BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 46, Number 4, October 2009, Pages 691–696
S 0273-0979(09)01262-2
Article electronically published on April 27, 2009

Probabilistic symmetries and invariance principles, by Olav Kallenberg, Probability and its Applications, Springer, New York, 2005, xii+510 pp., US$94.00, ISBN 978-0387-25115-8

This book takes an in-depth look at one of the places where probability and group theory meet. On the surface, probability (the mathematics of randomness) and group theory (the mathematics of symmetry) seem at opposite poles. The present account of sets of probability measures invariant under a group shows that there is a beautiful interface.

It is easiest to explain the subject by considering its first important development, de Finetti's theorem on exchangeability. Let $\mathcal{X}$ be the set of infinite binary sequences endowed with its usual product structure. This is the probabilist's classical coin-tossing space since a typical point $x = 011001\ldots$ can be thought of as a mathematical model of flips of a coin. Let P be a probability measure on $\mathcal{X}$. For example, if $0 \leq \theta \leq 1$ is fixed, $P = P_\theta$ can be specified by assigning to the cylinder set $(x_1, x_2, \ldots, x_n, * * * \ldots)$ the measure $\theta^j(1-\theta)^{n-j}$ with $j = x_1 + x_2 + \cdots$. This is the measure corresponding to "flip a coin independently with probability of heads θ". A measure P is *exchangeable* if it is invariant under permuting coordinates. That is to say, for every $n,\ x_1, x_2, \ldots, x_n$ and all permutations $\sigma \in S_n$ (the symmetric group on n lettters)

$$P(x_1, x_2, \ldots, x_n) = P(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}).$$

Thus $P(10) = P(01),\ P(001) = P(010) = P(100)$, etc., where we write $P(10)$ for the probability assigned to the cylinder set $\{x : x_1 = 1, x_2 = 0\}$.

Observe that P_θ is exchangeable and that the exchangeable probabilities form a convex set. De Finetti's Theorem identifies the extreme points of this convex set and shows that every exchangeable probability is a unique barycenter of the extreme points.

Theorem (de Finetti's Theorem for zero/one sequences)**.** *Let P be an exchangeable probability measure on coin tossing space. Then there exists a unique probability measure μ on $[0,1]$ such that*

$$P = \int_0^1 P_\theta \mu(d\theta).$$

This theorem was first proved by Bruno de Finetti as a contribution to the philosophy underlying probability and statistics. Some of this background is given in Section 2 below. The statement is so simple and elegant that it is natural to seek generalizations. What about, e.g., 3-values or Polish space values? What about more general groups or semi-groups? For example, what are the extreme points of the set of probability measures on $\mathbb{R}^\infty$ that are invariant under the orthogonal group O_n for all n? These and related extensions are the main subject matter of Olav Kallenberg's book.

2000 *Mathematics Subject Classification.* Primary 60G09.

The book presents a deep, mathematically careful amalgamation of a large literature (372 references directly related to exchangeability). It is a model of scholarship. The proofs are readable and complete. There is some new material due to the author and much that will appear new because the author has polished up a hard-to-locate gem. In this review, I will try to explain why exchangeability is interesting, give some of its applications, and explain some of the author's major contributions.

1. Background

De Finetti was a major contributor to the foundations of probability and statistics. He suggested a basic reinterpretation of things such as, "What does it mean to say that the chance of heads in a coin flip is about $\frac{1}{2}$?" The *mathematics* of coin tossing has been well understood since Pascal and Fermat in 1650. One introduces a mathematical model for n repeated flips of a coin; the chance of the outcome $x_1, x_2, \ldots, x_n$ is $\theta^j(1-\theta)^{n-j}$ with $j = x_1 + x_2 + \cdots + x_n$. Bernoulli asked the inverse question: If we observe j heads out of n trials, what can we say about θ? Bayes and Laplace postulated an *a priori* distribution $\mu(d\theta)$ which quantifies what is known about θ. Then, the chance of observing $x_1, x_2, \ldots, x_n$ is

$$\int_0^1 \theta^j (1-\theta)^{n-j} \mu(d\theta). \tag{1}$$

All of this begs the question, "What on earth is θ?" De Finetti, along with Ramsey and later Savage took a very different view of the basics. They were willing to assign probabilities to *observables*, such as the next n flips of a coin. They were less than happy assigning probabilities to unobservable abstractions such as θ. Thus, for de Finetti, $P(x_1, x_2, \ldots, x_n)$ represents a person's subjective probability for the next n tosses. This is something to be determined by previous experience and introspection. The question now arises, What is the connection between this subjective interpretation and the Bayes–Laplace formulation (1)?

De Finetti's Theorem shows that they are equivalent. Moreover, just from the assumption that P is symmetric, the mathematics build a parameter space $[0,1]$, a parameterized family of measures P_θ, and a prior distribution $\mu(d\theta)$. It is a remarkable theorem.

De Finetti extended the theorem to finite-valued and then $\mathbb{R}^d$-valued observations. Hewitt and Savage extended it to fairly general topological spaces via an early version of Choquet theory. Here is a modern version. Let P be a probability on a countable product of a Polish space $\mathcal{A}$. Suppose P is exchangeable, $P(A_1 \times A_2 \times \cdots \times A_n) = P(A_{\sigma(1)} \times A_{\sigma(2)} \times \cdots \times A_{\sigma(n)})$ for all n, $A_1, A_2, \ldots, A_n$ in $\mathcal{A}$ and σ. Then there is a unique probability μ on $\mathcal{P}(\mathcal{A})$ (the set of probability measures on $\mathcal{A}$) such that

$$P(A_1 \times A_2 \times \cdots \times A_n) = \int_{\mathcal{P}(\mathcal{A})} \prod \theta(A_i)\mu(d\theta). \tag{2}$$

In the language of convex sets and Choquet theory, the set of exchangeable measures is a convex simplex with extreme points the product measures. The formula (2) shows that every exchangeable P is a unique mixture (barycenter) of extreme points. There is a small but healthy subject, nonparametric Bayesian statistics, which constructs natural measures μ on $\mathcal{P}(\mathcal{A})$ and shows how to work with them. An introduction to the subject of "measures for measures" is in Ghosh–Ramamoorthi [6].

The nonparametric version requires integrating over a huge space. Suppose the basic space is $\mathbb{R}$; it is natural to ask what additional symmetry assumptions are required to get down to the basic models of a statistician's toolkit, the normal distribution (or other standard families). The first theorems here are due to David Freedman. Suppose P is a probability measure on $\mathbb{R}^\infty$ which is orthogonally invariant: $P(A_1 \times A_2 \times \cdots \times A_n) = P(\Gamma(A_1 \times \cdots \times A_n))$ for all n, all intervals A_i and all $\Gamma \in O_n$. Then there exists a unique probability measure μ on $[0, \infty)$ with

$$P(A_1 \times \cdots \times A_n) = \int_0^\infty \prod \Phi_\sigma(A_i)\mu(d\sigma). \tag{3}$$

In (3), $\Phi_\sigma(A) = \frac{1}{\sigma\sqrt{2\pi}} \int_A e^{-t^2/2\sigma^2} dt$ is the usual scale family of Gaussian distributions. One gets a mixture of location-scale families by restricting to invariance under the subgroup of O_n fixing the line from 0 to $(1, 1, \ldots, 1)$. Orthogonal invariance, in many variations, is a theme running through the book.

A third theme, contractability, is also simple to illustrate. Ryll and Nardzewski showed that P is exchangeable if and only if any subsequence of n coordinates has the same measure as the first n. Thus the theory extends to measures invariant under a semi-group, such as removing an infinite subsequence; these are called "contractable". This is varied and developed in surprising ways in the book under review.

2. Higher dimensional arrays

A major theme of the book is de Finetti's Theorem for *arrays* where invariance of only rows and columns are postulated. Thus one has a measure on, say, zero/one matrices which is invariant as

$$P(x_{ij}; 1 \le i, j \le n) = P(x_{\sigma(i)\tau(j)}; 1 \le i, j \le n)$$

for all n, all $x_{ij} \in \{0, 1\}$ and σ, τ permutations in S_n. These problems arose in Bayesian analysis of variance. It was expected that there would be a slight extension of the representation (1) with perhaps a row parameter for each i and a column parameter for each j. The story turned out differently. To explain, let me give a probabilistic description of a way of producing a row/column exchangeable array. Begin with an *arbitrary* Borel function $\varphi : [0, 1]^2 \to [0, 1]$. Pick independent uniform random variables $U_i, V_j,\ 1 \le i, j \le n$ on $[0, 1]$. To make the (i, j) entry of the array, flip a $\varphi(U_i, V_j)$ coin, writing one or zero as it comes up heads or tails. The construction is evidently symmetric and extendable to infinite arrays. Call it a φ-process and write P_φ for the associated measure. The probabilist David Aldous and logician Douglas Hoover independently proved that any row/column exchangeable P is a mixture of φ-processes

$$P = \int P_\varphi \mu(d\varphi).$$

The uniqueness of μ was a difficult problem. After all, if φ is transformed by a measure-preserving transformation in each variable, the same distribution is induced. Hoover produced a model-theoretic proof for a correct version of uniqueness. The probabilists struggled to find their own proof. This was finally achieved by Kallenberg; a full report is in Section 7.6 of the book under review. We note that clarifying uniqueness was only a small part of Kallenberg's work. His main result is to extend the Aldus–Hoover theorem to the contractable case.

These theoretical investigations had unexpected applications. We mention two briefly: work on the psychology of vision and work on graph limits.

The psychology of vision. The perception psychologist Bela Julesz studied what properties of a pattern made the difference between foreground and background. When are patterns visually distinguishable? Clearly, if one pattern is denser than another (more ink on the page), the eye will see the difference. Even if the density is the same, if one pattern is random and another is clumpy or clustered, the eye sees the difference. A long series of experiments seemed to show that the eye mainly sees these first- and second-order statistics. Abstracting, he conjectured that if two patterns had the same first- and second-order statistics, the eye would find them visually indistinguishable.

In one more careful version he generated stochastic arrays $X_{ij},\ 1 \leq i, j \leq n$. Two arrays have the same kth order statistics if $P(X_{ij} = x_{ij}, (i,j) \in s) = P(Y_{ij} = x_{ij}, (i,j) \in s)$ for all s with $|s| = k$. A host of probabilists and engineers had tried and failed to produce counterexamples to the Julesz conjecture: if two row/column exchangeable arrays have the same first- and second-order statistics, they will be visually indistinguishable.

I heard about this problem while David Freedman and I were trying (and failing) to prove the Aldous–Hoover theorem. We knew about φ-processes and could show they were extreme points. We could not show there were no other extreme points. Consider the following two processes. The first is fair coin tossing (thus $\varphi(x, y) \equiv \frac{1}{2}$). The second has φ as pictured:

0	1
1	0

The second process has a simple description: fill out the first row of an array by flipping a fair coin. For the other rows, flip a fair coin once; if it comes up heads, copy the original top row. If it comes up tails, copy the opposite of the original top row (mod 2). This process has the same first-, second-, and third-order statistics as fair coin tossing. By its construction, it results in a stripey pattern, clearly distinguishable from coin tossing. This is shown in Figure 1.

When we told Julesz, he had a wonderful reaction: "Thank you. For twenty years Bell Labs has paid me to study the Julesz conjecture. Now they will pay me for another twenty years to understand why it is wrong." The story goes on; see [4] for references and details.

Graph limits. A very recent appearance of the Aldous–Hoover theorem comes from the emerging theory of graph limits. Laszlo Lovasz, with many coauthors, has been developing a limiting theory for large dense graphs to answer questions such as When are two large graphs close? and When does a sequence of graphs converge? A very appealing theory using subgraph counts has emerged. This unified things such as Erdös–Renyi random graphs, quasi-random graphs, graph testing, Szemerédi's regularity lemma, extremal graph theory, and much else. The subject is in rapid growth, but the surveys by Borgs et al. [1] are most useful.

It turns out that the natural limit of a sequence of graphs is no longer a graph, but rather a symmetric function $\varphi : [0,1]^2 \to [0,1]$. One gets a realization of the

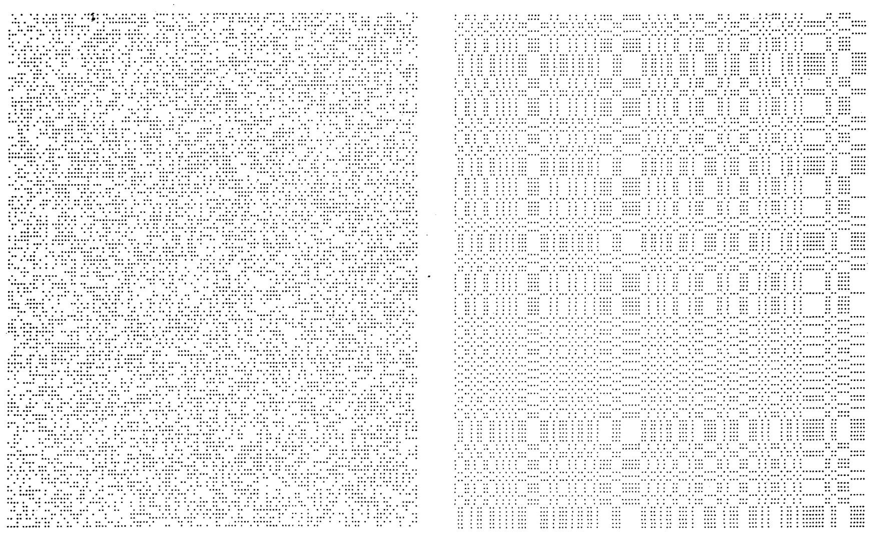

FIGURE 1. A counterexample to the Julesz conjecture.

limiting object by choosing a sequence $U_i,\ 1 \leq i < \infty$, uniformly and independently in $[0,1]$, forming an infinite symmetric zero/one matrix by flipping a $\varphi(U_i, V_j)$ coin for $i < j$, and taking the resulting symmetric matrix as the adjacency matrix of a graph. Of course, if $\varphi(x,y) \equiv p$ one gets Erdös–Renyi random graphs, but every convergent sequence of graphs converges to a mixture of such φ-graphs.

The evident parallel between graph limits and the Aldous–Hoover theorem was made explicit in joint work with Svante Janson [5]. The graph theorists' many variations (weights on vertices and edges, bipartite, directed or hypergraphs) are all special cases of the exchangeable theory. Of course, the graph theorists bring fresh questions and new tools and results. The mix is being worked out as you read this!

The above stories motivate the study of exchangeable arrays. The book also develops versions of the theorems for arrays invariant under the orthogonal group, and versions undertaking subsequences of the rows and columns. Random arrays can be replaced by random functions $f(x,y)$ with, say, x, y in $[0,1]$. There are versions for higher dimensional arrays. One of my favorite extensions, due to Kallenberg, considers processes indexed by C_2^∞, the infinite hypercube consisting of infinite binary sequences which terminate with all zeros. The symmetry group G of C_2^∞ consists of the semi-direct product of S_∞ and C_2^∞. One may ask for the structure of random processes $\{X_x\}_{X \in C_2^\infty}$ which are invariant under G. This includes many of the previous problems by looking at various "slices". Kallenberg gives an elegant parameterization, analogous to the original Aldous–Hoover theorem.

3. Summary

The book contains many gems. One striking example concerns extensions of Bob Connelly's game of "Say Red" [2]. In the original, an ordinary deck of cards is well shuffled, the cards are turned up one at a time and you are permitted to say "red" at any time. If the next card turned up is red, you win \$1. If black, you lose \$1. You must say red sometime. If you say red before any cards are turned up, your chance of winning is $\frac{1}{2}$. What is the expected value under the optimal strategy? If you have not seen this before, you will be surprised to learn that the expectation is $\frac{1}{2}$ for any strategy. What if you can make a sequence of, say, five dollar-sized bets on red at times of your choosing as the cards are turned up? What if you can choose the amounts bet (subject to the total bet being \$5)? The reader will have to look in Chapter 5 for this as well as sweeping generalizations to decoupling inequalities. These extend the classical identities of Wald. Putting all of these useful, slightly magical identities into a unified framework, bringing them to their natural level of generality, and relating them to the world of exchangeability is a superb contribution.

The book offers complete coverage of the topics included. Surveys of other points of view and other extensions of de Finetti's basic theorem to partial exchangeability can be found in articles in the same issue as [5] as well as in [3]. These extensions treat problems such as "When is a probability measure a mixture of Markov chains?" and "How can mixtures of other standard families (e.g., Poisson or exponential distributions or Wishart matrices) be characterized?" The general theory translates these questions into the language of projective limits and leans on statisticians' versions of the D-L-R equations of statistical physics. While these extensions are not treated in the present book, it does contain a good set of references.

Kallenberg has written a definitive, wonderful account of the topics treated. The care, clarity, depth, and scholarship are truly admirable.

References

1. C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), no. 6, 1801–1851. MR2455626
2. Robert Connelly, *Say red*, Pallbearers Review **9** (1974), 702.
3. P. Diaconis and D. Freedman, *Partial exchangeability and sufficiency*, Statistics: applications and new directions (Calcutta, 1981), Indian Statist. Inst., Calcutta, 1984, pp. 205–236. MR786142 (86i:60097)
4. Persi Diaconis and David Freedman, *On the statistics of vision: the Julesz conjecture*, J. Math. Psych. **24** (1981), no. 2, 112–138. MR640207 (83e:92071)
5. Persi Diaconis and Svante Janson, *Graph limits and exchangeable random graphs*, Rend. Mat. Appl. (7) **28** (2008), no. 1, 33–61. MR2463439
6. J. K. Ghosh and R. V. Ramamoorthi, *Bayesian Nonparametrics*, Springer Series in Statistics, Springer-Verlag, New York, 2003. MR1992245 (2004g:62004)

Persi Diaconis
Stanford University