Shape theory and topological statistics

Shape Theory and Topological Statistics

Kemal Gürsoy

July 30, 2020

Rutgers University, Department of MSIS
100 Rockafeller Road, Piscataway, NJ, 08854
e-mail: kgursoy@business.rutgers.edu

Abstract

In this work we consider a survey on the geometric statistics and shape theory to understand how we may utilize shape information to learn from data.

Keywords: geometry, statistics, topology, shape, data

1 Introduction

Henri Poincaré [68, 69, 70, 71, 72] categorized problems as either uninteresting problems, such as problems that definitely have a solution or do not have a solution, or as interesting problems, that are productive problems such that identifying them as solvable or not solvable problems would not be sufficient, since they also produce new questions and problems that have to be investigated for their own sake. Geometric statistics is one of the interesting methods for extracting information from the real life experience and construct generalized expectations for life experiments, based on the shape of the data set information.

2 Intuitive Construction

2.1 Experiment Desing

Assume that there are finitely many projects to activate, one at a time, until the activated project is completed, such that.

A1) Each project i has a positive reward, R_i, upon its completion.

A2) The completion time of each project i is a positive and conditionally independent random variable $\tau_i \sim F_i(x_i, t_i)$, subject to a probability law. This probability law is identified by its distribution function, F_i, based on the state of the project, x_i, and the activation time of the project, t_i.

A3) The expected reward of a project i, $E[R_i e^{-\alpha\tau_i}|x_i, t_i]$, depends upon its completion time, τ_i, where $\alpha \in (0, 1)$ is the time-discount factor, for all projects.

2.2 An Experiment Selection Policy

Let there be a collection of finitely many, N, projects to activate. One project will be activated after completing another project, and we want to activate all of them. We wish to collect the best possible expected reward by completing all the projects. Without loss of generality, assume that for a pair of projects, (project i, project j), the expected discounted reward of the project i is greater than the expected discounted reward of the project j.

A selection policy orders activation times of these projects, the project i first and the project j second, based on their expected rewards, as follows:

$$E[R_i e^{-\alpha\tau_i} + R_j e^{-\alpha(\tau_i+\tau_j)}] > E[R_j e^{-\alpha\tau_j} + R_i e^{-\alpha(\tau_j+\tau_i)}].$$

We can arrange terms, by using the linearity property of the expectation operator. Thus, we have:

$$E[R_i e^{-\alpha\tau_i}] + E[R_j e^{-\alpha(\tau_i+\tau_j)}] > E[R_j e^{-\alpha\tau_j}] + E[R_i e^{-\alpha(\tau_j+\tau_i)}].$$

We may rearrange terms, by using the independence assumption of the completion times.

Now we have:

$$E[R_i e^{-\alpha\tau_i}] + E[R_j e^{-\alpha\tau_i}]E[R_j e^{-\alpha\tau_j}] > E[R_j e^{-\alpha\tau_j}] + E[R_i e^{-\alpha\tau_j}]E[R_i e^{-\alpha\tau_i}].$$

Hence, by arranging terms for these individual projects, we obtain:

$\frac{E[R_i e^{-\alpha\tau_i}]}{E[1-e^{-\alpha\tau_i}]} > \frac{E[R_j e^{-\alpha\tau_j}]}{E[1-e^{-\alpha\tau_j}]}$.

This comparison identifies that the project i must be activated first and the project j must be activated later, according to their expected discounted rewards comparison, in this ratio form.

2.3 The Optimal Sequential Design Policy

An ordering policy selects projects based on the diminishing values of the ratio of their expected discounted rewards, $\frac{E[R_i e^{-\alpha\tau_i}]}{E[1-e^{-\alpha\tau_i}]}$. Let $g_i = \frac{E[R_i e^{-\alpha\tau_i}]}{E[1-e^{-\alpha\tau_i}]}$ be a *project activation index*, for the project i. Now, order the project activation indices, $\{g_{[i]}\}$, based on the diminishing values of the computed project activation index, such that $g_{[1]}$ is the maximum and $g_{[N]}$ is the minimum of g_i values, for all projects.

Theorem

The project activation policy that activates projects in the order of project [1], project [2], project [3], . . . , project [N-1], project [N], which is identified by the ordered indices, $g_{[1]} > g_{[2]} > g_{[3]} > \ldots > g_{[N-2]} > g_{[N-1]} > g_{[N]}$, provides the optimal expected discounted total reward of the projects.

Sketch of the proof

First activate an inferior reward value project. This will delay the activation of the superior reward value project. Therefore, this action may not provide the best expected discounted total reward.

3 Background Information

Let there be an algebraic polynomial, p, with two variables, say x and y. The set $\{(x,y)|p(x,y) = 0\}$ for $x \in \mathcal{R}$ and $y \in \mathcal{R}$ identifies a curve in $\mathcal{R}^2$. When p has a known degree n, what are the possible topological structures of this curve? As an example, when $n = 2$ the curve may be a circle, an ellipse, a hyperbola, a parabola, a pair of lines, or the plane. When we introduce a line at infinity, we simplify categorization by constructing the projective plane, since ellipse, hyperbola, and parabola have the same structure in the projective plane, being identified only by the position of the circle with respect to the line at infinity. But, when $n > 2$, the problem begins to get complicated, at least for this day and age. R. Descartes successfully investigated the case for $n = 3$ and $n = 4$, see Descartes [21]. A curve of degree n has no more than $g + 1 = \frac{(n-1)(n-2)}{2} + 1$ connected components, where g is the genus of the corresponding Riemann surface, formed by the complex solutions of the curve in the complex projective plane, $\mathcal{CP}^2$. Every connected and closed orientable surface is a surface of genus g, where g is the number of handles that have to be added to a sphere so that the surface is generated, see Alexandrov [3]. A real curve with the degree of n has no more than $g + 1$ components, where each component is diffeomorphic to the circle S^1. Now, imagine the graph of a real polynomial of degree n, as a surface $z = p(x, y)$, in $\mathcal{R}^3$. Wherever the surface is locally convex, the support plane is an elliptical space. Wherever the surface is locally a saddle, the support plane is a hyperbolic space. The elliptical and hyperbolic spaces are divided by a curve of parabolic points, and the curve of these parabolic points can be identified by the Hessian of the surface being equal to zero. Also, the sign of this Hessian is equivalent to the sign of the Gaussian curvature of the graph, see Cartan [15], also see Cartan and Eilenberg [16]. As a consequence, the topological structure of the complex sphere in projective space $\mathcal{CP}^3$ is identified as a four-dimensional manifold, that is diffeomorphic to the Cartesian product of two spheres, by utilizing an affine space $x^2+y^2+z^2 = 1$. Topological construct is necessary, otherwise we will end with absurd geometrical generalizations of rational parametrization, see Arnold [5]. For example all smooth algebraic curves of degree n in $\mathcal{CP}^2$ have the same genus, independent of the curve. In fact, in the real projective plane not only the topological classification of curves for a given degree, but more importantly, the topological classification of the polynomials that identify these curves must be constructed.

The study of links and knots in topology opened many new interesting problems; most are not answered yet. A knot is a smooth embedding of a circle in $\mathcal{R}^3$. A link is an embedded finite union of disjoint circles, see Atiyah [7]. The topological structure of a smooth real function is associated with a graph with connected components of the level hypersurfaces of this function. A smooth function with n variables is known as a Morse function, when every critical point is nondegenerate; that is the Hessian is not equal to zero at this point, see Cartan [15]. Any Morse function has finitely many critical points on any given compact domain. Moreover, every smooth function, through all its derivatives, can be identified by a Morse function. This implies that any randomly selected smooth function is a Morse function, see Arnold [5]. For a Morse function on a compact manifold, for example on the sphere, we accumulate every component of every level hypersurface to a point, and the resulting space is a graph. Here, the vertices identify the critical points of the function, and an edge that connects two vertices identifies a path that does not cross level hypersurfaces that are generated by other critical points. For a nondegenerate Morse function, this graph is a tree, where leaves identify equilibrium points (maxima, or minima, or saddle points). Hence, topological structures identify diffeomorphisms as well. Evidently, level hypersurfaces will be constructed by the generations of this tree, thus homeomorphisms would be constructed for the local equilibrium points, for the given orbits. Alternatively, smooth functions on tori generate graphs with loops; therefore analyzing extreme modes would be more complicated than looking for the extreme hanging branches of a tree.

When we investigate statistics on the graph of a function, we could utilize the tree structure, by ordering the branches with respect to the height of the leaves and utilizing level sets, see Prékopa [73]. For these relationships, it is intuitively interesting to investigate the evolutionary process of topological structures and identify what causes the significant property changes. Moreover, statistical investigations for the topology of periodic functions would provide a strong link between families of trigonometric polynomials with a given degree on a torus, and a spectrum that would yield properties of connectedness of regions on a complex manifold of harmonic waves.

A *field* is an algebraic construct with two binary operations on a set, namely addition and multiplication. Galois fields are fields with finitely many element sets, see Artin [6]. The number of elements of a Galois field is an integer power of a prime number. Hence, for a pair consisting of a prime number and a positive integer power, there is a Galois field with fixed number of elements, up to isomorphism. This finite isomorphism may be identified by random permutations of the cardinality of the generating set, see Arnold [5].

Analysis, topology, geometry, and algebra converged together and collectively evolved into a global role, to generate *interesting* problems. As an example, consider representing connections in differentiable manifolds with gauge theory, see Atiyah [7].

3.1 A Space of Dimensions

We may translate information from one dimension to another one, with some degree of loss in the information.

Model assumptions

A1) Assume an infinite (Platonic) line, see Euclid [29].

Pick a location on this line and fix it, call it the origin. Select a constant unit for a step size. Starting from the origin, walk on this line in the positive direction, where each step has a marked foot location i, for $i = 1, 2, \dots$.

Also, if you start from the origin and walk in the negative direction, this walk identifies the position of the negative foot position, for $i = -1, -2, -3, \dots$.

A2) Assume that the *foot location i* on this line identifies the position of an i *dimensional space.*

A3) Assume that every position on this real line, not only the footprints, identifies a corresponding dimensional space.

In this construct, the origin is the space of all zero dimensional objects, a.k.a. points.

Also, through the walk on this line, the positions of the dimension spaces are related. Hence, information may be translated from one dimension to another dimension, but some degree of translation loss would happen to the quality of information. Consequently, projections and anti-projections may be used to translate information, see Hartshorne [45].

For example imagine that there is a cube, a three dimensional object, of unknown volume. If we wish to find the magnitude of the volume of this cube, we may find the *footprint* of a surface of this cube, as a square, a two dimensional object that is representing the projection of the cube on a plane.

Then, by measuring the length of a side of this square, we may find the magnitude of the volume of the original cube. Therefore, the two dimensional projection of a three dimensional object may preserve the original information, such as the magnitude of the volume.

Furthermore, if we find the *footprint* of one side of this square, as a line segment, that is a one dimensional object, this second projection still preserves the original information, such as the magnitude of the volume of the cube. But, if we construct the projection, i.e., the *footprint*, of this line segment as a point, a zero dimensional object, then we shall loose the original information about the magnitude of the cube's volume. Moreover, consider a one dimensional object, a line, representing the required information, and there is a projection of this line on a point. If this point is missing, then this unknown point represents a projection of a negative one dimensional object. That is, there exists one dimensional information, but we do not have its representation. Therefore, we may imagine the negative dimensions as missing information for that dimension. Intuitively, we may say that the distance between the points of this space of dimensions is a factor of the degree of preserving the translated information.

Alternatively, if we have a projection representing the original entity, then this projection may not have sufficient information to identify the original entity. For example, imagine that we have a square, as a two dimensional projection of a three dimensional object, possibly a cube, and we wish to find the magnitude of the volume of this three dimensional object. If the object is a single cube, then we have the necessary and sufficient information to find the magnitude of the volume of this single cube, by using the obvious anti-projection. But, if we do not know whether there is a single cube, or a stack of similar cubes, towering over this square, then the information provided by this square is not sufficient to find the magnitude of the volume of this tower of cubes. In this case, we need to know how many of the similar cubes are stacked together. Or, perhaps changing our point of view and moving to another plane to investigate the two dimensional information may help. For example, if the original projection information is in the $\{(x, y)\}$ plane, then we may move into the $\{(x, z)\}$ plane, as well as into the $\{(y, z)\}$ plane and investigate these two dimensional images, generated by the three dimensional object in the $\{(x, y, z)\}$ space. Or, you may twist and turn the projective plane with respect to the three dimensional object, but you must need more information about these translations. Moreover, if you use nonlinear projections, then preserving the original information would be even more complicated with the necessity of additional information about these nonlinear translations.

3.2 Statistics for the Space of Dimensions

We can estimate the properties of a space of dimensions, based on the empirical evidence, as follows. Assign a probability measure, P, to every point of this space. If possible, by using the Lebesgue measure with $\int_{\{X_i\}} dP(X_i) = 1$, and measure the probability of an event, A, with $\int_{\{X_i\}\in A} dP(X_i)$, see Lebesgue [60], and Kolmogorov [57]. Also see Alexadroff [2], Federer [31], and Kendall [55].

Construct the k^{th}-moment of a point, by using this probability measure, $E[X_i^k] = \int_{\{X_i\}\in A} X_i^k dP(X_i)$, for all $k \in \mathcal{R}$. Use a data set, $\{X_1, X_2, \ldots, X_n\}$, for estimating these moments, $[\sum_{i=1}^n X_i^k]/n$, for all k. Hence, identify all the moments of the space by the estimated sample moments. This will estimate the behavior patterns of the space, except for the events with zero probability.

It is known that the Lebesgue measure is blind to the magnitude of points. For example, if you wish to measure the magnitude of the weight of a loaf of bread, slice this loaf of bread and weigh each slice, then add the weight of the individual slices for all slices to obtain the weight of the loaf of this bread, according to the Lebesgue measure definition. But, summing of the weights of all these slices ignores the weight of the bread crumbs made while the slicing the loaf, and these bread crumbs also have their own weights, greater than zero. Thus, ignoring the weight of the collection of all these bread crumbs would underestimate the actual measure of the total magnitude, sometimes significantly. Therefore, if points are important to estimate the probability measure for a space, then the Lebesgue measure would not be suitable to construct a probability law. In that case, construct a probability measure by using a suitable structure as a provider of the probability law, see Kallenberg [51], and Matilla [61].

3.3 Random Graphs and Percolation

The stochastic nature of global networks forces them to work under heterogeneous random environments, see Erdös and Rényi [28]. Given a collection of such random processes, we may impose a structure on them, see Prékopa [73, 74, 75], and Arnold [4]. We also identify patterns on them, see Gürsoy and Katehakis [43]. This working characteristic may be captured and represented by a stochastic percolation model, where the random movements take place in an unordered random medium. An important conjecture, projections of random Cantor sets was proved for fractal percolations, see Falconer [30]. It shows that a nonempty realization of a fractal percolation almost surely contains an interval, and thus enables us to construct a probability measure on a random space, see Rams and Simon [78].

If the medium were homogeneous and deterministic, then a stochastic diffusion model could be more suitable, see Itô and McKean [48]. The graph, G, is an object defined on a set of vertices (nodes), V, and a set of edges (arcs), E, that relate (connect) vertices. Graphs have properties identified by how these relations occur. For example, directed graphs (digraphs) or nondirected graphs, identify asymmetric or symmetric relations, respectively. It is standard notation to indicate a vertex, i, that is connected to another vertex, j, by i$\sim$j, which says i and j are adjacent. A random graph is a graph such that a pair of vertices are adjacent, subject to a probability law. That is, i$\sim$j with probability $p(i, j|\theta)$, $\forall$ i,j $\in$ V and with a parameter θ. We can denote a random graph by G(p), where p is the probability distribution of adjacency. Erdös observed that, among many other properties of random graphs, the degree to which a random graph be connected depends upon p, such that as $p(i, j|\theta) \to 1$, $\forall$ i,j $\in$ V, the graph will converge to be full, see Erdös [27], also see Erdös and Rényi [28].

Another observation was that the random graph would have properties with probability (moving from zero to one) according to a function of the cardinality of V, $|V| = n$. This is called a threshold function and is defined as follows. A function q(n) is a threshold function for a property M of the graph G. Comparing componentwise, and using P(A) as the probability measure of an event A:

If p$\gg$q(n), then P(G(p) has property M) $\to$ 0.

If p$\ll$q(n), then P(G(p) has property M) $\to$ 1.

This definition of a threshold function obeys Kolmogorov's "zero-one" law, such that: If $\lim_{n\to\infty}$ P(G(q(n)) has the property M)=0, then the property M holds with probability zero (almost nowhere). If $\lim_{n\to\infty}$ P(G(q(n)) has the property M)=1, then the property M holds with probability one (almost surely), according to the probability law ruling over $q(n)$, see Kolmogorov [57].

There is a relationship between irrational numbers and the zero-one law, such that: If $\alpha \in (0, 1)$ is an irrational number, then q(n) = $n^{-\alpha}$ satisfies the zero-one law. If $q(n) = n^{-\alpha+o(1)}$ with $\alpha \in (0, 1)$ being an irrational number, then for large n, $q(n)$ satisfies the zero-one law. For proofs, see Spencer [91] and the references therein. Hence, one expects the failure of the zero-one law for $\alpha \in (0, 1)$ being a rational number. Yet, there are many strange behaviors of random graphs observed for a rational α, when limiting probabilities are considered, see Spencer [91].

A property of graphs, P, can be called first order expressible when it can be written as a first order language. We can define a distance, $d \geq 0$, on a graph such that $d(x, y)$ is the length of the shortest path between two vertices, x and y on that graph, where $d(x, x) = 0$. The ϵ-neighborhood of a vertex, x, can be defined with respect to this distance measure, as follows. Let $x, y \in G(q(n))$. An $\epsilon > 0$ neighborhood of the vertex x is $b(x, \epsilon) = \{y|d(x, y) \leq \epsilon, \forall y \in G(q(n))\}$.

For every n, let there be a probability space, generated by all possible graphs $\{G(q(n))\}$, indexed by time, t. It is said that the sequence of probability spaces, $G_t(n)$, satisfies the zero-one law for every first order sentence, S_t, such that $lim_{n\to\infty} P(\{G_t(n) \models S_t\})$ is 0 or 1.

Vertices on a graph are called independent if no pair of vertices are connected by an edge. A graph is called r-chromatic if its vertices are colored by r distinct colors, such that no pair of connected vertices have the same color. It is said that a probability distribution function q of a random graph evolves as $G_t(q(n))$ moves from having a set of independent vertices to being complete.

Let $\frac{1}{n} \ll p \ll \frac{logn}{n}$, then $G_t(q(n))$ evolves subject to the zero-one law, see Spencer [91]. As p evolves in space-time, it generates a sequence of threshold functions. When $p = \frac{1}{nlogn}$, a random graph will be cycle free. Hence, it will be a random tree, $T(n)$.

For the following definitions, please see Erdös and Rény [28]. Assume a rooted tree, T, (a dedicated node is fixed as the root, r) with standard terminology of "root," "parent," "child," "generation." We define the depth of a vertex v as the distance from the root, $d(v, r)$, on T.

A random graph will be subject to percolation when $p = \frac{1}{n}$. When $p = \frac{logn}{n}$, a random graph will be connected. If $p = \frac{logn}{n} + \frac{c}{n}$, for any real constant, c, then $lim_{n \to \infty} P(\{G_t(q(n))$ is connected$\}) = exp(-exp(-c))$. Here $G_t(q(n))$ is connected means that $G_t(q(n))$ almost surely has no isolated vertices.

Assume a random rooted tree, $T(n)$, is subject to Bernoulli trials, that is, every node j can be connected to its parent node i with probability p or not connected with probability $1-p$. For large n, $T(n)$ will be asymptotically subject to a Gaussian probability distribution with $\mu = np$ and $\sigma^2 = np(1-p)$. Thus, the limiting distribution of the ϵ-neighborhood of a vertex v is asymptotically normal.

The decision process employed here uses the connectivity of a random graph. This raises the topological issue that a first order logic cannot capture and address this property. However a finite graph is a natural entity to construct topological structures, hence, it is essential to employ as an approach to our problem, since it provides a structure with a finitely possible number of random graphs, evolving in space-time.

A topological space, $\mathcal{T}$, is said to be unicoherent when any two collectively exhaustive sets in it have a connected intersection. A distance can be defined on a random graph, $G(q(n))$, as follows. For every pair of vertices, i and j, on $G(q(n))$, let $d(i, j)$ be $P(|i - j|)$, the probability measure as the distance metric between i and j. This definition has produced many fruitful results, Shelah [89].

The common approach is discretization of the random graphs and utilization of the discrete lattices generated, where the percolation on these lattices requires combinatorial approaches. In fact percolation can be seen as the connectivity properties of random graphs, see Penrose [67], and lattice percolation captures and explains properties of random graphs on the integer lattice, $\mathcal{Z}^d$, embedded in the Euclidean real space, $\mathcal{R}^d$. Let $0 \leq p \leq 1$ be the probability of success for a family of Bernoulli trials, $\mathcal{Z}^p$, such that a neighborhood of x is called open when $\mathcal{Z}^p(x) = 1$ (closed otherwise). Also let $\mathcal{B}^p$ denote the random set of open sites, a Bernoulli process.

The r-cluster, $r > 0$, at the origin for $\mathcal{B}^p$ is the open neighborhood in $\mathcal{B}^p$, containing the origin. Let X be a denumerable set and $(B_i(x), x \in X)$ be an indexed family of Bernoulli processes or a random field. For $k = 0, 1, \ldots$, $(X_x, x \in \mathcal{Z}^d)$ the random field is k-independent, when $A, B \subset \mathcal{Z}^d$ and $\| a - b \| > k$, $\forall a \in A$ and $\forall b \in B$.

The above random field translates into a natural partition of $\{e_1, e_2, \ldots, e_n\}$, where $e_i \in \mathcal{Z}^d$ is the ith canonical base, such that $\frac{1}{n} \sum_{i=1}^{n} X_i$ converges to $E[X]$ in distribution (multidimensional ergodicity). The percolation represents the behavior of large neighborhoods on a random graph, such as areas accessible by radio transmitters. The percolation probability is a measure of likelihood such that the origin is included of the random graph, with a cut-off value that depends upon the dimension and distance metric. The percolation probability is bounded above by the Galton-Watson branching process, and its behavior is observed with respect to the Eucliden l_2 norm, see Penrose [67].

Let us fix a general Euclidean l_p norm, denoted by $\| \, . \, \|_p$, for $1 \leq p \leq \infty$, and also let $B_k(G)$ denote the order of the largest component of G. If the mean vertex degree has a value greater than a critical value of one, then a *giant component* emerges, see Janson et al. [49].

Let $B_k(X) \in \mathcal{R}^d$ with a set of vertices X be a form $B_k(X) = \prod[x_k, y_k]$, for $k = 1, 2, \ldots, d$, in the Lebesgue sense, see Lebesgue [60]. Also let $\pi_k : \mathcal{R}^d \to \mathcal{R}$ be the projection on the k^{th} coordinate. Hence, the k-crossing for $B_k(X)$ is a set of vertices, $\{x_k, y_k \in X\}$, with $\{\| \, \pi_k(x_k) - \pi_k(y_k) \, \|_p \leq \frac{r}{2}\}$, such that there is a path which is embedded in the k-frame of $B_k(X)$. Thus, we have an upper bound for the percolation probability, $\lim_{n\to\infty} \sup n^{-(d-1)/d} \log \sum_{k\geq n} p_k < 0$, with an insight that there exists a $\lim_{n\to\infty} n^{-(d-1)/d} \log p_k$. This is an unproven conjecture; only cases for $d = 2$ and $d = 3$ are known, see Penrose [67].

Let Ω be measurable and $X, Y \in \Omega$. If $f : \Omega \times \Omega \to \mathcal{R}$ is σ-finite on $\Omega \times \Omega$, then $Var[f(X,Y)|Y] \leq Var[f(X,Y)]$ almost everywhere, by the monotone convergence, see Kallenberg [51].

Given a finite random graph, there is a family of problems that can be approached similarly. For example, the maximum flow problem and the maximum covering problem can be seen as members of the same general linear group over $G(n)$, subject to optimizing a function of this graph, for example, cost or profit. Also, the connectivity problem has a simple interpretation and strong solution approaches for finite random graphs, such that a connectivity threshold map of finite random graphs may be identified by the diameter of the minimal spanning tree on them. Some work has been done with this as an application to wireless networks and computer network security, see Gupta and Kumar [40]. A particular point of view for connectivity is quite suitable in this context, such that a random graph is k-connected if for each pair of vertices, there exist at least k independent paths connecting the vertex pairs, for a nonnegative integer k (if k is zero, then the graph is disconnected). Similarly, a random graph is k-edge-connected when any pair of vertices is connected by at least k-edge-disjoint paths (that is paths with no common edges), see Bollobas [11]. Also a k-separating-pair of a random graph is a pair of disjoint vertex sets such that there are connected random subgraphs induced by this pair. Hence, a random graph is k-connected when it is k-edge-connected or it has a k-separating pair. A large number of results emerged from the partitioning of graphs into subgraphs. For example, if for some natural number k, the set of all k-subsets of $\mathcal{N}$ (natural numbers) is divided into finitely many classes, then $\mathcal{N}$ has an infinite subset with its k-subsets belong to the same class, see Ramsey [79]. Ramsey's theorem describes graph partitioning in terms of colorings, such as duochromatic structures that have red-blue colors. For any two natural numbers, s and t, the Ramsey number, $R(s,t)$, is defined to be the minimum number for which every graph of order $R(s,t)$ has the clique number at least s, or independence number at least t. There are conjectures about possible upper bounds for the Ramsey number, such as those using the clique numbers of Paley graphs, see Bollobas [11].

Let us fix f to be the probability density function over Ω, and let the diameter for a metric space be defined as usual, $D(\Omega) = \sup\{\| x - y \|_p | x, y \in \Omega\}$. Also, let Ω be a product space of finite intervals, including the origin, such that $\Omega = [0, \omega_1] \times [0, \omega_2] \times \ldots \times [0, \omega_d]$, where $\omega_j > 0$, for $j = 1, 2, \ldots, d$. There are strong relationships between connectivity and the influence of vertices, according to the law of large numbers, see Penrose [67]. Also see Borodzik et all [12] and Wasserman [99].

We may wish that a network should provide dissemination of information with maximum reliability. Here, connectedness probability can be seen as the probability of operability; and even if some subset of the network probably fails, we want the network to be able to perform with the remaining components.

3.4 Empirical Bayesian Networks

The parameter estimation, based on empirical data $\{X_1, \ldots, X_n\}$, and subject to a probability density $f(x|\theta)$, where the parameter θ has its own (usually unknown) probability distribution, H, is the basis for Bayes estimators. When the data are first used to estimate the unknown H, instead of assuming a prior distribution, the parameter estimation method is known to be the empirical Bayes and its application area is usually known as the compound decision theory, see Robbins [81], and Neyman [65].

Causality can be captured by statistical analysis, see Prékopa [76], or by Bayesian digraphs, see Jensen [50], and Pearl [66]. Bayesian networks have been constructed as structures to represent causal models, even if Bayesian networks do not imply causality, per se. We can say that a Bayesian network is a digraph that inherits the parent-and-child relationship in its conditional probability distribution of connectedness, such that $P(Child|Parents, \theta)$ is the prior probability of connectedness on a Bayesian network.

When we have empirical evidence available, we can first estimate the unknown probability distribution of the parameter θ by using the available data. Thus we can coin the phrase "empirical Bayesian networks." In turn, these networks are used to generate posterior probabilities, see Robbins [81], based on the realized evidence. In particular, real world relationships evolve in space-time, thus the random graph better be constructed on a space-time domain. For example, when the connectedness probabilities are in nature with the Markov property; that is for future forecasting, present time has sufficient information, but the events are not completely observable, then one can use hidden Markov models, see Elliot et al. [26]. Or, one can use a model adapted to the new evidence through the Bayesian construct and refine its σ-field, sequentially in space-time, see Robbins and Monroe [82]. In the long run, this adaptation process will identify persistent states and filter out the effects of transient states. Therefore, most probable evolutionary behavior for a random graph will be estimated, based on the incoming stream of evidence.

3.5 Evolving Empirical Bayesian Graphs

Evolution is a natural, as well as a causal, process where the environment shapes the behavior of a system so that the system adapts to the environmental conditions, even if they are also changing. An empirical Bayesian graph captures and explains the evolution of random networks, based on empirical evidence, without assuming a prior probability law.

The homology identifies structural similarities, such as similar limbs of different animals. For example, arms of a human and wings of a dove share a similar skeletal structure, yet they evolved to adapt to their environmental conditions. A transformation between homological groups, cohomology, induces similarities that identify changes from one form into another. Thus, cohomologies are transformations that represent evolving graphs.

Behavior of a random graph is subject to a probability law; and an evolving random graph can behave according to a hidden Markov chain that represents the partially observable stochastic dynamic system's behavior, where one cannot observe every state, but can still be able to forecast the behavior, for example by using an extended Kalman filter, see Gülcür et al. [39].

Periodicity becomes an important property, where probability distributions on a sphere will be beneficial to construct the zero-one law for periodic Markov chains. In fact, it is an open problem to consider periodic, or semi-periodic probability densities on random graphs, see Penrose [67].

4 Dimension, Algebra, Combinatorial Topology

The connection with dimensions and combinatorial structures through algebra was provided by the Stanley's conjecture, that states that *Cohen-Macauley simplicial complex is partitionable*, see Stanley [92]. Simplicial complexes represent many-to-many relations, and they play natural roles in the random graphs as a closed family of subsets of vertices. Thus, they represent a topological structure of a simplex, identified by a geometric dimension. Hence, the question is how to identify Cohen-Macauley complexes by geometric properties? This extensive investigative process has built many avenues in algebra, geometry, and topology. Especially, partitioning of complexes produced combinatorial problems. For example constructible Cohen-Macauley rings in algebraic geometry are the basis of combinatorial construction of simplicial complexes on vertex sets, but with difficulties of translating the depth of a piece of graph information into a related dimension information. Although constructible topologies imply the Cohen-Macauley property, the inverse relationship is not true. Recently, a combinatorial construct, known as the Stanley depth, opened a door for commutative algebra approaches, see Stanley [92]. The Stanley depth generates a useful algebraic upper bound for the combinatorial problem, and supports constructible topologies, see Garsia and Stanton [32]. This recent progress shows that many different mathematical fields, such as algebra, geometry, topology, combinatorics, and number theory, can productively get together and generate interesting problems.

4.1 Finite Groups and Combinatorial Structures

The invariant theory of finite groups and finite fields plays an important role in algebraic topology, especially the modular case, where the characteristic of the field divides the order of the group, see Adem and Milgram [1], Benson [9], Cartan and Eilenberg [16], Cohen [18], Hartshorne [46], Atiyah and McDonald [8], and Eisenbud [25]. The effort to unify invariant theory, action of groups on rings, and the fixed subrings of the action will provide algebraic analogies to combinatorial constructs. As a consequence of the above efforts, we obtain constructive proofs that yield algorithms to solve finite group problems. Pick a field $\mathcal{F}$, then a finite dimensional vector space V over $\mathcal{F}$, $\mathcal{F}[V]$ will be the algebra of homogeneous polynomial functions or forms on V. The homogeneous component of $\mathcal{F}[V]$ of degree d, denoted by $\mathcal{F}[V]^d$, has $S^d(V^*)$ as the d-th symmetric power of V^*, where V^* is the dual of V. Let $b_1, \ldots, b_n \in V$ be a basis, then the elements of $\mathcal{F}[b_1, \ldots, b_n]$ are homogeneous polynomials in the linear forms with coefficients in $\mathcal{F}$, where the elements of the polynomial algebra $\mathcal{F}[V]$ are functions. Let G be a finite group and $\rho : G \to GL(n, \mathcal{F})$ be a representation of G. Thus G acts on $\mathcal{F}^n$ through linear transformations, the set of G-invariant polynomials, $\mathcal{F}[V]^G = \{f \in \mathcal{F}[V] | gf = f, \forall g \in G\}$. Since $\mathcal{F}[V]^G \subseteq \mathcal{F}[V]$ is a finite extension of the algebra, $\mathcal{F}[V]^G$ is totally finite in the sense that $\bigoplus_i (\mathcal{F}[V]^G)_i$ is a finite dimensional $\mathcal{F}$-vector space. The invariants of the symmetric group, S_n, acting in its canonical representation as a permutation group of $x_1, \ldots, x_n$, would be $\mathcal{F}[x_1, \ldots, x_n]^{S_n} = \mathcal{F}[e_1, \ldots, e_n]$, where $e_1, \ldots, e_n$ are the elementary symmetric polynomials in $x_1, \ldots, x_n$. The invariants of the alternating group, A_n, in its tautological representation as $\mathcal{F}[x_1, \ldots, x_n]^{A_n}$ are generated as an algebra by $e_1, \ldots, e_n$ and the polynomial, P, obtained by summing all the elements of A_n that orbit the monomial $x_1^1.x_2^2 \ldots x_{n-1}^{n-1}$. But they are not algebraically independent of a polynomial of $e_1, \ldots, e_n$, P^2. The Steenrod algebra, which is a way to organize information and provides an instrument to construct new invariants from old ones, is a basis for inductive recursions utilized to construct the combinatorial structures.

The finiteness is very important for the invariant theory, where several types of structural, homological, and combinatorial finiteness theorems constitute the basic finiteness theorems. Homological finiteness focuses on the length of Syzygy chains and the finiteness of various homological dimensions that can be associated with a ring of invariants. Combinatorial finiteness is concerned with the rate of growth of the sequence of integers, i_k and the dimension of the space of homogeneous invariant polynomials $S^d(V^*)^G$ of degree d. These integers are associated with the Poincaré series $\mathcal{P}(\mathcal{F}[V]^G, t) = \sum_{k=0}^{\infty} i_k t^k$, where $i_k = dim_{\mathcal{F}}(\mathcal{F}[V]^G_k)$. If H is a subgroup of a finite group G, and V is a finite-dimensional G-representation, then the relative transfer from H to G,

$Tr^G_H : \mathcal{F}[V]^H \to \mathcal{F}[V]^G$, is defined as

$Tr^G_H(f)(x) = \sum_{gH \in G/H} g(f)(x) = \sum_{gH \in G/H} f(g^{-1}(x)) \ \forall x \in V$.

This is the basic form of the Hilbert-Noether theorem which states that a finite dimensional representation of a finite group generates a totally finite algebra. Its constructive version with an upper bound on the number and degrees of generators, known as Noether bounds, is found in Schmid [85].

The basic combinatorial finiteness theorem for the field of complex numbers, $\mathcal{F} = \mathcal{C}$, has been known for a long time, and it associates the Poincaré series of $\mathcal{C}[V]^G$ with the average of reciprocals of the characteristic polynomials of the elements of G. A sequence of elements $a_1, \ldots, a_n$ in a graded commutative algebra over $\mathcal{F}$ is a homogeneous system of parameters if the quotient algebra $A/(a_1, \ldots, a_n)$ is totally finite. A system of parameters is algebraically independent, thus, it is a finite extension of a polynomial algebra.

Noether normalization theorem

Let A be a finitely generated graded algebra over a field $\mathcal{F}$. Thus there exists a system of parameters for A, and the following integers are equal, see Eisenbud [25].

i. The smallest integer k such that there exists k elements $a_1, \ldots, a_k \in A$ with $A/(a_1, \ldots, a_k)$ finite.

ii. The largest integer m such that there exists m algebraically independent elements in A.

iii. The length l of the longest increasing chain $I_1 \subset \ldots \subset I_l \subseteq A$ of prime ideals in A.

The permutation group has the following theorem with a constructive proof that provides a very useful algorithm to solve our problem, see Hochester and Eagon [47].

Prime characterization theorem

Let $\rho : G \hookrightarrow GL(n, \mathcal{F})$ be a representation of a finite group G over a field $\mathcal{F}$. If $|G|$ is prime to the characteristic of $\mathcal{F}$, then $\mathcal{F}[V]^G$ is a Cohen-Macauley algebra. Here we utilize the above homological finiteness theorem to construct a geometric divide-and-conquer method as follows, see Singer and Thorpe [90].

Step 1) A system of parameters has been generated over the field $\mathcal{F}$, also known as the primary generators of $\mathcal{F}[V]^G$.

Step 2) Find a basis for $\mathcal{F}[V]^G$ by using the primary generators.

Computational efficiency requires that the basis search effort be reduced, and Noether's bounds may be useful for this purpose. The permutation representations form another class where we can find good upper and lower bounds. Moreover, we have algorithms to compute the ring of invariants, see Garsia and Stanton [32], also see Hochester and Eagon [47].

Let X be a finite G-set, $|X| = n$, defined by $\rho : G \hookrightarrow S_n$. The following holds, $\mathcal{F}[e_1, \ldots, e_n] = \mathcal{F}[X]^{S_n} \subseteq \mathcal{F}[X]^G \subseteq \mathcal{F}[X]$, between the rings of invariants. The elementary symmetric polynomials $e_1, \ldots, e_n \in \mathcal{F}[X]^{S_n}$ are a system of parameters and $\mathcal{F}$ is a Cohen-Macaulay algebra. Hence $\mathcal{F}[X]$ is a free finitely generated $\mathcal{F}[X]^{S_n}$-module and the Poincaré series of $\mathcal{F} \otimes_{\mathcal{F}[X]^{S_n}} \mathcal{F}[X]$ has degree $\begin{pmatrix} n \\ 2 \end{pmatrix}$. Hence $\mathcal{F}[X]^{S_n}$ is zero in nonhomogeneous degrees larger than $\frac{n(n-1)}{2}$. Thus $\mathcal{F} \otimes_{\mathcal{F}[X]^{S_n}} \mathcal{F}[X]^G$ is also totally finite and zero in degrees larger than

$\begin{pmatrix} n \\ 2 \end{pmatrix}$. Consequently $e_1, \ldots, e_n$ is also a system of parameters for $\mathcal{F}[X]^G$ which is generated as an $\mathcal{F}[X]^{S_n}$-module by polynomials of degree less than or equal to $\frac{n(n-1)}{2}$ and as an algebra by polynomials of degree at most $\max\{n, \frac{n(n-1)}{2}\}$. This bound is better than $|G|$, Noether's bound, and the bound $\max\{n, \frac{n(n-1)}{2}\}$ for permutation representations of degree n is sharp since the ring of invariants of the alternating group A_n is generated by the elementary symmetric functions $e_1, \ldots, e_n$, where the discriminant has degree $\frac{n(n-1)}{2}$. Therefore the longest regular sequence in $\mathcal{F}[V]^G$ has a length of at most $dim_{\mathcal{F}}(V)$, which is finite. This upper bound is equal to the homological codimension of $\mathcal{F}[V]^G$. If $\rho : G \hookrightarrow GL(n, \mathcal{F})$ is a representation over a finite group, then the Steenrod operations and action of G on $\mathcal{F}[V]$ commute, thus $\mathcal{F}[V]^G$ is mapped into itself by Steenrod operations as a combinatorial operator. One would like to compute the cohomology of orbit space for Poincaré series, which is in fact equivalent to the computation of the group cohomology of the torsion-free discrete group, which is a matroidal structure where we identify solutions to it by a greedy algorithm. That is, fix a vertex at random and include it in a set, then delete this vertex and its neighbors. Continue picking vertices at random and including them in the selection set, while deleting them and their neighbors. This is a random algorithm that will generate the probability measure identical to the one that was generated by a greedy algorithm, and that yields the expected order of a k-connected independent subgraph, on a given finite random graph.

Hence we have the basis for recursive constructions in the permutation representations that provide good bounds for a divide-and-conquer algorithm, Gürsoy [41], to compute and identify optimal policies for the collective generalized influential reward processes on random graphs, see Bollobas [11], Erdös [27], Kelley [54], Spencer [91], Penrose [67], and Gürsoy [42].

5 Selecting Ensemble of Experiments

The problem is to decide how to utilize available "projects = experiments," in order to maximize the collective utility of them. Consequently, relevant facts will be obtained and processed by a background knowledge, in order to generate sufficient and necessary information to make the best possible decision. In the rich and long history of decision making methods, there is one method that is relevant for this work, and it is known as the sequential analysis, see Dodge and Romig [22], Wald [97, 98], Thompson [93], and Robbins [80]. Sequential analysis refers to deciding when is the best time to terminate an experiment, and to take actions accordingly, see Chow et al. [17], and Blackwell and Ferguson [10]. The goal is to allocate resources sequentially into several activities such that the overall benefits collected during the life cycle of all activities will be maximized. The multi-armed bandit problem concerns these types of problems, see Gittins and Jones [36], and Gittins [33]. A multi-armed bandit refers to a collection of independent binomial reward processes; and a multi-armed bandit problem is how to select one process and to decide how long to activate it, in order to collect the maximum expected average reward. It is assumed that the state of the activated process will change according to a probability law, yet the states of the inactive processes will not change. The state-space is the Cartesian product of the individual bandits' state spaces, and the σ-field is generated by the tensor product of individual bandits' σ-fields, parametrized by time. Thus, the evolution of a multi-armed bandit process is governed by a semi-Markov decision process, with a binary action space. As one process is activated, the probability law that governs this process is learned, to some extent. Therefore, choosing between bandits that are known versus bandits that are unknown introduces the dilemma of choice; thus it may be tempting to activate the bandit with which we have the most experience, for less risk and immediate gain. But, a bandit with which we have not experimented yet may have a better chance to be the most profitable one. That is the dilemma of taking actions that yield immediate reward versus actions whose rewards would only appear in the future, see Kiefer and Wolfowitz [56], and Lai and Robbins [59]. There is an approach to solve these classes of problems, based on the Hardy-Littlewood maximal function theorem, see Hardy et al. [44], and Atiyah [7]. It states that among a set of measurable functions over an index set I, $\{f_i | i \in I\}$, the maximal function is found by comparing the volume generated by these functions, subject to a positive measure, say $\mu > 0$, relative to the volume generated by the measure itself. Then the supremum of this ratio identifies the maximal function $Mf = \sup_{i \in I}\{\int d\mu(f_i)/\int d\mu\}$.

An economical index based on ensemble averages of economical entities and maximal function theorem was constructed by Wald [95]. Another application of this theorem is to solve the multi-armed bandit problems by constructing a positive measure of performance for bandits that depends upon their history and to allocate resources at each decision moment to a bandit that has the highest performance measure at that time. This performance measure is known as the Gittins index, see Gittins and Jones [36]. If there are a finite number of bandits, say N, then let $\alpha \in (0, 1)$ be a discount factor, X_i^t be the state of the bandit i at time t, and $R(X_i^t)$ be the bounded reward generated by activating the bandit i at time t, starting at an initial time s, for $i = 1, 2, 3, \ldots, N$. We could now connect the empirical evidence and optimal decision making by using mathematical constructs, as presented by Wald [96], Schwartz [86, 87], Prékopa [73, 74, 75], Robbins and Siegmund [83], Seip [88], Arnold [4], and Nahmod [62], among others.

We are trying to maximize the discounted total expected reward by selecting an activation order of the projects, $E[\sum_{i=1,N}\sum_{t=s,T}\alpha^t R(X_i^t)|F(X_i^s)]$.

Hence, the Gittins index for the bandit i, at time s, to identify this selection would be $G_i(s) = \sup_{\tau > s} E[\sum_{t=s,\tau-1}\alpha^t R(X_i^t)|F(X_i^s)]/E[\sum_{t=s,\tau-1}\alpha^t|F(X_i^s)]$, where $F(X_i^s)$ is the σ-field generated by the history of X_i^s and τ is the stopping time of the activated bandit i.

Thus, $\tau = arg\sup_{\tau > s}\{E[\sum_{t=s,\tau-1}\alpha^t R(X_i^t)|F(X_i^s)]/E[\sum_{t=s,\tau-1}\alpha^t|F(X_i^s)]\}$. This formulation helps to compare performance of the bandits and then select the one that will provide the maximum average reward rate, together with its activation time duration. Ordering bandits by using their Gittins indices, from the highest index to the lowest index, provides an optimal policy for activating them, see Gittins [33]. For more comprehensive information, please also see Gittins [34, 35]. A generalization of the multi-armed bandit problem is to let bandits influence each other, while keeping the independence assumption for the state transition probabilities for the bandits. A multiplicative influence for the rewards was introduced by Nash, with a performance measure different than Gittins index, known as the Nash index, see Nash and Gittins [64], also see Nash [63]. If $0 < Q(X_j^t) < 1$, then this gives the original Nash index construct, which indicates that there is a diminishing collaboration between bandits. Let $-1 < Q(X_j^t) < 1$ be the influence factor of bandit j at a state X_j^t that is not activated at time t. Here, we generalized this multiplicative effect between bandits, and this influence factor may be constructed as a function of the covariances between bandits, such that if $Q(X_j^t)$ is negative; then this implies that there is an adverse effect between the bandits, such as a competitive influence. Hence, cooperation and competition among projects is captured by this generalization, see Gürsoy [42]. Consequently, for $i \neq j$, the discounted influential expected reward would be defined as $E[\sum_{i=1,N} \sum_{t=s,T} \alpha^t R(X_i^t) \Pi_{j=1,N} Q_j(X_j^t) | F(X_i^s)]$.

Also, we could define the generalized Nash index for the bandit i as follows, $N_i(s) = \sup_{\tau \in T} \{E[\sum_{t=s,\tau-1} \alpha^t R(X_i^t) | F(X_i^s)] / E[Q(X_i^s) - \alpha^t Q(X_i^\tau) | F(X_i^s)]\}$.

6 Optimal Policy

If one is capable of activating more than one project at a time, then the search for an optimal policy that would generate the best possible expected collective return is a reasonable enterprise. The possibility of finding an optimal selection policy, for this search, was conjectured by Whittle [101, 102]. The multi-armed bandit problem, together with the Nash index policy, also provides a fruitful approach for choosing a subset of competing projects out of a set of feasible projects. Hence, the collective activation of influential armed-bandits model would be.

Assumptions

(i) There are finitely many, N, statistically independent processes (projects) to choose from and each one has a state of nature, X_i^t, at a time t in a finite time horizon, for $i = 1, 2, \ldots, N$.

(ii) Each selected project provides a positive reward on activation, $R(X_i^t) > 0$, subject to a discount factor, $\alpha \in (0, 1)$, that identifies the present value of the reward.

(iii) Each nonselected project influences the reward of the selected projects with a finite multiplicative factor, $Q(X_j^t)$, at time t, for $j = 1, 2, \ldots, N$.

(iv) The selected project changes its state, but nonselected ones do not change their states.

The above model expresses that each selected project provides a reward according to a probability distribution where a project also influences the reward of other projects through a multiplicative influence factor, known as the generalized reward-decision processes, Gittins [35]. Here, as an extension of this model, an ensemble of processes is selected at a decision time and their collective time-discounted rewards provides the decision making criterion. Let us select $1 \leq k \leq N$ projects at a time. Also, let I be the index set of the activated bandits and J be the index set of inactive bandits. Thus, the collective expected return, subject to competition at time t, will be $\Pi_{j \in J} Q_j(X_j^t) \sum_{i \in I} \alpha^t E[R_i(X_i^t)|F(X_i^t)]$.

Let Family-1 be defined as a collection of bandits with the following property $E[Q(X_j^s) - \alpha^t Q(X_j^\tau)|F(X_j^s)] < 0$, for $\tau > s$, where the time set would be $T = \{\tau | E[Q(X_j^s) - \alpha^t Q(X_j^\tau)|F(X_j^s)] < 0\}$, for $\tau > s$.

Also, let Family-2 be the collection of bandits with the following property $E[Q(X_j^s) - \alpha^t Q(X_j^\tau)|F(X_j^s)] \geq 0$, for $\tau > s$, where the time set would be $T = \{\tau | E[Q(X_j^s) - \alpha^t Q(X_j^\tau)|F(X_j^s)] \geq 0\}$, for $\tau > s$.

This introduces a preference relation between the two families such that Family-1 is preferred to Family-2. According to this construct, there is an optimal policy that maximizes the overall expected discounted rewards, as follows.

a) Select a bandit in Family-1 with maximum Nash index; find its stopping time; set the initial time, s, to this stopping time for the next project; activate it; finish it; and repeat.

b) When Family-1 is emptied, then apply the selection process to Family-2, until it is emptied.

A proof of the optimality of this policy was provided by Nash [63].

The finiteness of the state-space and the time-horizon enable us to construct a combinatorial algorithm with finite steps. The rate of growth is related to the dimension of the space, according to the Hilbert-Noether theorem, see Schmid [85]. There is an upper bound on the number of generators. Hence, the orbit space, or power set, of the bandits collective is constructed by permuting its coordinates, and the identifier of this tensor algebra has a finite degree that is equivalent to the construction of a group cohomology of a discrete torsion-free group that is isomorphic to a finite lattice structure, generated by integers. This is also known as the matroidal structure. Hence, there exists a maximum and a minimum element in this finite lattice. Consequently, a greedy algorithm would be sufficient to construct the basis of this group, see Edmonds [23], and Welsh [100]. Here, searching for the longest chain, according to the probability law, identifies the maximum expected total discounted reward. Based on this information, we can introduce the following lemma:

Lemma

There is an optimal activation policy for an ensemble of interacting projects, such that this policy would be:

Without loss of generality, set the starting time for all activities to be synchronized, and put all projects into a candidate list.

Step-1: Order projects from maximal Nash index to minimal Nash index. Find the maximal Nash index project; allocate it for the selected projects set; then discard it from the candidate list; and repeat until you have chosen the necessary number of projects, k, if more than k number of projects are available.

Otherwise, select all of the projects for an ensemble and activate them, according to their index-based ordering.

Step-2: Find the minimum stopping time of these selected projects, and set the new starting time for the next coming ensemble of projects to this minimum stopping time.

Step-3: Repeat, until all the projects are terminated.

Proof of the Lemma

Since we want to generate the maximum expected discounted collective rewards, one can see that the above algorithm generates an optimal selection by using an interchange argument. That is, if you change the order of activation between any early projects with any of the later ones, then it is impossible to exceed the original maximal expected total reward, due to a time discount. Hence, this sequence of projects are bundled into an ensemble of k-projects, by preserving the original optimal ordering and activating them together initially; and then the projects in the ensembles are activated by their optimal order, as soon as an activation time is available. Therefore, activating each ensemble does not change the original optimal activation sequence for individual projects; hence changing the order of activation of these projects would result a reduction in the total expected discounted reward. The existence of the maximal total expected discounted reward is provided by the finite integer lattice structure [6], but the maximal element may not be unique. In the case of multiple maximal elements, the first optimal policy that is identified would be sufficient, since we do not want to allocate our time to search for all possible optimal policies.

7 Some Remarks

In this work an optimization method for sequentially choosing some from many interacting projects has been considered. The multi-armed bandit problem is the predecessor of our work. Another approach would be to construct a multivariate stochastic differential equations model and solve it based on empirical boundary conditions that satisfy certain model-filtration assumptions, see Karatzas and Shreve [52]. Moreover, competition or cooperation is a realistic form of interactions among projects. Hence, we have incorporated a multiplicative influence factor, expressing competition or cooperation among projects, in the collective discounted reward structure, by following the work of Nash [63]. On the other hand, our model assumes perfect information for sequentially activating these ensembles of projects, which may not be a realistic assumption. A model that would incorporate the missing information may be a better approach, see Kumar and Varaiya [58]. Or, a restless bandit model that differs from others by letting the changing of the state of un-activated bandits as well as changing states of the activated ones might be considered, see Glazebrook [37], and Glazebrook et al. [38]. The algorithm presented here is an optimal selection policy among the ensemble of influential projects, by choosing the highest expected discounted total reward, that provides:

a) a sequence of ensemble of experiments with their activation times, and

b) how long they will be active.

The computational complexity of Nash indices is discouraging at best. If you set the multiplicative influence factor to be one, then the influences among projects will disappear, and consequently the decision problem can be approached by the Gittins index method.

There are computationally efficient algorithms designed to compute Gittins indices; see Katehakis and Veinott [53]. Also, see Varaiya et al. [94] for algorithms that are suitable for practically computing the discounted expected rewards. Another approach is to construct bounds for the optimal expected discounted reward by utilizing a myopic policy, or a greedy algorithm, see Edmonds and Karp [24], such that it selects an ensemble of projects that will provide the maximum collective influential reward at a decision time, without paying any attention to the future alternatives, until all the project ensembles are scheduled to be activated. This myopic selection would provide a lower bound to the optimal reward. Also, when we set the discount factor to be equal to one, the total influential reward is independent of the projects activation time sequence, and we apply the myopic policy to schedule activation of the ensemble of projects, then the total nondiscounted influential reward achieved by this policy would provide an upper bound for the optimal total discounted influential reward. These bounds will help in reducing the number of computations for the expected revenues when the number of projects gets larger, and this may be useful to build a dynamic programming approach to activate the ensemble of projects, see Gürsoy [42].

The above approach provides an optimal schedule for activating the collection of projects, one bundle at a time, and can be generalized further if one wishes to introduce different discount factors for different projects, such as different interest rates, see Brown and Smith [13]. This approach is suitable to handle flexible cases when there is a change in the original set of projects, such as abandoning some projects for future activations, or considering new projects for future activations, by restarting the optimization algorithm at the present time.

In this work, the optimization algorithm was constructed based on the expected discounted values. However, averages may not be suitable for making decisions in the rare event case. Thus, a performance measure that is different than the average value may be necessary. Such an approach to an optimal policy that maximizes the probability of collecting the best reward could be constructed by using stochastic programming and minimizing regret, see Prékopa [77]. Finally, you may see Rusmevichientong and Tsiklis [84] for a linearization approach of the multi-armed bandit problem.

Advances in the machine learning have renewed interest in the multi-armed bandit problem. For some of the interesting recent research please see [20], [14], and [19].

Therefore, the design of a mixture of experiments turns out to be a stochastic optimization problem.

References

[1] A. Adem and R. J. Milgram. *Cohomology of finite groups.* Springer-Verlag, New York, 1994.

[2] P. Alexandroff. *Elementary Concepts of Topology.* Dover, New York, 1961.

[3] P. S. Alexandrov. *Combinatorial topology.* Dover, New York, 1998.

[4] V. I. Arnold. Smooth function statistics. *Func. Anal. Other Math.*, 1:111–118, 2006.

[5] V. I. Arnold. *Experimental mathematics.* MSRI/Amer. Math. Soc., USA, 2015.

[6] E. Artin. *Galois theory.* Dover Publications, New York, 1998.

[7] M. Atiyah. *The geometry and physics of knots.* Cambridge University Press, Great Britain, 1990.

[8] M. Atiyah and I. G. Macdonald. *Introduction to commutative algebra.* Addison-Wesley, Menlo Park, 1969.

[9] D. Benson. *Polynomial invariants of finite groups.* Cambridge University Press, London, 1993.

[10] D. Blackwell and T. S. Ferguson. The big match. *Annals of Mathematical Statistics*, 39:159–163, 1968.

[11] B. Bollobas. *Random Graphs.* Cambridge University Press, Cambridge, England, 2001.

[12] M. Borodzik, A. Nemethi, and A. Ranicki. Morse theory for manifilds with boundary. *Algebraic and Geometric Topology*, 16:971–1023, 2016.

[13] D. B. Brown and J. E. Smith. Optimal sequential exploration: bandits, clairvoyants and wildcats. *Oper. Res.*, 61:644–665, 2013.

[14] A. N. Burnetas, O. Kanavetas, and M. N. Katehakis. Asymptotically optimal multi-armed bandit policies under a cost constraint. *Probability in the Engineering and Information Sciences*, 31(3):284–316, 2017.

[15] H. Cartan. *Formes différentielles.* Hermann, Paris, 1967.

[16] H. Cartan and S. Eilenberg. *Homological algebra.* Princeton University Press, Princeton, 1956.

[17] Y. S. Chow, H. E. Robbins, and D. Siegmund. *Great expectations: The theory of optimal stopping.* Houghton Mifflin, Boston, 1971.

[18] P. M. Cohen. *Algebra.* Springer, New York, 1989.

[19] W. Cowan, Y. Honda, and M. N. Katehakis. Normal bandits of unknown means and variances: asymptotic optimality, finite horizon regret bounds, and solution to an open problem. *Journal of Machine Learning Research (JMLR)*, 18:1–18, 2018.

[20] W. Cowan and M. N. Katehakis. Multi-armed bandits under general depreciation and commitmentt. *Probability in the Engineering and Information Sciences*, 29(1):51–76, 2015.

[21] R. Descartes. *Discourse de la méthode*. Ian Maire, Leiden, 1637.

[22] H. F. Dodge and H. G. Romig. A method of sampling inspections. *Bell Systems Technical Journal*, 8:613–631, 1929.

[23] J. Edmonds. Submodular functions, matroids and certain polyhedra. In *Combinatorial structures and their applications*, New York, 1970. Gordon and Beach.

[24] J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. In *Combinatorial structures and their applications*, New York, 1970. Gordon and Beach.

[25] D. Eisenbud. *Commutative algebra*. Springer-Verlag, Berlin, 1995.

[26] R. J. Elliot, L. Aggoun, and J. B. Moore. *Hidden Markov Models*. Springer-Verlag, New York, 1995.

[27] P. Erdös. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.

[28] P. Erdös and A. Rényi. On the evolution of random graphs. *Bull. Inst. Inter. Statistics*, 38:343–347, 1961.

[29] Euclid. *The Elements: Volumes 1, 2, 3*. Dover, New York, 1956.

[30] K. J. Falconer. Projections of random Cantor sets. *J. Theor. Prob.*, 2:65–70, 1989.

[31] H. Federer. *Geometric Measure Theory*. Springer-Verlag, New York, 1969.

[32] A. K. Garsia and D. Stanton. Group actions of Stanley-Reisner rings and invariants of permutation groups. *Advences in Mathematics*, 51:107–201, 1984.

[33] J. C. Gittins. Bandit processes and dynamic allocation indices. *Journal of Royal Statistics Society*, B:148–177, 1979.

[34] J. C. Gittins. *Multi-armed bandit allocation indices*. Wiley, Chichester, 1989.

[35] J. C. Gittins. Indices on thin ice. In F.P. Kelly, editor, *Probability, Statistics and Optimization*. John Wiley, 1994.

[36] J. C. Gittins and D. M. Jones. A dynamic allocation index for sequential design of experiments. In *Colloquia Mathematica Societatis Janos Bolyai*. Hungary, 1972.

[37] K. D. Glazebrook. Indices for families of competing Markov decision processes with influences. *Annals of Applied Probability*, 3:1013–1032, 1993.

[38] K. D. Glazebrook, D. J. Hodge, and C. Kirkbridge. Monotone policies and indexability for bidirectional restless bandits. *Adv. in App. Probability*, 45:57–91, 2013.

[39] H.Ö. Gülcür, G. Akman, and K. Gürsoy. Extended Kalman filter design for time-varying system identification: An adaptive controller. *IFAC Transactions*, 1:452–473, 1982.

[40] P. Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In W.M. McEneany, G. Yin, and Q. Zhang, editors, *Stochastic Analysis, Control, Optimization and Applications: A Volume to Honor W.H. Fleming*, pages 547–566. Birkhauser, Boston, 1998.

[41] K. Gürsoy. Stochastic matroids. Unpublished, 1994.

[42] K. Gürsoy. *Branch and Bound Methods for Sequentially Choosing Some Among Several Competing Projects*. PhD thesis, Rutgers University, New Jersey, 1997.

[43] K. Gürsoy and M. N. Katehakis. On maximizing the availability of two component series systems in discrete time. *American Journal of Mathematical and Management Sciences*, 23:61–73, 2003.

[44] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, Cambridge, second edition, 1994.

[45] R. Hartshorne. *Foundations of projective geometry*. Benjamin, New York, 1967.

[46] R. Hartshorne. *Algebraic geometry*. Springer-Verlag, New York, 1977.

[47] M. Hochester and J. A. Eagon. Cohen-Macaulay rings, invariant theory, and the generic perfection of determinantal loci. *Americal Journal of Mathematics*, 93:1020–1058, 1971.

[48] K. Itô and H. P. McKean. *Diffusion processes and their sample paths*. Springer-Verlag, second edition, 1974.

[49] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. John Wiley, New York, 2000.

[50] F. V. Jensen. *Bayesian networks and decision graphs*. Springer-Verlag, New York, 1956.

[51] O. Kallenberg. *Random measures.* Akademie-Verlag, Berlin, 1975.

[52] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.* Springer-Verlag, New York, second edition, 1996.

[53] M. N. Katehakis and A. F. Veinott Jr. The multiarmed bandit problem: Decomposition and computation. *Mathematics of Operations Research*, 12(2):262–268, 1987.

[54] F. P. Kelly. *Reversibility and stochastic networks.* Wiley, Chichester, 1979.

[55] D .G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and Shape Theory.* Wiley, New York, 1999.

[56] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Ann. of Math. Statistics*, 23:462–466, 1952.

[57] A. N. Kolmogorov. *Foundations of the theory of probability.* Chelsea, New York, 1933.

[58] P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification and Adaptive Control.* Prentice Hall, 1986.

[59] T. L. Lai and H. E. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, pages 4–22, 1985.

[60] H. Lebesgue. *Measure and the integral.* Holden-Day, San Fransisco, 1966.

[61] P. Matilla. *Geometry of sets and measures in Euclidean spaces.* Cambridge University Press, Great Britain, 1999.

[62] A. R. Nahmod. The nonlinear Shrödinger equation on tori: Integrating harmonic analysis, geometry, and probability. *Bull. of the Amer. Math. Soc.*, 53:51–85, 2016.

[63] P. Nash. A generalized bandit problem. *Journal of Royal Statistical Society*, B–42:165–169, 1980.

[64] P. Nash and J. C. Gittins. A Hamiltonian approach to optimal stochastic resource allocation. *Adv. in App. Probability*, 9:55–68, 1977.

[65] J. Neyman. Two breakthroughs in the theory of statistical decision making. *Rev. Inter. Stat. Inst.*, 30:11–27, 1962.

[66] J. Pearl. *Causality.* Cambridge University Press, New York, 2000.

[67] M. Penrose. *Random Geometric Graphs.* Oxford University Press, New York, 2003.

[68] H. Poincaré. Analysis situs. *Journal de l'École Polytechnique*, 1:1–123, 1895.

[69] H. Poincaré. The relations of analysis and mathematical physics. *Bull. Amer. Math. Soc.*, 4:247–255, 1898.

[70] H. Poincaré. *La science et l'hypothèse.* Flammarion, Paris, 1902.

[71] H. Poincaré. *La valeur de la science.* Flammarion, Paris, 1905.

[72] H. Poincaré. *La science et la méthod.* Flammarion, Paris, 1908.

[73] A. Prékopa. On stochastic set functions i. *Acta Math. Acad. Sci. Hung.*, 7:215–263, 1956.

[74] A. Prékopa. On stochastic set functions ii. *Acta Math. Acad. Sci. Hung.*, 8:337–374, 1957.

[75] A. Prékopa. On stochastic set functions iii. *Acta Math. Acad. Sci. Hung.*, 9:375–400, 1958.

[76] A. Prékopa. Estimation of cause-effect relationship under noise. *Journal of Applied Probability*, 31:343–350, 1994.

[77] A. Prékopa. *Stochastic programming.* Kluwer Academic Press, 1995.

[78] M. Rams and K. Simon. The dimension of projections of fractal percolations. *J. of Stat. Phys.*, 154:633–655, 2014.

[79] F. P. Ramsey. On a problem of formal logic. *Proc. London Math. Soc.*, 30:264–286, 1930.

[80] H. E. Robbins. Some aspects of the sequential design of experiments. *Bull. of Amer. Math. Soc.*, 5:527–535, 1952.

[81] H. E. Robbins. An empirical Bayes approach. *Proc. Third Berkeley Symp. Math. Stat. Prob.*, 1:157–163, 1956.

[82] H. E. Robbins and S. Monroe. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[83] H. E. Robbins and D. Siegmund. A convergence theorem for nonnegative almost supermartingales and some applications. In *Optimization methods in statistics*, pages 233–257, New York, 1971. Academic Press.

[84] P. Rusmevichientong and J. N. Tsiklis. Linearly parametrized bandits. *Math. of Oper. Res.*, 35:395–411, 2014.

[85] B. J. Schmid. *Finite groups and invariant theory.* Springer-Verlag, Berlin, 1991.

[86] L. Schwartz. *Geometry and probability in Banach spaces.* Springer-Verlag, Berlin, 1981.

[87] L. Schwartz. *Semimartingales and their stochastic calculus on manifolds.* Universite de Montreal, Montreal, 1984.

[88] K. Seip. *Interpolation and sampling in spaces of analytic functions.* American Mathematical Society, Rhode Island, 2004.

[89] S. Shelah. Zero-One laws for graphs with edge probabilities decaying with distance. Unpublished, 1996.

[90] I. M. Singer and J. A. Thorpe. *Lecture Notes on Elementary Topology and Geometry.* Scott, Foresman Co., Illinois, 1967.

[91] J. Spencer. *The strange logic of random graphs.* Springer, New York, 2003.

[92] R. P. Stanley. *Combinatorics and Commutative Algebra, 2nd edition, V. 41 of Progress in Mathematics.* Birkhäuser Inc., Boston, 1996.

[93] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrica*, 25:285–294, 1933.

[94] P. P. Varaiya, J. C. Walrand, and C. Büyükkoc. Extensions of the multiarmed bandit problem: the discounted case. *IEEE Transactions on Automatic Control*, AC30:426–439, 1985.

[95] A. Wald. A new formula for the index of cost of living. *Econometrica*, 7:319–331, 1939.

[96] A. Wald. On a statistical generalization of metric spaces. *Proceedings of the National Academy of Sciences*, 29:196–197, 1943.

[97] A. Wald. *Sequential analysis.* John Wiley, New York, 1947.

[98] A. Wald. *Statistical decision function.* John Wiley, New York, 1950.

[99] L. Wasserman. Topological data analysis. *Annual Review of Statistics and Its Applications*, 5:501–532, 2018.

[100] D. Welsh. *Matroid theory.* Academic Press, London, 1976.

[101] P. Whittle. Arm acquiring bandits. *Annals of Probability*, 9(2):284–292, 1981.

[102] P. Whittle. Restless bandits: activity allocation in changing world. *Journal of Applied Probability*, 25:287–298, 1988.