Summary of Research

by

Antar Bandyopadhyay

Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, New Delhi Centre
7 S. J. S. Sansanwal Marg
New Delhi 110016, INDIA

Phone : +91 (0)99990 09125
Fax : +91 (0)11 4149 3981
antar@isid.ac.in
http://www.isid.ac.in/∼antar

1 General Overview

My main research interest is in theoretical and applied probability mainly focusing on discrete problems arising out of combinatorics and statistical physics. In particular, I am interested in random graphs, probability on tress, combinatorial optimizations, recursive distributional equations, branching random walks, percolation theory, interacting particle systems, Markov chains and their mixing behavior, random walks in random environment and the urn models. I also keep strong interest in theoretical statistics and application of statistics. The next section describes my contributions so far in chronologically backward sequence. The Sections 3 and 4 list the ongoing work and future plan respectively.

2 Research Contributions till Date

2.1 Negatively Reinforced Urn Models

Jointly with my former Ph.D. student Gursharn Kaur [9], in this work we have introduced a new class of urn schemes, which we refer as negatively reinforced urns. For negative but linear reinforcements [9], we have derived the exact almost sure behavior of the configuration of the urn at the strong law of the large number scaling, Detailed central limit theorem type results for the fluctuations around the almost sure limit have also been derived [9]. The results have been also extended for the associated *color count statistics* in the same model.

2.2 Infinite Color Urn Models and Representation Theorem

Jointly with my another former Ph.D. student Debleena Thacker [18], we have derived *Marginal* and *Grand Representation Theorems* for the classical and also for the *infinite color* generalized Pólya urn schemes. These theorems essentially establish that any balanced urn can be viewed as a *branching Markov chain* on a random tree, namely, the *random recursive tree*, also known as, *uniform random tree.* These powerful representation theorems can provide quick proofs of most of the classical results, as well as, provides generalization to infinite color schemes, bridging the two most well studied urn schemes, namely the finite color generalized Pólya urn scheme and the Blackwell-MacQueen scheme. It is expected that these theorems will have significant impact on the study of urn models and their applications. In recent days there has been several research papers by various authors around the world studying more details of the model introduced by us [20, 22]. Earlier in two papers [17, 16], we proved Gaussian limit for similar models where replacement matrix is the transition matrix of bounded increment random walks on d-dimensional integer lattice. There we also derived finer asymptotic of the type of *local limit theorems* [17], *Berry-Essen bounds* and *large deviation bounds* [16].

2.3 Airspace Safety Monitoring Data Analysis

Starting from the year 2011, in this statistical modeling and data analysis project jointly with Deepayan Sarkar, and officials of the *Airports Authority of India (AAI)* we are conducting the analysis of the *Traffic Sample Data (TSD)* routinely collected by the *Flight Information Regions (FIRs)* for the month of December in every calender year. December is considered to be a high-traffic month for civilian air traffic across the world. The goal of the project is to analyze this complex and large data set using statistical techniques to estimate and report about the safety of the civilian airspace in and around the oceanic region of India, which includes the Bay of Bengal, the Indian Ocean and the Arabian Sea. The analysis is conducted in the context of lateral and longitudinal collision risks. The analysis of the TSD collected by the three Indian FIRs, namely, Kolkata, Chennai and Mumbai, along with the foreign FIRs such as, Kuala Lumpur, Rangoon, Colombo, Male and Karachi are used to certify the International Target Level Safety (TLS) is met, which is at most 5×10^{-9}-number of fatal accidents in the region per flight hours. The reports are presented each year to the *Regional Airspace Safety Monitoring Advisory Group (RASMAG)* and placed for formal acceptance by the*International Civil Aviation Organization (ICAO)*, the UN body entrusted with civil aviation and safety. Our data analytic methodology has been well accepted by the civil aviation community and as a recognition India was entrusted to have its own *En-route Monitoring Agency (EMA)* under AAI, which was created in 2012 and is named BOBASMA and is now located at the Chennai FIR. The work has been hailed extraordinary and our newly developed velocity based methodology has now been accepted and adopted by several other international air safety monitoring agencies, including Japan, Singapore Thailand and China. There has been newspaper reports on the achievements of the work in several national and regional dailies.

The analysis of the latest December 2018 data is now complete and will be presented in the upcoming RASMAG/24 meeting at Bangkok in July this year.

2.4 Variance Estimation for Tree Order Restricted Models

In this work jointly with Sanjay Chaudhuri [6], we discuss estimation of the common variance of several normal populations with tree order restricted means. We discuss the asymptotic properties of the maximum likelihood estimator of the variance as the number of populations tends to infinity. We consider several cases of various orders of the sample sizes and show that the maximum likelihood estimator of the variance may or may not be consistent or be asymptotically normal.

2.5 Random Geometric Graphs on Fractals

Jointly with Farkhondeh Sajadi [11], we consider the question of connectivity for *random geometric graphs*, where there is no density for underlying distribution of the vertices. In fact, we consider the other extreme, where the vertices are distributed on a fractal set, namely, on the Cantor set in one dimension. Let n i.i.d. *Cantor* distributed points on $[0, 1]$. We show that [11] for this random geometric graph, the connectivity threshold R_n, converges almost surely to a constant $1 - 2\phi$ where $0 < \phi < 1/2$, which for standard Cantor distribution is $1/3$. We also show that $\|R_n - (1 - 2\phi)\|_1 \sim 2\, C\,(\phi)\; n^{-1/d_\phi}$ where $C\,(\phi) > 0$ is a constant and $d_\phi := -\log 2 / \log \phi$ is the *Hausdorff dimension* of the generalized Cantor set with parameter ϕ.

2.6 Discrete Time Processes on Random Structures

One of my main interest is to study discrete time processes on random structures. In a recent work jointly with my former Ph.D. student Farkhondeh Sajadi [13], we looked into a simple SIR type virus spread model on a general network and found asymptotic for expected number of infected vertices without making any specific assumptions on the underlying network structure. On a different type of problem Sajadi and myself [12] worked on a greedy approximation algorithm for the random Traveling Salesman Problem in the *mean-field* setup, that is when the underlying network has *independent and identically distributed (i.i.d.)* lengths. We showed that the greedy algorithm of moving the "nearest not visited so far", performs fairly well, in the sense that it is at most a constant times $\log n$ away from the optimal, where n is the number of cities visited.

2.7 One Dimensional Learning from Neighbor Chain

In this work jointly with with Rahul Roy and Anish Sarkar [10], we consider a model of a discrete time "interacting particle system" on the integer line where infinitely many changes are allowed at each instance of time. We describe the model using chameleons of two different colors, *viz*., red (R) and blue (B). At each instance of time each chameleon performs an independent but identical coin toss experiment with probability α to decide whether to change its color or not. If the coin lands head then the creature retains its color (this is to be interpreted as a "success"), otherwise it observes the colors and coin tosses of its two nearest neighbors and changes its color only if, among its neighbors and including itself, the proportion of successes of the other color is larger than the proportion of successes of its own color. This produces a Markov chain with infinite state space $\{R, B\}^{\mathbb{Z}}$. This model was studied by Chatterjee and Xu in 2004 in the context of diffusion of technologies in a set-up of myopic, memoryless agents. In their work they assume different success probabilities of coin tosses according to the color of the chameleon. In this work we consider the symmetric case where the success probability, α, is the same irrespective of the color of the chameleon. We show that starting from any initial translation invariant distribution of colors the Markov chain converges to a limit of a single color, i.e., even at the symmetric case there is no "coexistence" of the two colors at the limit. As a corollary we also characterize the set of all translation invariant stationary laws of this Markov chain. Moreover we show that starting with an i.i.d. color distribution with density $p \in [0, 1]$ of one color (say red), the limiting distribution is all red with probability $\pi(\alpha, p)$ which is continuous in p and for p "small" $\pi(p) > p$. The last result can be interpreted as the model favors the "underdog".

2.8 Cluster Size Distribution for Percolation on General Cayley Graphs

In this work jointly with Jeffrey Steif and Adám Timár [15] we consider i.i.d. bond percolation model on general Cayley graphs and show that the probability (at any parameter value $p \in [0, 1]$) of the cluster of the origin has exact size n, decays at a well-defined exponential rate, which may be zero! For general graphs, we relate this rate being strictly positive in the supercritical regime with the amenability/nonamenability of the underlying graphs.

2.9 Random Walks in Random Environments (RWRE)

In an earlier work jointly with Ofer Zeitouni [19], we proved a long standing conjecture about (quenched) invariance principle for *dynamic* RWRE on integer lattice with high dimension. Recently with two of my colleagues, Siva Athreya and Amites Dasgupta [2] we considered a *static* RWRE model on Cayley trees where the environment is independent and identically distributed, in the sense that it is invariant under the action of an appropriate finitely generated group. We showed that such walks are always transient and in fact often have positive speeds.

2.10 PAS for Counting Problems

In two papers jointly with David Gamarnik [7, 8], we proved that for counting the number of independent sets and also for number of proper q-colorings *polynomial time approximation schemes (PAS)* exits when the largest degree of the underlying graph is bounded and the girth of the graph is large. These works also provided asymptotic for such combinatorial quantities, which then was used to derive the *free energy* of the associated *hard-core*-type statistical physics models. The works are also considered non-trivial examples of the use of *Aldous-Steele objective method*.

2.11 Recursive Distributional Equations

Jointly with David J. Aldous in [1] and subsequently independently in [3, 4], we developed a general theory for studying distributional fixed point equations, which arise in the context of Combinatorial Optimization problems. We derived sufficient conditions under which a solution can be represented by a measurable map on a suitable probability space. The work [1] is deeply related to *Aldous-Steele objective method* and a non-rigorous argument of statistical physics known as *replica symmetry cavity method*. It has attracted a very large number of applications from both theoretical and applied point of view. It is one of the heavily cited papers.

2.12 How to Combine Fast Heuristic Markov Chain Monte Carlo with Slow Exact Sampling

In this work jointly with David J. Aldous [5], we show that Given a probability law π on a set S and a function $g : S \to R$, suppose one wants to estimate the mean $\bar{g} = \int g d\pi$. The Markov Chain Monte Carlo method consists of inventing and simulating a Markov chain with stationary distribution π. Typically one has no a priori bounds on the chain's mixing time, so even if simulations suggest rapid mixing one cannot infer rigorous confidence intervals for $\bar{g}$. But suppose there is also a separate method which (slowly) gives samples exactly from π. Using n exact samples, one could immediately get a confidence interval of length $O(n^{-1/2})$. But one can do better. Use each exact sample as the initial state of a Markov chain, and run each of these n chains for m steps. We show how to construct confidence intervals which are always valid, and which, if the (unknown) relaxation time of the chain is sufficiently small relative to m/n, have length $O(\log n/n)$ with high probability.

3 On Going Research Projects

3.1 Strong Convergence of Infinite Color Balanced Urns Under Uniform Ergodicity

In this work jointly with Svante Jonson and Debleena Thacker, we consider the generalization of the Pólya urn scheme with possibly infinite many colors as introduced in [16, 17, 18]. For countable many colors, we prove almost sure convergence of the urn configuration under *uniform ergodicity* assumption on the associated Markov chain. The proof uses a stochastic coupling of the sequence of chosen colors with a *branching Markov chain* on a weighted *random recursive tree* as described in [17, 20]. Using this coupling we estimate the covariance between any two selected colors. In particular, we reprove the limit theorem for the classical urn models with finitely many colors.

3.2 SLLN and CLT for Random Walks in I.I.D. Random Environment on Non-Abelian Free Groups

In this work jointly with Siva Athreya, Amites Dasgupta and Neeraja Sahasrabudhe, we consider the random walk on an *independent and identically distributed (i.i.d.)* random environment on a Cayley graph of a finitely generated non-abelian free group. The model was earlier introduced by Athreya, Bandyopadhyay and Dasgupta [2], who showed that such a walk is almost surely transient. In this work we use *regeneration time* techniques to prove an annealed SLLN for the walk showing that the walk has a almost surely constant and positive speed. We further show that the walk admits an annealed CLT under usual diffusive scaling.

3.3 Negative Reinforcement via Choice of Two

Using the ideas developed in [9], Kaur has extended the results for more general type of negative reinforcements [21]. Her techniques used *stochastic approximation theory* [21]. As a follow up work jointly with Kaur, we consider a negatively reinforced Pólya urn model with two choices, where two distinct colours are selected uniformly at random from the set of k colours and then a ball is added according to a decreasing weight function. This model can be seen as a simplified version of the model where the colour with the least proportion is reinforced into the urn. We also give a comparison of this model to the random reinforcement model and negatively reinforced with more than 2 choices. In particular, we show a two choice paradigm for negatively reinforced Pólya urn models. This in some sense establish rigorously a phenomenon known as "*power of two choices*" in computer science literature. The work also provides a practical solution for implementation of negatively reinforced urn models.

3.4 De-preferential Random Graphs Models

Jointly with Subhabrata Sen [14], we introduce a growing sequence of random graphs with negative reinforcements, where vertices with higher degrees are less preferred by the future vertices. This is useful in modeling resource constrained networks, for example, the food chain of an ecosystem. We derive asymptotic for the degree of a fixed vertex, as well as, the asymptotic degree distributions. The results shows drastically different phenomenon in comparison to more well studied *preferential attachment random graph models.*

3.5 A New Approach to Branching Random Walk through Smoothing Transformations

In this on going work jointly with my curent Ph.D. student Partha Pratim Ghosh, we are exploring the possibility of deriving asymptotics of the right-most position of a *branching random walk (BRW)*, using a coupling with a more well studied process derived out of a statistical technique called smoothing transformation. We show that under fairly general conditions on the underlying point process asymptotics of the right-most position can be derived for a "modified" version of the BRW. We think, that this method may help in deriving the asymptotics of the original BRW.

4 Future Research Projects

4.1 Introduction

In recent days various *urn schemes* and their many generalizations have been a key element of study for random processes with reinforcements. Starting from the seminal work by Pólya [23], various types of urn schemes with finitely many colors have been widely studied in literature. In recent years, there have been a new development for *balanced* urn schemes, namely, infinite color generalizations [17, 16, 18]. In [18] Bandyopadhyay and Thacker have considered any balanced urn schemes as a *measure valued process* on the set of colors S and thus generalizing urn schemes beyond only finitely many colors. In that case, the replacement mechanism is given by a Markov kernel, say R on S. If S is finite, it can be viewed as a *replacement matrix*, like in the classical Pólya scheme or its many finite color generalizations. It is shown that any such urn process can be represented as a *branching Markov chain* on the *random recursive tree.* Many asymptotic results can then be derived using this representation. In view of the above developments, the we wish to propose the following general project, which we term as *Interacting Urn Schemes.*

4.2 Model

Suppose $G := (V, E)$ be a *(undirected) graph*, where V is the set of vertices and E denotes the set of all edges. We will say G is finite or infinite depending on whether V is finite or infinite. If G is infinite, then we further assume that G is *locally finite*, that is, $d_v := \#\left\{u \in V \,\middle|\, \{u, v\} \in E\right\} < \infty$ for all $v \in V$. d_v is referred as the *degree* of vertex v in G. For two vertices $u, v \in V$, we will write $u \sim v$ iff $\{u, v\} \in E$.

Let S denotes the set of colors and R be a Markov kernel on S. Naturally, if S is finite then R can be presented as a matrix $\left(\left(R\left(s, s'\right)\right)\right)_{s,s' \in S}$ listing the probability mass functions $R\left(s, \cdot\right)$, as the row vectors corresponding to each color, say, $s \in S$.

Now, for each $v \in V$, let $U^{(v)} \equiv \left(U_n^{(v)}\right)_{n \geq 0}$ denote the *urn process* at the vertex $v \in V$, starting with the initial configurations $\left(U_0^{(v)}\right)_{v \in V}$ and with replacement mechanism R. The processes is defined as follows:

$$U_{n+1}^{(v)}(ds) = U_{n+1}^{(v)}(ds) + \frac{1}{d_v} \sum_{u \sim v} R\left(Z_n^{(u)}, ds\right), \tag{1}$$

where "$Z_n^{(v)}$ *is the* $(n+1)$*-th color of the random ball drawn at time* $n+1$ *at the vertex* v", that is,

$$\mathbf{P}\left(Z_n^{(v)} \in ds_v \text{ for } v \in V' \,\middle|\, \mathcal{F}_n\right) = \prod_{v \in V'} \frac{U_n^{(v)}(ds_v)}{n+1}, \tag{2}$$

for any finite subset V' of V, where $\mathcal{F}_n := \sigma\left(\left(U_0^{(v)}, U_1^{(v)}, \cdots, U_n^{(v)}\right)_{v \in V}\right)$.

It is worth to note here that the graph G essentially defines the *interactions* between these urns and the equation (1), says that interactions are "*local*", in the sense, that only *neighboring* urns can influence each others. Also, from the equation (2), we note that at each time n and given all the configurations of the urns $\left(U_n^{(v)}\right)_{v \in V}$, the balls are drawn independently from their respective urns. It is also important to note that, for any G (be it finite or otherwise) at any time point n, all the urns are going to change *simultaneously*. Thus the entire interacting urn process as defined above is necessarily a discrete time process, which need not have any continuous time analog.

4.3 Motivation

Our main motivation is to get a non-trivial and rigorous example of a *self-organized criticality*. A classical example of a self-organized criticality is from *statistical physics*, where they consider the so called, *Abelian sandpile model*. Unfortunately, any rigorous treatment of the same is very challenging whenever the dimension is two or more. And there are very few rigorous results till date. The model we propose here seems to be much more "tractable because of the *Grand Representation Theorem* derived in [18]. Our main goal will be to derive asymptotic limit of $\frac{U_n^{(v)}}{n+1}$ jointly for all $v \in V$, under appropriate notion of convergence.

4.4 Three Specific Models

We will mainly be interested in studying the following three specific models. In all the three cases the limiting distributions may be considered as non-trivial examples of *self-organized criticality*.

4.4.1 Interacting Pólya Urns on Integer and Bethe Lattice

Suppose R is the *identity kernel*, that is, $R(s, dt) = \delta_s(dt)$, where δ_s denotes the *Dirac Delta Function* at $s \in S$. For S finite we then get the Pólya replacement matrix, namely, the identity matrix. We will further consider G as the d-dimensional Euclidean lattice, $\mathbb{Z}^d$ with $d \geq 1$, or the d-array infinite regular tree, $\mathbb{T}_d$ with $d \geq 3$, also known as the *Bethe Lattice*. We conjecture that in both cases, the process $\left(\left(U_n^{(v)}\right)_{v \in V}\right)_{n \geq 0}$ has an *almost sure* limit, say, $(X_v)_{v \in V}$ where the marginally $X_v \sim$ *Dirichlet process* with parameter determined by the initial configurations $U_0^{(v)}$. Further we conjecture that the joint limiting distributions should be correlated in a "translation invariant" manner (under the action of the the appropriate group which keeps the underlying graph G invariant). The limit should also have some "*local averaging*" property similar to that displayed in the equation (1) above. Needless to say that we expect that the limiting distributions for $\mathbb{Z}^d$ and $\mathbb{T}_d$ to differ significantly.

4.4.2 Interacting Random Walk Urns on Integer and Bethe Lattice

Motivated by the work [21], we would further like to study the process for R which is the kernel for a *random walk* in $\mathbb{R}^m$ for some fixed $m \geq 1$. Like in the previous case, we will like to investigate the problem for both $G = \mathbb{Z}^d$ for some $d \geq 1$ and also for $G = \mathbb{T}_d$ for $d \geq 3$. We conjecture that the limiting distribution will be a Gaussian process on G with correlation structure determined by the structure of G.

4.4.3 Mean Field Interacting Pólya Urns

Now suppose R is as in Subsection 4.4.1 but $G \equiv G_N := K_N$ be the complete graph on N-vertices. In some sense, here we deliberately making interactions "*global*", by letting all the urns to "influence each other" at every time step. For this case, we again believe that there should be a limiting distribution. However, unlike in the previous two cases, for this limit we conjecture that for "large" N the limit will be "highly correlated" Dirichlet processes.

References

[1] David J. Aldous and Antar Bandyopadhyay. A survey of max-type recursive distributional equations. *Ann. Appl. Probab.*, 15(2):1047–1110, 2005.

[2] Siva Athreya, Antar Bandyopadhyay, and Amites Dasgupta. Random walks in i.i.d. random environment on Cayley trees. *Statist. Probab. Lett.*, 92:39–44, 2014.

[3] Antar Bandyopadhyay. A necessary and sufficient condition for the tail-triviality of a recursive tree process. *Sankhyā*, 68(1):1–23, 2006.

[4] Antar Bandyopadhyay. Endogeny for the logistic recursive distributional equation. *Z. Anal. Anwend.*, 30(2):237–251, 2011.

[5] Antar Bandyopadhyay and David J. Aldous. How to combine fast heuristic Markov chain Monte Carlo with slow exact sampling. *Electron. Comm. Probab.*, 6:79–89 (electronic), 2001.

[6] Antar Bandyopadhyay and Sanjay Chaudhuri. Variance estimation for tree-order restricted models. *Statistics*, 48(5):1122–1137, 2014.

[7] Antar Bandyopadhyay and David Gamarnik. Counting without sampling. New algorithms for enumeration problems using statistical physics. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 890–899. ACM, New York, 2006.

[8] Antar Bandyopadhyay and David Gamarnik. Counting without sampling: asymptotics of the log-partition function for certain statistical physics models. *Random Structures Algorithms*, 33(4):452–479, 2008.

[9] Antar Bandyopadhyay and Gursharn Kaur. Linear de-preferential urn models. *Adv. in Appl. Probab.*, 50(4):1176–1192, 2018.

[10] Antar Bandyopadhyay, Rahul Roy, and Anish Sarkar. On the one dimensional "learning from neighbours" model. *Electron. J. Probab.*, 15:no. 51, 1574–1593, 2010.

[11] Antar Bandyopadhyay and Farkhondeh Sajadi. The connectivity threshold of random geometric graphs with Cantor distributed vertices. *Statist. Probab. Lett.*, 82(12):2103–2107, 2012.

[12] Antar Bandyopadhyay and Farkhondeh Sajadi. On the nearest-neighbor algorithm for the mean-field traveling salesman problem. *J. Appl. Probab.*, 51(1):106–117, 2014.

[13] Antar Bandyopadhyay and Farkhondeh Sajadi. On the expected total number of infections for virus spread on a finite network. *Ann. Appl. Probab.*, 25(2):663–674, 2015.

[14] Antar Bandyopadhyay and Subhabrata Sen. De-Preferential Attachment Random Graphs. (submitted), 2016.

[15] Antar Bandyopadhyay, Jeffrey Steif, and Ádám Timár. On the cluster size distribution for percolation on some general graphs. *Rev. Mat. Iberoam.*, 26(2):529–550, 2010.

[16] Antar Bandyopadhyay and Debleena Thacker. Rate of convergence and large deviation for the infinite color Pólya urn schemes. *Statist. Probab. Lett.*, 92:232–240, 2014.

[17] Antar Bandyopadhyay and Debleena Thacker. Pólya urn schemes with infinitely many colors. *Bernoulli*, 23(4B):3243–3267, 2017.

[18] Antar Bandyopadhyay and Debleena Thacker. A new approach to pólya urn schemes and its infinite color generalization. (cimmunicated) available at https://arxiv.org/pdf/1606.05317, 2019.

[19] Antar Bandyopadhyay and Ofer Zeitouni. Random walk in dynamic Markovian random environment. *ALEA Lat. Am. J. Probab. Math. Stat.*, 1:205–224, 2006.

[20] Svante Janson. A.s. convergence for infinite colour Pólya urns associated with random walks. (https://arxiv.org/abs/1803.04207), 2018.

[21] Gursharn Kaur. Negatively reinforced balanced urn schemes. *Adv. in Appl. Math.*, 105:48–82, 2019.

[22] Cécile Mailler and Jean-François Marckert. Measure-valued Pólya urn processes. *Electron. J. Probab.*, 22:Paper No. 26, 33, 2017.

[23] G. Pólya. Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré*, 1(2):117–161, 1930.