Ergodic theory of random transformations, by Yuri Kifer. Progress in Probability and Statistics, vol. 10, Birkhäuser, Boston, Basel, Stuttgart, 1986, 210 pp., $34.00. ISBN 0-8176-3319-7

Traditionally ergodic theory has been the qualitative study of iterates of an individual transformation, of a one-parameter flow of transformations (such as that obtained from the solutions of an autonomous ordinary differential equation), and more generally of a group of transformations of some state space. Usually ergodic theory denotes that part of the theory obtained by considering a measure on the state space which is invariant or quasi-invariant under the group of transformations. However in 1945 Ulam and von Neumann pointed out the need to consider a more general situation when one applies in turn different transformations chosen at random from some space of transformations. Considerations along these lines have applications in the theory of products of random matrices [2, 3], random Schrödinger operators [2], stochastic flows on manifolds [6], and differentiable dynamical systems.

Mathematically the set up is as follows. Let M be a space, $\mathscr{B}$ a σ-algebra of subsets of M and let $\mathscr{T}$ be a collection of measurable transformations of M into M. For example, if M is a topological space we could choose $\mathscr{T}$ to be the space, $C(M, M)$, of all continuous transformations of M into M, and if M is a smooth manifold we could take $\mathscr{T}$ to the space, $D(M, M)$ of all smooth transformations of M into M. Suppose $\mathscr{T}$ is equipped with a σ-algebra so that the map $(f, x) \to f(x)$ of $\mathscr{T} \times M \to M$ is measurable. Let m be a probability measure on $\mathscr{T}$. We want to study the action on M of compositions of elements of $\mathscr{T}$ chosen independently with distribution m. So consider the direct product space $\Omega = \mathscr{T}^{\mathbb{N}}$ equipped with the direct product measure $p = m^{\mathbb{N}}$, where $\mathbb{N}$ denotes the natural members. The elements of Ω are sequences $w = (w_1, w_2, w_3, \dots)$ of members of $\mathscr{T}$. There is a natural transformation, $S: \Omega \to \Omega$, of Ω called the shift map and defined by $S((w_1, w_2, w_3, \dots)) = (w_2, w_3, \dots)$. The shift preserves the probability p (*i.e.* $p(S^{-1}A) = p(A)$ for every measurable subset A of Ω) and p is ergodic for S (*i.e.* if A is a measurable subset of Ω and $S^{-1}A = A$ then $p(A) = 0$ or 1). Consider the skew-product transformation $T: \Omega \times M \to \Omega \times M$ defined by $T(w, x) = (Sw, w_1(x))$ where $w = (w_1, w_2, \dots) \in \Omega$ and $x \in M$. Iterating gives $T^n(w, x) = (S^n w, w_n \circ w_{n-1} \circ \cdots \circ w_1(x))$ for $n \geq 1$, and the second coordinate gives the action of the randomly chosen maps on M.

This induces on M a discrete-time Markov Process with the probability, $P(x, B)$, of moving from the point $x \in M$ to a point of the measurable subset $B \subset M$ in one unit of time given by $P(x, B) = m(\{f \in \mathscr{T} \mid f(x) \in B\})$. For some applications, such as stochastic stability of diffeomorphisms [5], it seems more natural to consider certain Markov processes on M rather than actions by random maps, so one should consider which Markov

processes can arise from random maps as above. If M is a Borel subset of a complete separable metric space then every Markov process is induced as above from a probability measure m on the family of all measurable maps of M into M. However if M is a topological space one has to put restrictions on M and the transition probabilities $P(x, \cdot)$ if one wants the Markov process to be induced, as above, from a probability on $C(M, M)$. When M is a smooth manifold there seem to be no results giving conditions on $P(x, B)$ that ensure it is induced from a probability on $D(M, M)$. However many different probabilities m on $\mathscr{T}$ can induce the same Markov process on M and these different probabilities give rise to random actions on M with vastly differing dynamical behaviour. By focussing attention on random actions one can define concepts such as entropy and Lyapunov exponents and so generalise the theory known for iterates of a single transformation. One can get much more detailed dynamical information from a random action on M than from a Markov process on M. This is because one can compare the orbits on M of different points under the random action. The random action on M, determined by m, can be viewed as follows. Starting from $x \in M$ one chooses $w_1 \in \mathscr{T}$ at random and moves from x to $w_1(x)$, then one chooses $w_2 \in \mathscr{T}$ at random and moves to $w_2 \circ w_1(x)$, etc. If a certain system were modelled by a single transformation $f: M \rightarrow M$ one could obtain a more realistic model by considering the random action on M determined by a probability on $\mathscr{T}$ concentrated on maps which are perturbations of f. This is because conditions of a repeated process differ slightly at each stage, and also because one may not know exactly the rules of the evolution of the process. Random actions are also models of numerical simulations of real systems because of round off approximations. They also arise when diffusion processes are considered as solutions of stochastic differential equations.

Let us return now to the study of $T: \Omega \times M \rightarrow \Omega \times M$, bearing in mind that we have fixed a probability measure $p = m^{\mathbb{N}}$ on Ω. It is natural to consider T-invariant probabilities μ on $\Omega \times M$ (i.e. $\mu(T^{-1}C) = \mu(C)$ for every measurable subset C of $\Omega \times M$) which project to p (i.e. if $\pi_1: \Omega \times M \rightarrow \Omega$ is the projection onto the first factor then $\mu \circ \pi_1^{-1} = p$) and apply the known ergodic theory of measure-preserving transformations to T and read off what this means for the action on M. Because we wish to consider the behaviour on M of the random action it is best to consider only those measures μ on $\Omega \times M$ of the form $p \times \eta$ where η is a probability on M. Ohno has shown that $p \times \eta$ is T-invariant iff $P^*\eta = \eta$ where P^* acts on probabilities on M by $(P^*\nu)(B) = \int_{\mathscr{T}} \nu(f^{-1}B) dm(f)$. Therefore $P^*\eta = \eta$ means $\eta(B) = \int_{\mathscr{T}} \eta(f^{-1}B) dm(f) \ \forall B \in \mathscr{B}$, so this does not require m-almost all f to preserve η but rather that the m-average of f's preserves η. Also $p \times \eta$ is ergodic for T iff η is P^*-ergodic in the sense that if $B \in \mathscr{B}$ has the property that $\eta(B \Delta f^{-1}B) = 0$ for m-almost $f \in \mathscr{T}$ then $\eta(B) = 0$ for 1. When applying theorems of ergodic theory to T one can sometimes strengthen the conclusion to obtain a nonrandom result by applying the following simple proposition. Let $P^*\eta = \eta$ and let $h \circ T = h$ $(p \times \eta)$-a.e. where $h: \Omega \times M \rightarrow R$ is bounded and measurable. If η has a barycentric decomposition as an integral over the P^*-invariant

and ergodic probabilities (which it always will if M is a Borel subset of a complete separable metric space) then $h(w, x)$ depends only on x $(p \times \eta)$-a.e. (*i.e.* $h(w, x) = \int h(w, x) dm(w)$ $(p \times \eta)$-a.e.).

The random version of Kingman's subadditive ergodic theorem is the following. Let η be a probability on M with $P^*\eta = \eta$ and let $h_n: \Omega \times M \to R$, $n \geq 1$, be a sequence of measurable functions with $\max(0, h_1) \equiv h_1^+ \in L^1(p \times \eta)$ and $h_{k+n} \leq h_k + h_n \circ T^k$ $(p \times \eta)$-a.e. $\forall n, k \geq 1$. If η can be written as a barycentric integral of ergodic P^*-invariant probabilities then $\lim_{n \to \infty} h_n/n = h$ $(p \times \eta)$-a.e. where $h: M \to R$ is a function on M with $h^+ \in L^1(\eta)$ and $h(fx) = h(x)$ for $(m \times \eta)$-almost all (f, x). If η is ergodic then h is constant η a.e. The fact that h is independent of points of Ω comes from the proposition mentioned above. A random version of Birkhoff's ergodic theorem can easily be deduced by taking h_n to be $\sum_{i=0}^{n-1} h_1 \circ T^i$.

The measure-theoretic entropy of the random action can be viewed as a special case of the relative entropy of skew-product transformations, due to Abramov and Rohlin, by considering the shift $S: \Omega \to \Omega$ as a factor measure-preserving transformation of $T: \Omega \times M \to \Omega \times M$ [1]. Therefore if η is a probability measure on M with $P^*\eta = \eta$ (so that $p \times \eta$ is T-invariant) and if ξ is a finite partition of M we can put

$$h(m, \eta, \xi) = \lim_{n \to \infty} \frac{1}{n} \int H_\eta \left(\bigvee_{i=0}^{n-1} w_1^{-1} \circ \cdots \circ w_i^{-1} \xi \right) dp(w),$$

where $H_\eta(\xi)$ denotes the usual entropy of the finite partition ξ using the probability η and the sumbol $\bigvee$ denotes the join of partitions. Then the entropy $h(m, \eta)$ of the random action, determined by m, on M equipped with the probability η is defined as the supremum of $h(m, \eta, \xi)$ over the space of all finite partitions of M. Versions of the Kolmogorov-Sinai theorem, of the formula for entropy of powers, and of the Shannon-McMillan-Brieman theorem can be proved, the latter result requiring the stronger hypothesis that m-almost all f preserve η.

Consider now the topological version of the situation. Let M be a compact metric space and let m be a probability on the space $C(M, M)$ of all continuous transformations of M into M. As before let $\Omega = C(M, M)^{\mathbb{N}}$, $p = m^{\mathbb{N}}$, let $S: \Omega \to \Omega$ be the shift map, and let $T: \Omega \times M \to \Omega \times M$ be defined by $T(w, x) = (S(w), w_1(x))$ where $w = (w_1, w_2, \dots) \in \Omega$. The notions of topological entropy and topological pressure can be defined for the random action determined by m, and they are special cases of relative topological entropy and pressure obtained from viewing $S: \Omega \to \Omega$ as a factor of T [7, 8]. They can be defined using open covers or separated sets or spanning sets. If α is an open cover of M then

$$h(m, \alpha) = \lim_{n \to \infty} \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} w_1^{-1} \circ \cdots \circ w_i^{-1} \alpha \right)$$

exists for p-almost all w and is constant p-a.e. Here the symbol $\bigvee$ denotes the join of open covers and if β is an open cover then $H(\beta)$ is the natural logarithm of the minimal number of members of β needed to cover M.

Then the topological entropy, $h(m)$, of the random action determined by m is the supremum of $h(m, \alpha)$ over all open covers of M. One has $h(m) \geq \sup\{h(m, \eta) | P^*\eta = \eta\}$ and this may be a strict inequality; of course $h(m)$ is the supremum of the relative measure-theoretic entropies over all T-invariant measures μ on $\Omega \times M$ that project to p on Ω (not just ones of the form $p \times \eta$) [7].

Now consider Oseledec's theorem, which has motivated the exciting advances in smooth ergodic theory in recent years and has led to a deeper understanding of the long-term behaviour of the iterates of a single diffeomorphism arousing hope that these ideas may explain complicated behaviour in real systems (such as turbulence) [4]. Oseledec's theorem can be studied in a purely measure-theoretic context but let us consider it only for the situation of diffeomorphisms. So now M is a compact manifold and m is a probability on the Borel subsets of the space $D(M, M)$ of C^1 diffeomorphisms of M equipped with the C^1-topology. As before $\Omega = M^{\mathbb{N}}$, $p = m^{\mathbb{N}}$, $S: \Omega \to \Omega$ denotes the shift and $T: \Omega \times M \to \Omega \times M$ is defined by $T(w, x) = (Sw, w_1(x))$, where $w = (w_1, w_2, \dots) \in \Omega$. Each $f \in D(M, M)$ has a tangent map Tf defined on the tangent bundle TM of M and we denote by $T_x f$ the restriction of Tf to the tangent space, $T_x M$, to M at $x \in M$. Hence $T_x f: T_x M \to T_{f(x)} M$ is a linear map for each $x \in M$.

Applying Oseledec's theorem to the measure-preserving transformation $T: \Omega \times M \to \Omega \times M$ gives the following, where the norm sign $\| \cdot \|$ comes from any Riemannian metric chosen on M. Let η be a probability on M with $P^*\eta = \eta$ and assume η is P^*-ergodic. Let $\iint \log^+ (\|T_x f\|) dm(f) d\eta(x) < \infty$. There is a Borel set $Y_\eta \subset \Omega \times M$ with $(p \times \eta)(Y_\eta) = 1$, an integer $s \geq 0$ and constants $-\infty < \alpha^{(s)} < \cdots < \alpha^{(0)} < \infty$ with the following properties. For every $(w, x) \in Y_\eta$ there are subspaces of $T_x M$, $0 \subset V_{(w,x)}^{(s)} \subset \cdots \subset V_{(w,x)}^{(0)} = T_x M$, such that $\forall v \in V_{(w,x)}^{(i)} \setminus V_{(w,x)}^{(i+1)}$ we have

$$\lim_{n \to \infty} \frac{1}{n} \log \|(Tw_n) \circ (Tw_{n-1}) \circ \cdots \circ (Tw_1)(v)\| = \alpha^{(i)},$$

where we define $V_{(w,x)}^{(i)}$ to be $\{0\}$ if $i \geq s + 1$. One has $T_x V_{(w,x)}^{(i)} \subset V_{T(w,x)}^{(i)}$. The numbers $\alpha^{(i)}$ are called the Lyapunov characteristic exponents with respect to m and η and $m^{(i)} = \dim V_{(w,x)}^{(i)} - \dim V_{(w,x)}^{(i+1)}$ is constant $(p \times \eta)$-a.e. and is called the multiplicity of $\alpha^{(i)}$. The subspaces $V_{(w,x)}^{(i)}$ depend only measurably on (w, x), and they are the only objects, in the conclusion of the theorem, that depend on w. Can one obtain a filtration of η-almost every $T_x M$ that is independent of w and retains most of the properties of the filtration $\{V_{(w,x)}^{(i)}\}$? Kifer has proved that one can. Suppose that we have the stronger assumption $\iint (\log^+ \|T_x f\| + \log^+ \|T_x f^{-1}\|) dm(f) d\eta(x) < \infty$. Then there is a Borel subset M_η of M with $\eta(M_\eta) = 1$, an integer $r \geq 0$, constants $-\infty < \beta^{(r)} < \beta^{(r-1)} < \cdots < \beta^{(0)} < \infty$, and for each $x \in M_\eta$ a filtration of $T_x M$ by subspaces $0 \subset W_x^{(r)} \subset W_x^{(r-1)} \subset \cdots \subset W_x^{(0)} = T_x M$ such that

$$\forall v \in W_x^{(i)} \setminus W_x^{(i+1)} \lim_{n \to \infty} \frac{1}{n} \log \|(Tw_n) \circ \cdots \circ (Tw_1)(v)\| = \beta^{(i)}$$

for p-almost all $w \in \Omega$. Here we take $W_x^{(i)} = \{0\}$ if $i \geq r + 1$, and the set of w where this convergence holds can depend on x and v. For each i the dimension of $W_x^{(i)}$ does not depend on $x \in M_\eta$ and $W_x^{(i)}$ depends measurably on $x \in M_\eta$. We have $T_x f(W_x^{(i)}) = W_{f(x)}^{(i)}$ for $(m \times \eta)$-almost all (f, x). What is the relationship between these two versions of Oseledec's theorem? It is clear that $r \leq s$ and for each j there is an i_j with $\beta^{(j)} = \alpha^{(i_j)}$. We have $i_1 < i_2 < \cdots < i_r$. Also $W_x^{(j)} \subset V_{(w,x)}^{(i_j)}$ $(p \times \eta)$ a.e., and for η-almost all $x \in M$,

$$\{v \in T_x M | p(\{w | v \in V_{(w,x)}^{(i_{j+1}-1)} \setminus W_x^{(j)}\}) > 0\} = \varnothing.$$

Versions of Ruelle's inequality (entropy is bounded above by the sum of the positive exponents) and the stable manifold theorem are given in Kifer's book. Also these results can be given in the context of stochastic flows and this is very important in the study of stochastic differential equations. Kifer also presents further results about the bundles $\{W_x^{(i)}\}$ and the largest exponent.

All the results described above are presented in a very readable form in Kifer's book. The style and pace of the book makes the reading enjoyable and rewarding. There are the inevitable small slips of notation and wrong references to the bibliography, but my only quibble is that there is no index. Kifer has managed to overcome the problem of overbearing notation, and anyone with a basic knowledge of ergodic theory is equipped to read the book. This theory is still in its infancy and one can look forward to more exciting developments and applications.

REFERENCES

1. L. M. Abramov and V. A. Rohlin, *Entropy of a skew-product transformation with invariant measure*, Vestnik Leningrad Univ. **17** (1962), 5–13; English Transl. Amer. Math. Soc. Transl. (2) **48** (1966), 255–265.

2. P. Bougerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, Progress in Probability and Statistics, vol. 8, Birkhäuser, 1985.

3. J. C. Cohen, H. Kester and C. M. Newan (eds.), *Random matrices and their applications*, Contemporary Math., vol. 50, American Math. Soc., Providence, R.I., 1984.

4. J.-P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Rev. Modern Phys. **57** (1985), 617–656.

5. Y. Kifer, *General random perturbations of hyperbolic and expanding transformations*, J. Analyse Math. **47** (1966), 111–150.

6. H. Kunita, *Stochastic differential equations and stochastic flows of diffeomorphisms*, Lecture Notes in Math., vol. 1097, Springer-Verlag, Berlin and New York, 1984, pp. 143–303.

7. F. Ledrappier and P. Walters, *A relativised variational principle for continuous transformations*, J. London Math. Soc. **16** (1977), 568–576.

8. P. Walters, *Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts*, Trans. Amer. Math. Soc. **296** (1986), 1–31.

PETER WALTERS
UNIVERSITY OF WARWICK