DISCUSSION ON PROFESSOR ORNSTEIN'S PAPER[1]

PROFESSOR PAUL C. SHIELDS (*Stanford University*). Professor Ornstein's paper describes some of the properties of Bernoulli processes. In particular he shows that the B-processes are precisely the closure of the mixing n-step Markov processes in the $\bar{d}$-metric, and that entropy is a complete invariant for B-processes. In this and elsewhere it has been shown that factors and roots of Bernoulli shifts are Bernoulli and that Bernoulli shifts are embeddable in flows.

If we take larger classes of automorphisms much pathology can occur (K-automorphisms may not have roots, the entropy 0 factor may not be a direct factor). We propose here a definition of regularity which may include all that is of physical interest and excludes much of this pathology. We say that a class **R** of automorphisms is ***regular*** if it satisfies

(R1) **R** is closed in the $\bar{d}$-metric.

(R2) Each transformation in **R** is embeddable in a flow.

(R3) If $T \in \mathbf{R}$, any factor of T is in **R**.

(R4) Any transformation in **R** of positive but not completely positive entropy is the direct product of an entropy zero factor and a K-automorphism.

The class of Bernoulli shifts is a regular class. The class of rotations of the circle is also a regular class, for one can easily show that any two distinct rotations are at least $\frac{1}{2}$ apart in the $\bar{d}$-metric. (For this class (R2) and (R4) are trivial, while Adler [1] has established (R3)). Of some physical interest is the closure $\mathbf{R}_M$ of the n-step Markov processes in the $\bar{d}$-metric. An n-step Markov process is the product of a Bernoulli shift and a finite rotation [2]. If two such processes are close enough in the $\bar{d}$-metric their rotation factors must coincide. Hence the class $\mathbf{R}_M$ is just the class of direct products of Bernoulli shifts with finite rotations and is therefore regular.

It would be of much interest to know whether some further classes of interest are regular. Among these are the following:

(a) The class of products of rotations of the circle and Bernoulli shifts. (This clearly satisfies (R1) (R2) and (R4)).

(b) The class of affine maps of a torus, or some other group of interest.

Simple physical descriptions of such classes (such as that of $\mathbf{R}_M$ as the $\bar{d}$-closure of the n-step Markov processes) as well as further properties of such classes need investigating (e.g., does (R2) imply (R4)?).

REFERENCES

[1] ADLER, R. (1964). Invariant and reducing subalgebras of measure preserving transformations. *Trans. Amer. Math. Soc.* **110** 350–360.

[2] ADLER, R., SHIELDS, P. and SMORODINSKY, M. (1972). Irreducible Markov shifts. *Ann. Math. Statist.* **43** 1027–1029.

---

[1] Professor Ornstein's paper was presented as a Special Invited Paper on June 20, 1972, at the Western Regional meeting of the Institute of Mathematical Statistics in Seattle, Washington. Professors P. C. Shields and R. M. Blumenthal were invited discussants at this meeting. The wider discussion contained herein was made possible by distributing advance copies of Professor Ornstein's manuscript to a group of interested persons. The Editor greatly appreciates Professor Ornstein's willing assistance in this as well as the interesting responses by the discussants.

PROFESSOR R. M. BLUMENTHAL (*University of Washington*). Let T be a measure-preserving invertible transformation on $(X, \mathcal{B}, P)$ and let p be a partition of X, that is $p: X \to R$ is $\mathcal{B}$ measurable and takes on the values $1, 2, \cdots, k$. Let p_T be the mapping $p_T(x) = \{p(T^i x)\}_{-\infty<i<\infty}$ from X, to sequence space, Σ, and let μ be the distribution, $\mu = p_T P$, in Σ of p_T; that is $\mu(A) = P(p_T^{-1}(A))$. It might be interesting to render Ornstein's definition and theorems entirely in terms of the induced measures μ since this is a viewpoint common in probability theory. I will make a few remarks on the very interesting metric $\bar{d}$ as a metric on measures μ on Σ.

The induced measure μ is of course stationary, that is invariant under the shift $(\tau x)_i = x_{i+1}$. Let μ and ν be stationary probability measures, ergodic or not, on Σ and let (μ, ν) denote the set of τ-invariant measures on Σ^2, $(\tau(x, y) = (\tau x, \tau y))$ with μ and ν for marginals. Define $\bar{d}(\mu, \nu) = \inf\{\theta(\{(x, y) \mid x_0 \neq y_0\})\}$, the infimum being over all measures θ in (μ, ν). This is Ornstein's metric $\bar{d}$ except that he writes it as a distance between pairs (T, p) and (S, q) rather than as a distance between the induced measures $p_T P$ and $q_S P$. If μ and ν are ergodic (relative to τ) then Ornstein puts in the additional requirement that only the ergodic measures in (μ, ν) be used in forming the infimum. But this is unimportant (as Ornstein points out in the appendix) since one can pass to the conditional measures, given the τ-invariant field, and these are ergodic always.

In case μ and ν are ergodic the definition of $\bar{d}$ can be rewritten: given such a μ call a point x in Σ a μ-sequence if for every set $A \subset \Sigma$ which depends on only finitely many coordinates,

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} I_A(\tau^i x) = \mu(A).$$

According to the ergodic theorem almost all (μ) x in Σ are μ-sequences. Given 2 points x, y in Σ set $\delta(x, y) = \limsup (1/n) \sum_{i=0}^{n-1} I(x_i \neq y_i)$. Then $\bar{d}(\mu, \nu) = \inf \delta(x, y)$, the infimum being over all x, y which are μ- and ν-sequences respectively. This highlights $\bar{d}(\mu, \nu)$ as the least fraction of time the μ printouts and ν printouts can differ. The definition of $\bar{d}$ can be generalized by letting Σ be any compact metric space, τ be any bimeasurable mapping from Σ to Σ (perhaps the identity), defining (μ, ν) as before and then setting $\bar{d}(\mu, \nu) = \inf\{\theta(f(x) \neq f(y)) \mid \theta \text{ in } (\mu, \nu)\}$ where $f: \Sigma \to R$ is measurable. Clearly $\bar{d}$ is nonnegative and symmetric. The triangle inequality is easy to check, but of course it is possible to have $\bar{d}(\mu, \nu) = 0$ for $\mu \neq \nu$. I do not know if this extension is at all useful.

Returning to the stationary measures on sequence space it is clear that the $\bar{d}$ topology is stronger than the ordinary vague topology of measures on the compact Σ. As one reflection of this write the entropy $e(\mu)$ of the stationary measure μ as

$$e(\mu) = \lim_n \frac{-1}{n} \int \log p_\mu(x_0, \cdots, x_{n-1}) \mu(dx)$$

where

$$p_\mu(a_0, \cdots, a_j) = \mu(\{x \mid x_0 = a_0, \cdots, x_j = a_j\}).$$

Standard theorems on entropy show that $e(\mu)$ is a continuous function of μ in the $\bar{d}$ topology, whereas it is only a lower semicontinuous function of μ in the vague topology.

PROFESSOR KONRAD JACOBS (*University of Erlangen, Germany*). *1.* The main results mentioned in the paper seem to be

(a) the introduction of the metric $\bar{d}$ and the closedness of the classes of B-resp. K-processes.

(b) Theorem 3.

(c) the joint result of Ornstein and Shields about the abundance of K-processes that are not B.

2. The Ornstein-Shields theorem seems to be the result of efforts to prove an analogue to A. Ionescu-Tulcea's result (transformations not admitting an equivalent σ-finite invariant measure are the rule and not the exception in a category sense, proving that Ornstein's famous example was a "natural" one) for processes that are not B but K. Do Ornstein and Shields have any results of category character?

3. Looking at Theorem 3 and trying to find a picture it might fit into in some future theory, I see the ergodic processes as a boundary of the set of all processes, the K-processes as blunt corners and the B-processes as sharp corners where the vague topology coincides with the 'fine topology' (as known in boundary theory) $\bar{d}$. Probably one could construct an artificial Markov process tending to the desired part of the boundary by simply replacing all arbitrarities in Ornsteins' and Ornstein-Shields' constructions by proper random choices. However, I feel unable to judge whether this idea of linking Ornstein's paper to boundary theory is promising.

4. Concerning the relation between ergodic theory and probability in general, my opinion at present runs as follows: Ergodic theory should try to explain why mechanical systems like ideal gases fulfill certain probabilistic laws. To prove that the systems are not only K but even B, is always a considerable step in that direction, since B-systems fulfill *all* laws of probability. On the other hand, the effect of these laws to a concrete situation may be blocked by the fact that the isomorphies established are measurable only and not e.g. continuous. On the other hand, physicists might be content to see only *some* probabilistic laws fulfilled, and those only to a certain degree. The situation is quite parallel to that in the v. Mises-Kolmogorov-Martin-Löf-Schnorr theory. Martin-Löf shows that 0–1-sequences fulfilling *all* probabilistic laws are abundant but not constructible by machines; Schnorr is about to classify constructible sequences according to the number of probabilistic laws and the degree to which they fulfill them. I could imagine that an exchange of thoughts with Schnorr could lead Ornstein to a class of systems and isomorphies which really justify the existence of what is called statistical physics.

PROFESSOR U. KRENGEL (*University of Gottingen, Germany*). Let

$$X = \{X_i, -\infty < i < \infty\} \quad \text{and} \quad Y = \{Y_i, -\infty < i < \infty\}$$

be two sequences of independent identically distributed random variables, where X_i takes the values ± 1 and ± 2 with probability $\frac{1}{4}$ each, and Y_i takes the value 0 with probability $\frac{1}{2}$ and the values 1, 2, 3, and 4 with probability $\frac{1}{8}$. Obviously these processes are probabilistically very different. E.g. X gives rise to a recurrent random walk and Y to a transient random walk. Yet, ergodic theorists say that these processes are the "same." In fact, they may write down a complicated transition matrix of a stationary Markov chain and say that this is again the "same" process. Does not it follow that their concept of "same" (the concept "isomorphic") has no probabilistic relevance? Yes and no! For the majority of probabilistic questions (e.g limit-distributions) isomorphy is meaningless. Still, there are several problems, which are solved or clarified using ergodic theoretic tools, and, in particular, entropy theory. Among these are coding problems and problems on tail-σ-algebras.

Historically, ergodic theory was an essential tool in information theory during its initial development. Then, starting with Kolmogorov's work on entropy and culminating with the fascinating work of Ornstein, concepts and theorems from information theory were used to solve outstanding problems in ergodic theory. Reading the present paper I got the feeling that the pendulum may now swing back again: For this, Ornstein's new and apparently quite natural and useful concept of distance of two processes may be a beginning. So far, coding in information theory usually means block-coding, i.e. coding of words of a fixed length. One requires that after the coding, the transmission through the channel and the decoding, one gets with probability $1 - \epsilon$ the completely correct block. If an error is made there is no difference between a 100% wrong decoded message and a 1% wrong decoded message. To me it would be more natural to code an infinite string of letters in a stationary way, and such that the probability that the ith encoded letter differs from the ith decoded letter is small, i.e. the encoded process is close in the Ornstein-metric to the decoded process. I would expect that this distance can be made arbitrarily small, if the capacity of the channel exceeds the entropy of the source. (This does not follow from the coding theorems because block coding is not stationary.) It would be interesting to know how close one can get if the capacity is only 99% of the entropy of the source. Note that in this case, classical coding theory just tells us that the information cannot be transmitted.

Ornstein's isomorphy constructions involve the arbitrarily distant future and the arbitrarily distant past. In information theory this is of course not allowed. Since the "ideal" coding necessary to obtain isomorphy can be approximated by coding procedures depending on a finite past and a finite future only, this does not matter (except for a delay) if one allows a small distance of the decoded process from the source-process.

It is clear, however, that constructions that depend only on the past would be more desirable since they preserve more of the structure of a process. This may be quite useful: For example, I have used past-dependent isomorphy constructions in *J. Math. Anal. Appl.* **35** (1971), to show that for continuous time (in contrast to the case of discrete time) there exist stationary processes with finitely many states that are deterministic and for which the reversed process is completely nondeterministic. I am confident that past-dependent constructions would have other probabilistic implications. Therefore it would be nice if some of Ornstein's results could be sharpened in this direction.

It should be mentioned that J. P. Conze, University of Paris, has already made significant contributions towards the problems suggested by Ornstein in Section 4. In particular he has proved that n-parameter Bernoulli-shifts with the same entropy are isomorphic. Some introductory work of Conze on this subject is contained in his paper: "Entropie d'un groupe abélien de transformations."

PROFESSOR W. KRIEGER (*Ohio State University*). This paper in a way advocates a program. A stationary process is viewed as a consistent assignment of probabilities to finite sequences of symbols. Then invariants that are usually studied in ergodic theory are associated with this assignment and the classification problem is posed. The scope of this program is considerable. Indeed, one has associated with such an assignment of probabilities not only the measure-preserving transformation (automorphism), but also, e.g. the endomorphism on the space of one-sided infinite sequences. One can proceed to consider the invariants of this endomorphism. The methods that are used in the study of the invariants of endomorphisms are somewhat different from the methods used for automorphisms. E.g. there are invariants for endomorphisms that are invariants of partitions and sequences of partitions that the endomorphism produces.

At the end of the paper, it was hinted that an extension of the theory beyond the first dimension is possible. Indeed, it is now known that the finite generator theorem and the isomorphism theorem for Bernoulli schemes with finite entropy hold for a large class of countable groups.

PROFESSOR G. MARUYAMA (*Tokyo University of Education, Japan*). As is stressed in Professor Ornstein's paper, the class of stationary processes is equivalent to that of measure preserving transformations. It is therefore natural that we can draw from Ornstein's theory conclusions which throw light on classical unsolved problems for stationary processes.

Among the class of Gaussian stationary processes, only characteristics effectively used for distinguishing between metrical types of generated automorphisms are spectral measures. If the spectral measures of the given two Gaussian processes are equivalent, i.e. absolutely continuous with respect to each other, then they generate isomorphic automorphisms. In its essence this criterion is based on the linear theory of Gaussian stationary processes. Suppose we are given two discrete-time Gaussian processes with spectral measures having densities relative to Lebesgue measure and suppose further that they are perpendicular, i.e. they have separate supports. Are the generated automorphisms isomorphic? The above criterion can not answer this question. However, Ornstein's theory, (reference [2] of his paper) provides a satisfactory answer that they are isomorphic as factors of a Bernoulli process (with infinite entropy).

There is another important application of his theory to the longstanding question: Under what conditions is it possible that a real stationary process $\{x_n\}_{-\infty}^{\infty}$ may be put in the form $z_n = f(\cdots, \theta^n\xi_{-1}, \theta^n\xi_0, \theta^n\xi_1, \cdots)$, a function of independent identically distributed ξ_n, associated with shifts $\theta^n\xi_k = \xi_{n+k}$? The problem is intimately related to Ornstein's theory. According to [8], only the regularity of $\{x_n\}_{-\infty}^{\infty}$, which is the same as saying that it is a K-process, is insufficient for this specified representability. Natural sufficient conditions that single out the class of weak Bernoulli processes have been introduced in [3]. Historically, various mixing conditions for stationary processes were proposed. Ornstein's condition is stronger than Rosenblatt's but weaker than Ibragimov's.

In view of the constructive examples of Markov processes or prediction theory of Gaussian stationary processes, it is likely that under suitable conditions the x_n-process should admit the more restrictive representation that

$$x_n = f(\cdots, \theta^n\xi_{-1}, \theta^n\xi_0).$$

This problem initiated by P. Lévy, owes much to Rosenblatt's fundamental work but remains unsolved. Needless to say, it belongs to the same domain as the isomorphism problem. Its complete solution will, more or less, require refinement of existing techniques by Ornstein.

PROFESSOR H. TOTOKI (*Kyoto University, Japan*). As a relevant item, I would like to point out the problem of isomorphism of stationary processes with the time sense preserved. Namely, let $(\mathcal{P}, T)$ and $(\mathcal{Q}, S)$ be stationary processes; then we ask under what circumstances is there an isomorphism ψ such that $\psi T = S\psi$ and $\psi(\bigvee_{-\infty}^{0} T^n\mathcal{P}) = \bigvee_{-\infty}^{0} T^n\mathcal{Q}$? Another equivalent formulation of the problem is the isomorphism problem of non-invertible measure-preserving transformations.

When $(\mathcal{P}, T)$ and $(\mathcal{Q}, S)$ are both independent processes, it is easy to see that they are isomorphic in the above sense iff the distributions of $\mathcal{P}$ and $\mathcal{Q}$ are of the same type. If $(\mathcal{Q}, S)$ is an independent process, then the problem is nothing but the representation problem of stationary processes by independent processes (cf. [1] Problem 1(*b*)). M. Rosenblatt ([1], Theorem 1) has given a condition for a Markov process $(\mathcal{P}, T)$ with finite $\mathcal{P}$ to be isomorphic (in the above sense) to an independent process. His condition can be naturally extended to the case when $\mathcal{P}$ is a countably infinite partition.

Along this line, we can also see that the continued-fraction transformation (itself) is not a Bernoulli transformation and a β-expansion transformation (itself) is Bernoulli iff β is an integer.

The results mentioned above will be published in [2].

REFERENCES

[1] M. ROSENBLATT (1959). Stationary processes as shifts of functions of independent random variables. *J. Mash. Mech.* **8** 665–681.

[2] KUBO, I. MURATA H. and TOTOKI, H. (1972). On the isomorphism problem for endomorphisms of Lebesgue spaces. (to appear.)

PROFESSOR BENJAMIN WEISS (*The Weizmann Institute of Science, Israel*). For brevity's sake I will confine myself to reporting on some new developments apropos Appendix 4. In collaboration with Yitzhak Katnelson much of the theory, sketched in the paper under discussion, has been extended to families of commuting transformations. The basic result may be described as follows. Let Z^d denote the d-dimensional integer lattice, $l \in Z^d$, $l = (l_1, \cdots, l_d)$. Suppose that to each $l \in Z^d$ is associated a measure preserving transformation, written ϕ^l, such that $\phi^l\phi^{l'} = \phi^{l+l'}$; such a group will be denoted by Φ. Two groups Φ_i, $i = 1, 2$, are isomorphic if there is an invertible measure preserving transformation θ from the space X_1, where Φ_1 acts, to X_2 such that $\phi_2{}^l\theta = \theta\phi_i{}^l$ for all l. A group Φ is Bernoulli if there is a partition α such that $\phi^l\alpha$ generates the measure algebra and the family $\{\phi^l\alpha\}$ is independent. An entropy can be introduced, so that the entropy of a Bernoulli-group is the same as the entropy of α, its independent generator.

THEOREM. *Two Bernoulli groups with the same entropy are isomorphic.*

It is not difficult to extend the notion of finitely determined partitions to groups—and the theorem can be extended to such processes. It would be of interest to determine whether or not the states that arise in statistical mechanics are finitely determined or not. Recent work of F. Spitzer and others have shown that the Gibbsian distributions are intimately connected with a kind of Markovian multi-parameter process, and one would expect that they are indeed finitely determined. In the same direction it would be of interest to find suitable "finitary" processes that approximate multi-parameter processes like the n-step Markov chains approximate the usual stochastic processes.

*   *   *   *   *

I wish to thank each of the discussants for their efforts and valuable suggestions. I feel that many promising directions for future research have been pointed out in this discussion and of course the possibility of future research is what gives life to a subject.

DONALD ORNSTEIN