A Generalized Urn Model With Multiple Drawing And Random Addition (I)

Rafik Aguech and Olfa Selmi

In this paper we consider a non balanced generalized Polya urn model. At each step, m balls are drawn at random from an urn containing white and black balls. Balls are added according the inspected colors. We consider two different models: the self reinforcement model and the opposite reinforcement model. We use the Stochastic approximation to obtain the almost sure convergence of the proportion of white balls in the urn, then we give the asymptotic of the expectation and the variance. Moreover we obtain a central limit theorem for the number of white balls in the case of opposite reinforcement urn model.

Estimating the Small Data in Big Data-Concentration and Regular Variation

Anna Ben-Hamou, Stéphane Boucheron and Mesrob I. Ohannessian

XXX
From the 20th century to the 21st, various disciplines have tried to infer something about scarcely observed events: zoologists about species, cryptologists about cyphers, linguists about vocabularies, and data scientists about almost everything. These problems are all about 'small data' within possibly much bigger data. Can we make such inference?

On σ-palindromes and their applications

Srečko Brlek, Sébastien Labbé, Nadia Lafrenière and Xavier Provençal

We study the structure of pseudo-palindromes on finite alphabets, i.e. words w satisfying $w = \sigma(\tilde{w})$ for some involution σ on the alphabet. We provide algorithms for the computation of σ-lacunas in w, that is the positions where the longest σ-palindromic suffix is not uni-occurrent. We also consider the problem of reconstructing words from a given fixed set of σ-palindromes. Finally we describe an algorithm for the generation of tiles that tile the plane by translation.

The Maximum Degree of a Random Delaunay Triangulation in a Smooth Convex

Nicolas Broutin, Olivier Devillers and Ross Hemsley

We give a new polylogarithmic bound on the maximum degree of a random Delaunay triangulation in a smooth convex, that holds with probability one as the number of points goes to infinity. In particular, our new bound holds even for points arbitrarily close to the boundary of the domain.

Connected System of Urns

Jacek Cichoń, RafałKapelko and Marek Klonowski

XXX

Fragmented Random Structures
Ed Coffman, Robert Margolies, Peter Winkler and Gil Zussman

Computers with hard-disk storage and networks with dynamic spectrum access illustrate resources that allow fragmented items – in these examples, items are files and spectra, respectively. In our discrete model of such systems, the resource is a sequence of slots. There is a queue of items awaiting allocations of the resource; the queue is served in FIFO order. Specified for each item are the number of slots needed by the item, and for what period of time. Under the key assumption that an item's allocation can not be changed prior to its departure, fragmentation in the form of alternating gaps and allocated resource builds up as items come and go, regardless of the allocation algorithm adopted. The improvements in resource utilization created by fragmentation are countered by the added cost of manipulating fragmented items, so how fragmentation evolves is an important performance issue. Within a baseline probability model of the system operating at capacity, we prove that, in the stationary limit under any practical algorithm, almost all items are completely fragmented, i.e., items of size i are fragmented into i disjoint slots. In the full paper, this result is balanced by many experimental results which show that, in reality, the times to approach steady-state fragmentation are typically exceptionally long, and hence that even nearly complete fragmentation is often of no concern.

Random Sampling of Ordered Trees According to the Number of Occurrences of a Pattern
Gwendal Collet, Julien David and Alice Jacquot

In this paper we describe an algorithm to automatically build a random generator for ordered trees according to their size and the number of occurrences of a given pattern. This work fits into the general framework of automatic building of random samplers that follow a probability distribution in which one can guarantee a set of observables (exactly or in average) on the generated objects, that we sketch in this paper. We also give a linear pattern matching algorithm which is an adaptation of the localization automaton on trees.

A generator of random convex polygons in a disc
Olivier Devillers, Philippe Duchon and Rémy Thomasse

Using the theory of random polytopes, we propose an algorithm that generates a random polygon in $\mathbb{R}^2$ as the convex hull of n points in a disc, without explicitly generating n points. As the size complexity is only $O(n^{1/3})$ and the time complexity only $n1/3\log^2 n$,this algorithm allows to generate random polygons with very large n.

A chaotic random convex hull
Olivier Devillers, Marc Glisse and Rémy Thomasse

The asymptotic behavior of the expected size of the convex hull of uniformly random points in a convex body in $\mathbb{R}^d$ is polynomial for a smooth body and polylogarithmic for a polytope. We construct a body whose expected size of the convex hull oscillates between these two behaviors when the number of points increases.

The Multiplicity Matching Parameter in Suffix Trees Built over an Arbitrary Finite Alphabet
Jeffrey Gaither and Mark Daniel Ward

The multiplicity matching parameter (or MMP) Mn of a string X gives the number of suffixes among the first n suffixes of X which agree maximally with X's $(n+1)$st suffix. The quantity M_n has a natural application to suffix trees, where it represents the number of suffixes among the first n that are rooted at the parent of the $(n+1)$st; and also to the Lempel-Ziv '77 data compression algorithm, where it gives the number of pointers into the relevant database at the $(n+1)$st transmission. In this work we derive the asymptotic distribution of the MMP in the case where the alphabet over which X is built is of arbitrary size. This extends the main result of the second author's thesis, which assumed a binary alphabet.

On the Asymptotic Probability of Forbidden Motifs on the Fringe of Recursive Trees
Mohan Gopaladesikan and Mark Daniel Ward

We use analytic methods to study the asymptotic probability of a family of motifs not occurring on the fringe of a random recursive tree. This probability is intimately related to the WhittakerM function. We study this asymptotic probability in two regimes, which lay the foundation for a larger multivariate asymptotic analysis in the future.

A new upper bound for Policy Iteration
Romain Hollanders, Balázs Gerencsér, Jean-Charles Delvenne and Raphaël M. Jungers

Solving Markov Decision Processes (MDPs) is a recurrent task in engineering. Even though it is known that solutions for minimizing the infinite horizon expected cost can be found in polynomial time using LP techniques, iterative methods like the Policy Iteration algorithm (PI) remain usually the most efficient in practice. This method is guaran- teed to converge in a finite number of steps. Unfortunately, it is known that it may require an exponential number of steps in the size of the problem to converge. On the other hand, many unknowns remain considering the actual worst case complexity. In this work, we provide the first improvement over the fifteen years old upper bound from Mansour and Singh (1999) by showing that PI requires at most $2 \cdot \frac{2^n}{n}$ iterations to converge.

Information in Strings: Enumeration of Eulerian Paths and Eulerian Components in Markov Sequences
Philippe Jacquet and Dimitrios Milioris

In this paper, we evaluate the number of Eulerian circuits that can be obtained by an arbitrary rotation in a Markovian string, i.e., corresponding to a given Markovian type. Since all rotations do not result in an Eulerian circuit, but several of them, called Eulerian components; we also investigate the number of Eulerian components that result from a random rotation in a Markovian string. We consider the asymptotic behaviour of those quantities when the size of the string n tends to infinity. In particular we show that the average number of components tends to be in $\log V$, where V is the size of a large alphabet, in the uniform case.

Random 312-Avoiding Permutations
Neal Madras and Lerna Pehlivan

Monte Carlo experiments in a recent paper of Atapour and Madras reveal some features of random 312-avoiding permutations when these random permutations are graphed as functions from $1, \ldots, N$ to $1, \ldots, N$. In light of these experiments we determine some probabilities explicitly under the uniform distribution on the set of 312-avoiding permutations of $1, \ldots, N$. We derive exact formulas for the probability that the graph of a random 312-avoiding permutation has one specified point below the diagonal as well as two specified points below the diagonal considering both increasing and decreasing cases. In all of these cases we find the asymptotic approximations to these probabilities for large N and with the restriction that the determined points apart from each other and far from the boundaries. We also generalized the two specified decreasing points to k specified decreasing points and we observe that that for large N the points below the diagonal look like trajectories of a random walk.

Steady-State Analysis of Reaction Networks
Vincent Picard, Anne Siegel and Jérémie Bourdon

The quantitative analysis of reaction networks is of major concern in systems biology. The Gillespie algorithm is a kinetic Monte-Carlo method that generates probabilistic trajectories for these networks and provides a dynamical understanding at the single-cell level. However the algorithm requires many parameters and can be computationally too expensive. Inspired by flux balance analysis approaches, we study the Gillespie trajectories when the system fulfills steady-state conditions. The resulting stochastic process is a random walk that admits a central limit theorem. We present some practical applications of this result such as model validation. The result can be extended to stochastic Petri nets or weighted Markov chains and emphasizes the relationship between average-case analysis of algorithms methods and concrete biological questions.

Equilibrium Points of an AND-OR Tree: under Constraints on Probability
Toshio Suzuki and Yoshinao Niida

We study a probability distribution d on the truth assignments to a uniform binary AND-OR tree. Liu and Tanaka [2007, Inform. Process. Lett.] showed the following: If d achieves the equilibrium among independent distributions (ID) then d is an independent identical distribution (IID). We show a stronger form of the above result. Given a real number r such that $0 < r < 1$, we consider a constraint that the probability of the root node having the value 0 is r. Our main result is the following: When we restrict ourselves to IDs satisfying this constraint, the above result of Liu and Tanaka still holds. The proof employs clever tricks of induction. In particular, we show two fundamental relationships between expected cost and probability in an IID on an OR-AND tree: (1) The ratio of the cost to the probability (of the root having the value 0) is a decreasing function of the probability x of the leaf. (2) The ratio of derivative of the cost to the derivative of the probability is decreasing function of x, too.

Quasi-Boltzmann Sampling for Iterated Substitutive Classes with Applications to Balanced Trees

Qiuyi Zhangi

For many combinatorial classes specifiable by symbolic operators, Boltzmann samplers have been designed as efficient algorithms to generate random objects in those classes. However, iterated substitutive combinatorial classes, which are defined via the use of the substitution operator of the form $\mathcal{C} = \mathcal{P} + \mathcal{C} \circ \mathcal{Q}$, has evaded similar analysis. To sample random objects in these classes, we propose a quasi-Boltzmann sampler, which samples according to a distribution that is almost exponential in size. We produce some average-case analysis of our algorithm and demonstrate the utility of our analysis in applications to the study of B-trees.