MaPhySto
The Danish National Research Foundation:
Network in Mathematical Physics and Stochastics

Miscellanea
no. 25 March 2004

Søren Asmussen (ed.):

Symposium on Probability and Algorithms

ISSN 1398-5957

Symposium
on
Probability and Algorithms

Symposium on Probability and Algorithms

MaPhySto

University of Aarhus
Friday September 26, 2003

Organized by Søren Asmussen

Typesetting & Layout by Lars Madsen
This publication was typeset using $\LaTeX 2_\varepsilon$ and the Memoir class.

Contents

This publication contains the abstract from the various talk given at the symposium in the order of presentation.

Further information on the symposium can be found at the symposium homepage

`http://www.maphysto.dk/events2/ProbAlg03/`

Node Depth Distribution in Random Binary Search Trees

Rudolf Grübel
Department of Mathematical Stochastics, Hannover

We investigate the distribution of the depth of a node containing a specific key or, equivalently, the number of steps needed to retrieve an item stored in a randomly grown binary search tree. Using a representation in terms of mixed and compounded standard distributions we derive approximations by POISSON and MIXED POISSON distributions; these lead to asymptotic normality results. We are particularly interested in the influence of the key value on the distribution of the node depth. Methodologically our message is that the explicit representation may provide additional insight if compared to the standard approach that is based on the RECURSIVE STRUCTURE of the trees, and that in order to exhibit the influence of the key on the distributional asymptotics a suitable choice of distance of probability distributions is important. Our results are also applicable in connection with the number of recursions needed in Hoare's selection algorithm FIND.

Performance Modeling of Future Communication Systems

Hans-Peter Schwefel
Department of Communication Technology, Aalborg

Performance models for IP networks have turned out to be a huge challenge for a number of reasons:

(1) PACKET BASED TRANSMISSION is more difficult to model than the circuit switched case;
(2) traffic tends to be very BURSTY;
(3) the impact of ROUTING and SCHEDULING has to be taken into account;
(4) the transport layer (i.e. Transmission Control Protocol, TCP, for a large part of today's traffic) has major performance impact, and
(5) there exists a wide range of applications generating HETEROGENEOUS TRAFFIC with varying performance (or Quality of Service) requirements.

As mobile access to IP networks is becoming increasingly popular, the problem of finding adequate performance models becomes even more complex. This talk focuses on three of the challenges, namely (1),(2), and (4): Previous results from MARKOVIAN MODELS using a MATRIX-EXPONENTIAL REPRESENTATION of so-called truncated POWER-TAIL distributions provide important insights into the performance impact of extreme burstiness (long-range dependence). However, the models and methods have to be modified to account for the dynamic packet rate adjustment as provided by TCP congestion control. One consequence is that the 'linear' approach of defining a traffic model which is subsequently fed into a network model cannot be employed any more. Three different model types are discussed in this context:

CONNECTION LEVEL MODELS, FIX-POINT MODELS, and INTEGRATED MODELS. As part of the latter class, an extension of the Markovian models from earlier in the talk is discussed, and the matrix-algebraic solutions are used to obtain numerical performance results. The talk concludes with an outlook on the challenges in wireless scenarios.

An Introduction to Extractors

Peter Bro Miltersen
BRICS, Aarhus

An extractor with MIN-ENTROPY THRESHOLD h and error ϵ is a POLYNOMIAL TIME computable map $F : \{0,1\}^R \times \{0,1\}^d \to \{0,1\}^n$, $d \ll n < R$ with the property that whenever X is a random variable on $\{0,1\}^R$ of min-entropy at least h and U is an random variable, independent of X and uniformly distributed on $\{0,1\}^d$, then $F(X,U)$ is a random variable with a distribution which is ϵ-close, in variation distance, to the uniform distribution on $\{0,1\}^n$.

We give a survey of known extractors and their applications. In particular, we highlight the application of RANDOMNESS-EFFICIENT MONTE CARLO INTEGRATION: If a set $A \subseteq \{0,1\}^n$ is given as a black box, we can estimate the density of A within fixed additive error, say $\epsilon = \frac{1}{100}$ and confidence $1 - 2^{-R+1.01n}$ in polynomial time, using only R random bits, i.e., each random bit used decreases the inconfidence by a factor of two. In contrast, the obvious sampling approach would use $\theta(n)$ random bits to decrease the inconfidence by a factor of two.

Local Computation with Probabilities

Steffen L. Lauritzen
Department of Mathematical Sciences, Aalborg

The lecture describes a series of algorithms for efficient computation with probabilities in GRAPHICAL MODELS. The type of algorithms have been developed in a number of special cases, but this lecture attempts to give a unified and abstract view.

The key to efficient computation is the exploitation of a factorization of the joint probability mass function into a product of simpler terms. The structure of the factorization is formally captured in a graph and graphtheoretic manipulations sets up the basic computational structure as a JUNCTION TREE OF CLIQUES, i.e. a tree of cliques of a CHORDAL GRAPH with the additional structure that the intersection of any pair of cliques is contained in all other cliques on the tree-path between them.

Computations are then organized in terms of message passing between cliques, thus reducing the basic complexity of the computation from being exponential in the number of variables to exponential in clique size.

Variants of the algorithm yield marginal probabilities of specific configurations, full clique marginals, conditional distributions of single variables given specific configurations, most probable configurations, etc., and enables efficient sampling from the same distribution, exploitable within several MONTE CARLO procedures.

Further abstraction of the algorithms unify them with algorithms for solving SPARSE EQUATIONS, DYNAMIC PROGRAMMING, PROBABILISTIC DECODING, etc.

QUICKSELECT REVISITED

Uwe Rössler
Department of Stochastics, Kiel

The algorithm QUICKSELECT or FIND, invented 1961 by Hoare, finds the l-th smallest in a set of n members. QUICKSELECT proceeds as follows:

- Choose by random a pivot element.
- Construct the set of smaller elements than the pivot and the set of larger elements.
- Recall QUICKSELECT appropriate for the subset containing the l-th smallest.

We give a complete literature review on the analysis of the random divide-and-and-conquer algorithms QUICKSELECT. The running time is proportional to the random number of comparisons. The analysis involves the worst and best case, the average, exponential moments and finally the asymptotic distribution. The asymptotic process is the FIND process with values in the space D of right continuous functions with left limits.

Further we present a refined approach to the limiting process FIND relying on the contraction method. We construct the asymptotic finite dimensional marginals via the recursive equations only. The existence of a measure on the space D follows by general results using superexponential estimates for the supremum of FIND.

This approach allows the analysis of more general algorithms or recursive equations involving processes with values in D .

A copy of the article is available at `http://www-computerlabor.math.uni-kiel.de/stochastik/roesler/`

Analysis of Algorithms by the Contraction Method

Ludger Rüschendorf
Department of Mathematical Stochastics, Freiburg

A general limit theorem for RECURSIVE SEQUENCES of divide and conquer type is presented in this talk. Basically it says that the recursive structure plus some information on first moment(s) allow to conclude that the recursive sequence converges to some limit and also to identify the limit as solution of some fixed point equation. This limit theorem applies to a large class of DETERMINISTIC and STOCHASTIC ALGORITHMS in particular to many random tree models. The limit theorem is based on some recent joint work with Ralph Neininger and uses some extensions of the contraction method.

Omniparametric Simulation for the Two-type Richardson Model

Fredrik Lundin
Department of Mathematical Sciences, Göteborg

In the field of particle systems and growths models simulation is an important tool. When explicit calculations are too complex or impossible to perform we may use simulations instead. In this situation it is important to have a technique for doing simulations for every possible parameter value, since we can have different probabilistic behaviour at certain parameter values. When such situations occur ordinary fixed parameter simulation does not suffice. In talk I describe an omniparametric simulation algorithm, that is, a simulation algorithm generating samples for all parameter values within the same sample. The technique is used to study the problem of ASYMMETRIC SIMULTANEOUS SURVIVAL in the two-type RICHARDSON MODEL.

Statistical Inference for Locally Scaled Point Processes

Eva B. Vedel Jensen
Department of Mathematical Sciences, Aarhus

A new class of models for inhomogeneous spatial point processes in $\mathbb{R}^k$ has recently been suggested in Hahn et al. (2003). A LOCALLY SCALED POINT PROCESS has the property that subsets of the point process with different point intensity only differ by a local scaling factor. A locally scaled version of a homogeneous TEMPLATE PROCESS can be constructed by replacing d-dimensional volume measures in $\mathbb{R}^k$ by locally scaled versions.

In this lecture, statistical inference for locally scaled point processes is discussed. It is assumed that the template process is parametrized by a parameter ψ and the scaling function by a parameter θ. Simultaneous maximum likelihood estimation of ψ and θ by MCMC is computational very demanding. It is investigated whether an alternative more simple two-step estimation procedure can be used instead. This procedure is based on a decomposition of the likelihood function in a Poisson likelihood only depending on θ and a likelihood representing the density of the locally scaled process with respect to the scaled Poission process. In the two-step procedure, the parameter θ is first estimated from the Poisson likelihood and next ψ is estimated from the second component of the likelihood with the estimate of θ inserted. This procedure can be given theoretical support and works well on simulated data.

U. Hahn, E.B.V. Jensen, M.-C. van Lieshout and L.S. Nielsen (2003). *Inhomogeneous spatial point processes by location-dependent scaling*. Adv. Appl. Prob. **35**, 319–336.

Threshold Phenomena, MCMC Convergence, and Related Monotonicity Issues in Probability Theory

Olle Häggström
Department of Mathematical Sciences, Göteborg

Many intriguing problems in probability theory are of the following kind. Consider a stochastic system with a parameter p that governs its dynamics, and some particular aspect A of the dynamics. For instance, p could be a drift parameter of some random motion, and A is the resulting asymptotic speed. Does A depend monotonically on p? Such questions are particularly natural to ask in systems exhibiting critical phenomena or phase transitions.

This talk will be a survey of such monotonicity issues. Specific models that will be discussed include PERCOLATION, GIBBS SYSTEMS such as the ISING AND WIDOM-ROWLINSON MODELS, MONTE CARLO MARKOV CHAINS, and RANDOM WALKS. As to the answers to the main question – whether A depend monotonically on p – we will encounter all four of the following answers, the last of which can be quite a surprise:

- Yes, monotonicity holds, and it admits a short and easy proof.
- Yes, it does hold, but the proof is not straightforward.
- The question is open.
- No, the desired monotonicity is demonstrably false.

Erlang Extrapolation for Ruin Probabilities

Søren Asmussen
Department of Mathematical Sciences, Aarhus

The classical CRAMÉR-LUNDBERG MODEL for insurance risk assumes claims to arrive according to a Poisson process, claim sizes to be i.i.d. with common distribution say B and premium to be paid at a coninuous rate, w.l.o.g. 1. The initial reserve is denoted by u, the ruin time $\tau(u)$ is the first time the reserve goes negative, and the probability of ruin with an infinite horizon or a finite horizon T are $\psi(u) = P(\tau(u) < \infty)$, resp. $\psi(u, T) = P(\tau(u) \leq T)$.

It has been known for long that the calculation of $\psi(u)$ is simple if B is phase-type (and basically only then). This means that B is the distribution of the absorption time in a finite MARKOV PROCESS, for example a mixture of exponentials, an ERLANG distribution (a gamma with integer parameter) or more generally any parallel/series/loop combination of exponential distributions. In fact, the ruin probability is then $\psi(u) = \boldsymbol{\alpha} \exp\{\boldsymbol{Q}u\}\boldsymbol{q}$ where $\boldsymbol{Q}$ is a matrix, $\boldsymbol{\alpha}$ a row vector and $\mathbf{q}$ a column vector, all of explicit form. No such simple formulas are known for the FINITE HORIZON RUIN PROBABILITY $\psi(u, T)$.

The idea of the work presented is to show that if H_k is an ERLANG r.v., then $E\psi(u, H_k)$ has the form $\boldsymbol{\alpha}_k \exp\{\boldsymbol{Q}_k u\}\boldsymbol{q_k}$ with a simple recursion for the parameters (the case $k = 1$ reduces to a known formula for the LAPLACE TRANSFORM of $\psi(u, T)$ w.r.t. T). Scaling H_k to have mean T and letting $k \to \infty$ yields the approximation $\psi(u) \approx E\psi(u, H_k)$ for k large enough; a numerical study indicates that $k = 6$–8 is usually large enough, providing a fast and easily implemented algorithm to compute $\psi(u, T)$.

Noting that $\psi(u) - E\psi(u, H_k)$ is of order $1/k$ by the central limit theorem allows to use RICHARDSON EXTRAPOLATION, improving the precision of the algorithm so much that in most cases 2 correct digits were obtained by taking just $k = 2$ or 3.

S. Asmussen, F. Avram and M. Usabel (2002).
Erlangian approximations for ruin probabilities. Astin Bull. **32**, 267–281.