right)}{\sum_{k_1,\ldots,k_{N(t)}} \prod_{j=1}^{N(t)} g_{k_j}(t_{(j)}) \left[x_{k_j} + \sum_{i=1}^{N(t)} \delta_{k_j k_i} \right] P\left(\Lambda = \mathbf{x} + \sum_{i=1}^{N(t)} e^{k_i} \right)}$$

Before concluding, we notice that the assumption of exchangeability for the endogenous variables $Z_1, ..., Z_n$, even though a strong one, is a direct extension of the conditions of being i.i.d. or conditionally i.i.d, which have often been considered in the literature (see e.g. [Richardson, Green (1997)], though for a different kind of mixture models)

References

A. Gerardi, F. Spizzichino, B. Torti (2000a), "Exchangeable mixture models for lifetimes: the role of occupation numbers". Stat. Prob. Lett. (to appear).

A. Gerardi, F. Spizzichino, B. Torti (2000b), "Filtering equations for the conditional law of residual lifetimes from a heterogeneous population". J. Appl. Prob. (to appear).

S. Richardson, P. J. Green (1997), "On Bayesian Analysis of Mixtures with an Unknown Number of Components". J.R. Stat. Soc., B, 731-792.

Matthew Stephens

Oxford University

Computationally-Intensive Inference in Molecular Population Genetics

Abstract: Recent experimental advances have lead to an explosion of data documenting diversity in modern natural populations. These kinds of data present a considerable computational challenge, even for sophisticated modern statistical inference methods. The problem is of considerable practical importance and has attracted recent attention, with the development of algorithms based on importance sampling (IS) and Markov chain Monte Carlo (MCMC).

We will begin our talk by introducing some of the models relevant to the study of molecular population genetic data. These models typically focus on (aspects of) the genealogical tree relating the sampled individuals, which is usually unobserved, and may be treated as "missing data". The very high dimension of this missing data is the main reason that these problems are so challenging. We will describe some specific IS and MCMC approaches which have been suggested for dealing with this missing data, and compare their performance on some of the simplest inference problems which arise in this field. The results of these comparisons suggest some insights for computationally intensive inference in problems with high-dimensional missing data which we hope will be of more general interest.

William Stewart

North Carolina State University, Raleigh

Numerical methods for Markov chains

ABSTRACT: In these lectures our attention is directed at computational methods for computing stationary distributions of finite irreducible Markov chains. We let q_{ij} denote the rate at which an n-state Markov chain moves from state i to state j. The $n \times n$ matrix Q whose off-diagonal elements are q_{ij} and whose i^{th} diagonal element is given by $-\sum_{j=1,j\neq i}^{n} q_{ij}$ is called the *infinitesimal generator* of the Markov chain. It may be shown that the stationary probability vector π, a row vector whose k-th element denotes the stationary probability of being in state k, can be obtained by solving the homogeneous system of equations $\pi Q = 0$. Alternatively, the problem may be formulated as an eigenvalue problem $\pi P = \pi$, where $P = Q\Delta t + I$ is the stochastic matrix of transition probabilities, (Δt must be chosen sufficiently small so that the probability of two or more transitions occurring in time Δt is small, i.e., of order $o(t)$). Mathematically, the problem is therefore quite simple. Unfortunately, problems arise from the computational point of view because of the large number of states which many systems may occupy. It is not uncommon for thousands of states to be generated even for simple applications.

We begin our discussion with an examination of the relative advantages and disadvantages of iterative and direct solution methods. We show that iterative methods are generally preferred, unless the infinitesimal generator has some special structure which makes a direct method more efficient. Next, we discuss direct methods and show how to implement them in a computationally efficient manner. Basic single vector iteration methods are also considered. In particular, we examine the power method, forward and backward Gauss-Seidel and SOR and preconditioned power iterations. Block single vector iterative methods are also considered. Following this, iterative methods that incorporate a subspace of vectors are considered. These methods go under the more generic name of *projection techniques*, and have been shown in comparison testing to be among the most effective for general Markov chain problems. The final methods considered for the computation of stationary solutions are decompositional methods described. These are valuable when the matrix is *nearly-completely-decomposable, NCD*, a situation which arises often in practice.

Dietrich Stoyan

TU Bergakademie, Freiberg

Statistical Characterization of Connectivity and Permeability of Porous Media

ABSTRACT: The study of transport phenomena in porous media is a very important research topic in physics and engineering. A big and still unsolved problem is the geometrical characterization of permeability. It is very difficult even if three-dimensional data are given, which may result from computer tomography and are usually lattice data. It is natural to use methods of random-sets statistics in this context, considering the set of pores as a sample of a random set. 'Classical methods' based on which use characteristics such as contact distribution and covariance functions yield interesting information on the size of pores, but are not able to characterize connectivity and percolation properties. Therefore, physicists have developed the concepts of local porosity distributions and local percolation probabilities.

The talk discusses these concepts and describes how the statistical analysis can be refined by dilating or eroding (or opening or closing) the system of pores by spheres of radius r. This yields valuable geometrical information, but can be seen also in the context of movement of spherical particles through the pores. An example for a characteristic arising in this context is the specific Euler characteristic seen as a function of the sphere radius r.

In the numerical calculations of the characteristics, the geometry of the lattice has to be considered; some operations which are unproblematic in Euclidean geometry have to be modified in the lattice case. Finally, new ideas of edge-correction of ratio estimators of random-set statistics are sketched.

Evaluation of first passage times of diffusion processes through boundaries by means of a totally simulative algorithm

Maria Teresa Giraudo, Laura Sacerdote and Cristina Zucca
Dept. of Mathematics University of Torino
V.C.Alberto 10 10123 Torino, Italy

ABSTRACT

In many contexts arising both from the theoretical and from the application point of view the necessity often arises to consider the first passage time of diffusion processes through a boundary rather than to describe the detailed evolution of such processes. Methods relying on simulation often appears to be the easiest approach to such problems, but they present hidden difficulties leading in many cases to unreliable results. Suitable improved techniques for the simulation of first passage times of diffusion processes have been recently introduced, relying on the evaluation at each time step of the crossing probabilities for the corresponding tied-down processes. Here we propose a revision of such methods based on the evaluation of the crossing probabilities via a pure Monte Carlo algorithm.

Keywords: Diffusion processes; First passage time; Simulation; Monte Carlo methods

1 Introduction

Diffusion processes are largely employed in literature to describe the dynamics of complex systems. As analytical or numerical techniques are available only in some instances, methods that make use of simulations appear to be the easiest approach to solve first passage time (FPT) problems for such processes. However, FPT simulation presents some hidden difficulties that can lead to unreliable results. Though a large literature exists about pathwise simulations of diffusion processes associated with stochastic differential equations (cfr. Honerkamp, J. (1994); Kloeden, P.E. and Platen, E. (1992) and references quoted therein), the work generally focuses on unbounded processes disregarding the problems arising when the sample paths are constrained by boundary conditions.

In order to minimize the error induced by possible undetected crossings of an absorbing boundary, in a previous paper (Giraudo, M.T. and Sacerdote, L. (1999)) a new technique for the estimation of FPTs for diffusion processes $X(t) = \{X(t),\ t \geq 0\}$ was

proposed which was based on the evaluation at each time step of the FPT probability for the corresponding tied-down processes taking at the ends of the time interval the values of the discretized process. Though highly reliable, the mathematical complexity of the algorithms involved makes them of difficult employment for those not mathematical researchers to whom they could be of great interest. Here we propose a revision of such method relying on the evaluation of the tied-down crossing probabilities by means of a pure Monte Carlo algorithm.

After a brief survey on the necessary mathematical background in Section 2, we introduce the new simulation method in Section 3 by describing in details the algorithm employed. An illustration of the features of the proposed method is finally done in Section 4 by means of some examples.

2 The FPT problem for diffusion processes

We limit ourselves to summarize here the basic definitions necessary to deal with the problem of simulation of FPTs for diffusion processes while referring to Karlin, S. and Taylor, H.M. (1981), Ricciardi, L.M. (1977) and Ricciardi, L.M. and Sato, S. (1990) for a detailed exposition.

Let $X(t) = \{X(t), t \geq 0\}$ be a time homogeneous one dimensional diffusion process defined over the diffusion interval $I = (l, r)$ where $l, r \in (-\infty, \infty)$ and let $\mu(x)$ and $\sigma^2(x)$ denote its drift and infinitesimal variance functions respectively. The Itô form of the stochastic differential equation (SDE) for the process $X(t)$ is (cf. Arnold, L. (1974))

$$\begin{cases} dX(t) = \mu\left[X(t)\right] dt + \sigma\left[X(t)\right] dW(t) \\ X(0) = x_0 \end{cases} \tag{1}$$

where $W(t) = \{W(t), t \geq 0\}$ is a standard Wiener process.

The conditional transition probability density function of $X(t)$ is defined as

$$f(x, t \mid y, \tau) = \frac{\partial}{\partial x} P(X(t) \leq x \mid X(\tau) = y), \quad \tau < t, x \in I, y \in I \tag{2}$$

Given a diffusion process $X(t)$ originated in x_0 at time $t_0 = 0$, the first passage time of $X(t)$ through a boundary $S > x_0$ is the random variable

$$T_S(x_0) = \inf\left\{t \geq 0 : X(t) \geq S; X(0) = x_0\right\}. \tag{3}$$

Its probability density function $g(S, t \mid x_0)$ is defined as

$$g(S, t \mid x_0) = \frac{dP(T_S(x_0) \leq t)}{dt} \tag{4}$$

and it can be obtained as the solution of the following Volterra second kind integral equation (cf. Giorno, V. et al. (1989)):

$$g(S, t \mid x_0) = -2\psi(S, t \mid x_0) + 2\int_0^t d\tau g(S, \tau \mid x_0)\psi(S, t \mid S, \tau), \quad x_0 < S. \tag{5}$$

Here

$$\psi(S,t \mid y,\tau) = \frac{dP(X(t) \leq S \mid X(\tau) = y)}{dt} + k(t) f(S,t \mid y,\tau), \tag{6}$$

where $k(t)$ is a suitable function that can be arbitrarily chosen to make the kernel of equation (2.5) regular at $t = \tau$.

Starting from the diffusion process $X(t) = \{X(t), t \geq 0\}$, we will denote as $X^{u,v}_{t.d.}(t) = \{X_{t.d.}(t), u \leq t \leq v\}$ the corresponding tied-down diffusion process constrained to take the values $X(u)$ and $X(v)$ at the time instants u and v, $u < v$, respectively and behaving otherwise as $X(t)$ for $u < t < v$. For the sake of simplicity, the FPT probability density function of $X^{u,v}_{t.d.}(t)$ will be denoted as $g_{t.d.}(S,t \mid \cdot)$. Given the drift $\mu(x)$ and infinitesimal variance $\sigma^2(x)$ of the unconstrained diffusion process, the analoguous functions for the corresponding tied-down process can be proved to be:

$$\begin{aligned} \mu^{u,v}_{t.d.}(x,t) &= \mu(x) + \frac{\sigma^2(x)}{f(b,v \mid x,t)} \frac{\partial f(b,v \mid x,t)}{\partial x}; \\ \left(\sigma^2\right)^{u,v}_{t.d.}(x,t) &= \sigma^2(x) \end{aligned}$$

where $f(x,t \mid y,\tau)$ is the conditional transition probability density function of the process $X(t)$.

3 The simulation algorithm

As the main purpose of this work is to develop a reliable totally simulative technique to estimate FPTs, we will not focus on the problems connected with the discretization of the SDEs involved. We just briefly recall that we will make use of a scheme proposed by Platen, E. and Wagner, W.(1983) achieving order of strong convergence $\gamma = 1.5$.

The most relevant source of error in the simulation of FPTs of diffusion processes through boundaries lies in disregarding the crossings that may happen inside each discretization interval. Trying to reduce such error, a new simulation technique was recently proposed in Giraudo, M.T. and Sacerdote, L. (1999). At each time step (τ_n, τ_{n+1}) of amplitude h, at the ends of which the process takes the values y_n and $y_{n+1} < S$ respectively, a suitable approximation of the FPTs probabilities $P_n(S, y_n, y_{n+1}) = \int_0^h g_{t.d.}(S,t \mid y_n)dt$ were evaluated. The random value of $T_S(x_0)$ was set equal to $\tau_n + \frac{h}{2}$ either when $y_{n+1} > S$ or in the case where $U_n < P_n$ where U_n is a $(0,1)$-uniformly distributed random number. Otherwise the simulation was allowed to go on and the procedure repeated at the successive step.

Here, in order to obtain a more manageable method, we propose a revision of the technique where the tied-down crossing probabilities are obtained by means of a totally simulative procedure.

At each discretization step (τ_n, τ_{n+1}) a suitable number $N_{t.d.}$ of simulation runs is carried on for the tied-down diffusion process originating in $X(\tau_n)$ and constrained to take the value $X(\tau_{n+1})$ at time τ_{n+1}. The relative frequency $M/N_{t.d.}$, where M is the

number of simulated tied-down sample paths out of $N_{t.d.}$ crossing S, is then taken as the probability to be compared with the random generated number as above.

In order to determine M with a good precision we make use of a sort of nested procedure. Precisely, if $y_{n+1} < S$, the value of $X_1 = X_{t.d.}^{0,h}(\tau_n + \frac{h}{2})$ is simulated. If $X_1 > S$, M is augmented by one. Otherwise, the procedure is repeated over the interval $\left(0, \frac{h}{2}\right)$ for the tied-down processes constrained in $X_{t.d.}^{0,h/2}(\tau_n), X_{t.d.}^{0,h/2}(\tau_n + \frac{h}{2})$ and, if no crossing has happened, over the interval $\left(\frac{h}{2}, h\right)$ for the tied-down processes constrained in $X_{t.d.}^{h/2,h}(\tau_n + \frac{h}{2}), X_{t.d.}^{h/2,h}(\tau_{n+1})$. If neither this step results in a crossing, the procedure is finally repeated in the same way successively for the nested intervals $\left(0, \frac{h}{4}\right), \left(\frac{h}{4}, \frac{h}{2}\right), \left(\frac{h}{2}, \frac{3h}{4}\right)$ and $\left(\frac{3h}{4}, h\right)$.

The succession of nested simulations is carried on $N_{t.d.}$ times for each discretization interval in order to obtain the required approximated value of $P_n(S, y_n, y_{n+1})$.

4 Numerical results

We will employ the simulation technique described in the previous Section in the two exemplificative cases of the Ornstein-Uhlenbeck (O.U.) and of the Feller diffusion processes. Even though analytical expressions for their FPT distributions are known only in a few instances, they can be computed numerically by solving suitable integrale equations (cf. Ricciardi, L.M. et al. (1984) and Giorno, V. et al. (1989)).

Henceforth we use the word "exact" for the results obtained via numerical evaluations while the word "simulated" is used for results arising from pathwise properties.

We briefly recall that the normalized O.U. process is solution of the SDE:

$$\begin{cases} dX(t) = -X(t)dt + \sqrt{2}dW(t) \\ X(0) = x_0 \end{cases} \tag{7}$$

with diffusion interval $I = (-\infty, \infty)$, while the SDE for the Feller diffusion process is

$$\begin{cases} dY(t) = (pY(t) + q)dt + \sqrt{2rY(t)}dW(t) \\ Y(0) = y_0 \end{cases} \tag{8}$$

where $p < 0$ and $r > 0$. The diffusion interval is $I = (0, \infty)$ and the lower boundary $x = 0$ changes its nature depending on the value of the parameters p, q and r. Here we always choose $q > r$, hence the origin is an entrance boundary (cf. Karlin, S. and Taylor, H.M. (1981)).

The infinitesimal coefficients for the two corresponding tied-down processes can be obtained by means of formulae (2.7).

In Tables I and II we report the confidence intervals for the mean FPT and for its variance obtained by means of the proposed simulation technique.

The O.U. process was originated in $x_0 = 0$ and constrained by an absorbing boundary in $S = 1$, while the Feller process with parameters $p = -1, q = 2, r = 1$ was originated in $y_0 = 2$ and constrained by an absorbing boundary in $S = 3$. Different values of h have been chosen to show the algorithm behavior as the discretization step increases. For

each value of h, $N = 1000$ sample paths were simulated, while the number of simulation runs for the tied-down processes at each step was $N_{t.d.} = 100$ for the O.U. process and $N_{t.d.} = 600$ for the Feller process. Other simulation batches have shown that in this latter case $N_{t.d.} = 200$ would have been sufficient to obtain satisfactory results. For the Ornstein-Uhlenbeck process the exact values of $E[T_S(x_0)]$ and of $Var[T_S(x_0)]$ are 2.09 and 5.84 respectively while for the Feller process $E[T_S(y_0)] = 1.41$ and $Var[T_S(y_0)] = 3.65$.

TABLE I: O.U. process

step h	$C.I.$(mean)	$C.I.$(variance)	$K.-S.$ significance test p-level
0.01	$[2.03, 2.32]$	$[5.24, 5.93]$	0.2
0.05	$[2.01, 2.29]$	$[6.48, 7.34]$	< 0.01
0.075	$[2.07, 2.37]$	$[5.50, 6.24]$	0.1
0.1	$[2.02, 2.31]$	$[5.22, 5.91]$	0.1
0.3	$[1.94, 2.24]$	$[5.43, 6.14]$	0.2

TABLE II: Feller process

step h	$C.I.$(mean)	$C.I.$(variance)	$K.-S.$ significance test p-level
0.01	$[1.34, 1.58]$	$[3.54, 4.00]$	0.2
0.05	$[1.35, 1.58]$	$[3.40, 3.85]$	0.2
0.075	$[1.39, 1.64]$	$[3.56, 4.03]$	0.1
0.1	$[1.35, 1.58]$	$[3.32, 3.76]$	0.1
0.3	$[1.29, 1.53]$	$[3.28, 3.72]$	0.1

In the last column of Tables I and II we reported the Kolmogorov-Smirnov significance test p-levels obtained by comparing the FPT distributions obtained via the simulations with the "exact" distributions obtained by means of the above mentioned numerical techniques.

The examples shown allow to pinpoint how the fully Monte Carlo algorithm proposed to approximate the tied-down FPT probabilities of the simulated processes in order to obtain fair estimates of the first passage time through a boundary leads to reliable results. A possible increase in computational time with respect to the method proposed in Giraudo, M.T. and Sacerdote, L. (1999), connected with the additional simulations of the tied-down process trajectories, should be compensated by the possibility of using larger discretization steps. However, a check done on the computational times required has shown that the method is convenient also from such viewpoint. Furher insights in this direction, as well as the determination of the order of error in the estimation of the FPT by means of this technique and of its computational effort, are the objects of our present research.

References

[1] Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*, Wiley.

[2] Giorno,V., Nobile, A.G. Ricciardi, L.M. and Sato, S. (1989). "On the evaluation of first-passage-time probability densities via non-singular integral equations", *Adv. Appl. Prob.* 21, 20-36.

[3] Giraudo, M.T. and Sacerdote, L. (1999). "An improved technique for the simulation of first passage times for diffusion processes", *Comm. Stat. Sim. 28,* no.*4*

[4] Honerkamp, J. (1994). *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis*, VCH.

[5] Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes*, Academic Press.

[6] Kloeden, P.E. and Platen, E. (1992). *The Numerical Solution of Stochastic Differential Equations*, Springer Verlag.

[7] Platen, E. and Wagner, W. (1983) On a Taylor formula for a class of Itô processes. Prob. Math. Statist. 3, no. 1, 37-51.

[8] Ricciardi, L.M. (1977). *Diffusion Processes and Related Topics in Biology*, Lectures Notes in Biomathematics, n. 14, Springer Verlag.

[9] Ricciardi, L.M., Sacerdote, L. and Sato, S. (1984). "On an integral equation for first-passage-time probability densities", *J. Appl. Prob.* 21, 302-314.

[10] Ricciardi, L.M. and Sato, S. (1990). "Diffusion processes and first-passage-time problems", in *Lectures in Applied Mathematics and Informatics*, Ricciardi, L.M. ed., Manchester University Press.

2 Workshop Program

Monday 17 January

09.00-10.00 Registration and coffee/tea

Chairman: Eva B. Vedel Jensen

10.00-10.10 **Ole E. Barndorff-Nielsen:**
Welcome.

10.10-11.00 **Ole G. Mouritsen:**
The third science — the computer experiment.

11.10-12.00 **Ole G. Mouritsen:**
The third science — the computer experiment.

12.30-14.00 Lunch

Chairman: Adrian Baddeley

14.00-14.50 **Ian Dryden:**
Stochastic deformation.

Coffee/tea

15.10-16.00 **Ian Dryden:**
Stochastic deformation.

16.10-16.40 **Olivier Perrin:**
Estimating a non-stationary spatial structure using simulated annealing.

16.45-17.15 **Günter Döge:**
Grand canonical simulations of hard-disk systems by simulated tempering.

17.15-18.30 Welcome Reception

Tuesday 18 January

Chairman: Søren Asmussen

9.00-9.50 **William Stewart:**
Numerical methods for Markov chains.

Coffee/tea

10.10-11.00 **William Stewart:**
Numerical methods for Markov chains.

11.10-12.00 **Paul Glasserman:**
Variance reduction techniques for simulating value-at-risk.

12.30-14.00 Lunch

Chairman: Ian Dryden

14.00-14.30 **Ole F. Christensen:**
Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo.

14.35-15.05 **Søren Feodor Nielsen:**
Simulated EM algorithms: A comparison.

Coffee/tea

15.30-16.00 **Laird Breyer:**
Automatic ways of coupling Markov chains.

Wednesday 19 January

Chairman: Søren Asmussen

9.00-9.50 **William Stewart:**
Numerical methods for Markov chains.

Coffee/tea

10.10-11.00 **William Stewart:**
Numerical methods for Markov chains.

11.10-12.00 **Paul Glasserman:**
Pricing American options by simulation.

12.30-16.00 Conference Lunch & Excursion

Coffee/tea

Chairman: Ute Hahn

16.30-17.00 **Niels Væver Hartvig:**
A stochastic geometry model for fMRI data.

17.10-17.40 **Fabio Spizzichino:**
Exchangeable heterogeneous populations and computation of probability distributions for vectors of "occupation numbers".

Thursday 20 January

Chairman: Dietrich Stoyan

9.00-9.50 **Adrian Baddeley:**
Conditional simulation.

Coffee/tea

10.10-11.00 **Adrian Baddeley:**
Conditional simulation.

11.10-12.00 **Anders Krogh:**
Applications of hidden Markov models in molecular biology.

12.30-14.00 LUNCH

Chairman: Anders Krogh

14.00-14.50 **Matthew Stephens:**
Computationally-intensive inference in molecular population genetics.

COFFEE/TEA

15.10-16.00 **Jotun Hein (with B. Knudsen):**
RNA secondary structure prediction using stochastic context-free grammars and evolutionary history.

16.05-16.35 **Dietrich Stoyan:**
Statistical characterisation of connectivity and permeability of porous media.

16.40-17.10 **Tomas Mrkvicka:**
Estimation variances for Poisson processes of compact sets.

Friday 21 January

Chairman: William Stewart

9.00-9.50 **Søren Asmussen:**
Matrix-analytic algorithms for many–server queues.

10.00-10.30 **Cristina Zucca (with M.T. Giraudo and L. Sacerdote):**
Evaluation of first passage times of diffusion processes through boundaries by means of a totally simulative algorithm.

3 List of participants

Søren Asmussen
Department of Mathematical Statistics
University of Lund
Box 118
S-221 00 Lund, Sweden
Email: asmus@maths.lth.se

Adrian Baddeley
Department of Mathematics
University of Western Australia
Nedlands WA 6907
Australia
Email: adrian@maths.uwa.edu.au

Ole E. Barndorff-Nielsen
MaPhySto
Department of Mathematical Sciences
University of Aarhus
8000 Aarhus C, Denmark
Email: oebn@imf.au.dk

Laird Breyer
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7E
DK-9220 Aalborg Ø, Denmark
Email: lbreyer@math.auc.dk

Ole F. Christensen
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7E
DK-9220 Aalborg Ø, Denmark
Email: olefc@math.auc.dk

Jose Luis Batun Cutz
CIMAT
Apartado Postal 402
36000-Guanajuato Gto.
Mexico
Email: batun@fractal.cimat.mx

Ian Dryden
Department of Statistics
University of Nottingham
Nottingham NG7 2RD
United Kingdom
Email: Ian.Dryden@maths.nottingham.ac.uk

Günter Döge
Inst. of Stochastics
Freiberg Univ. of Mining and Technology
Bernard-von-Cotta-Str. 2
D-09596 Freiberg, Germany
Email: doege@orion.hrz.tu-freiberg.de

Paul Glasserman
Graduate School of Business
Columbia University
New York
NY 10027, U.S.A.
Email: pg20@columbia.edu

Hans Jørgen G. Gundersen
Stereological Research Laboratory
University of Aarhus
DK-8000 Aarhus C
Denmark
Email: stereo@svfcd.aau.dk

Ute Hahn
Department of Mathematics and Statistics
The University of Western Australia
WA 6907 Nedlands
Australia
Email: uhahn@maths.uwa.edu.au

Niels Væver Hartvig
Department of Theoretical Statistics
University of Aarhus
DK-8000 Aarhus C
Denmark
Email: vaever@imf.au.dk

Jotun Hein
Department for Ecology and Genetics
University of Aarhus
DK-8000 Aarhus C
Denmark
Email: jotun@imf.au.dk

Asger Hobolth
Department of Theoretical Statistics
University of Aarhus
Ny Munkegade
8000 Aarhus C, Denmark
Email: asho@imf.au.dk

Eva B. Vedel Jensen
Department of Theoretical Statistics
University of Aarhus
DK-8000 Aarhus C
Denmark
Email: eva@imf.au.dk

Anders Krogh
Bioteknologisk Sekvensanalyse
Danmarks Tekniske Universitet
Bygn. 208
DK-2800 Lyngby, Denmark
Email: krogh@cbs.dtu.dk

Debicki Krzysztof
Mathematical Institute
University of Wroclaw
Pl. Grunwaldzki 2-4
50-384 Wroclaw, Poland
Email: debicki@math.uni.wroc.pl

Ludolf Meester
Delft University of Technology
Mekelweg 4
NL-2628 CD Delft
The Netherlands
Email: l.e.meester@its.tudelft.nl

Klaus Mosegaard
Niels Bohr Intitute for Astronomy,
Physics and Geophysics
Juliane Maries Vej 30
DK-2100 Copenhagen Ø, Denmark
Email: klaus@gfy.ku.dk

Ole Mouritsen
Center for Biomembrane Physics
Deparment of Chemistry, Building 206
Technical University of Denmark
DK-2800 Lyngby, Denmark
Email: ogm@kemi.dtu.dk

Tomas Mrkvicka
Mathematical Institute
Charles University
Sokolovska 83
CZ-186 00 Praha 8, Czech Republic
Email: tomas.mrkvicka@seznam.cz

Jesper Møller
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7E
DK-9220 Aalborg Ø, Denmark
Email: jm@math.auc.dk

Søren Feodor Nielsen
Department of Theoretical Statistics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø, Denmark
Email: feodor@stat.ku.dk

Zbigniew Palmowski
Mathematical Institute
University of Wroclaw
Grunwaldzki 2/4
50-384 Wroclaw, Poland
Email: zpalma@math.uni.wroc.pl

Anne-Mette Krabbe Pedersen
Department for Ecology and Genetics
University of Aarhus
8000 Aarhus C
Denmark
Email: annemet@pop.bio.aau.dk

Olivier Perrin
Université Toulouse 1
GREMAQ, Manufacture des Tabacs
21 Alle de Brienne
31000 Toulouse, France
Email: perrin@cict.fr

Matthew N. Sathekge
The Judge Institute of Management Studies
University of Cambridge
Trumpington Street
Cambridge CB2 1AG, UK
Email: mns22@eng.cam.ac.uk

Hanspeter Schmidli
Department of Theoretical Statistics
University of Aarhus
DK-8000 Aarhus C
Denmarks
Email: schmidli@imf.au.dk

Fabio Spizzichino
Department of Mathematics
University of Rome "La Sapienza"
Piazzale A. Moro 5
00136 Rome, Italy
Email: spizzichino@axrma.uniroma1.it

Valeria Spizzichino
Department of Chemistry
University of Rome "La Sapienza"
Via Papiniano, 46
I-00136 Rome, Italy
Email: spyzzy@tiscalinet.it

Matthew Stephens
Department of Statistics
University of Oxford
1 South Parks Road
Oxford OX1 3TG, England
Email: stephens@stats.ox.ac.uk

William Stewart
Department of Computer Science
Box 8206
North Carolina State University
Raleigh, NC 27695-8206, U.S.A.
Email: billy@csc.ncsu.edu

Dietrich Stoyan
Institut für Stochastik
TU Bergakademie Freiberg
Bernhard-von-Cotta-Str. 2
D-09596 Freiberg, Germany
Email: stoyan@hrz.tu-freiberg.de

Rasmus Waagepetersen
Department of Agricultural Systems
Research Centre Foulum
DK-8830 Tjele
Denmark
Email: rasmus.waagepetersen@agrsci.dk

Dvoralai Wulfsohn
Institut for Maskinteknik
Aalborg University
Pontoppidanstræde 101
DK-9220 Aalborg Ø, Denmark
Email: dw@ime.auc.dk

Cristina Zucca
Department of Mathematics
University of Torino
V.C. Alberto 10
I-10123 Torino, Italy
Email: zucca@dm.unito.it