7 August 1959, Volume 130, Number 3371 SCIENCE

Stochastic Models of Population Dynamics

A few simple chance mechanisms may combine to reproduce many manifestations of a complex phenomenon.

Jerzy Neyman and Elizabeth L. Scott

Before proceeding to the main subject of this article, it may be useful to indicate briefly the meaning which we ascribe to the terms *model* and, in particular, *stochastic model.* The question is, what do we mean when we say that we build a stochastic model of a given class of phenomena?

In considering this question, it is necessary to remember that, ordinarily, phenomena of interest are not open to immediate observation in all of their details. Most frequently, we observe only a few external manifestations of the phenomena. The figures on births and deaths, employment and unemployment, and so forth, collected by the Bureau of the Census exhibit several manifestations of the complex phenomenon designated by the phrase "the dynamics of the population of the United States." Again, the observations of Thomas Park represent an interesting manifestation of a complex phenomenon of struggle for existence. These observations refer to two noninterbreeding and nonaggressive species of flour beetles. If these species are forced to share the same habitat, then one, not always the same, invariably becomes extinct. On the other hand, and this is another manifestation of the same phenomenon, each species, if kept separately in the same kind of habitat, is capable of maintaining a considerable population, the size of which fluctuates about a point of apparent equilibrium.

In each of the two cases, the dynamics of the population of the United States and the struggle for existence of flour beetles, there are a variety of observations yielding a number of different manifestations of a complex phenomenon. However, the complete picture of the phenomenon remains hidden.

Another element to bear in mind while approaching the term *model* is the existence, for each of us, of a certain category of phenomena, or mechanisms, which, partly through experience and partly through imagination, appear to us familiar and, therefore, simple. Although, in some cases, this simplicity is only apparent and our familiarity not very deep, we are inclined to believe the contrary and to think in terms of these familiar mechanisms. For this particular reason, we shall call these mechanisms "elementary" mechanisms.

A classical example of an elementary chance mechanism, or chance phenomenon, is the stability of the frequency of "heads" in a series of tosses of a coin. Although not many of us have done any extensive coin tossing in our life, such experience as we do have, combined with imagination, make this particular mechanism a very familiar one.

With these two prerequisite concepts, (i) the essentially hidden complex phenomena of which we observe only certain manifestations and (ii) simple mechanisms, which are broadly familiar through experience and imagination, we may now explain the sense in which we use the terms *mathematical model* and *stochastic model.*

When we speak of a mathematical model $M(\Pi)$ of a class of phenomena Π we mean a system of mathematically expressed hypotheses representing Π as a particular combination of several elementary mechanisms. These mechanisms may be either deterministic or indeterministic. If a particular mechanism implies that every variable of the system contemplated is a function of the others, then this mechanism is called deterministic. Otherwise, that is, if at least one of the variables is a chance or a random variable, then the mechanism is indeterministic. The word *stochastic* is used as a synonym of *indeterministic.*

As a consequence of the above explanations, a stochastic model of population dynamics means a system of mathematically expressed hypotheses, representing the development of a population as a particular combination of elementary mechanisms, at least some of which are chance mechanisms.

At this point one might wonder about a distinction between mechanisms which are and which are not "elementary." As we see it, the distinction is both simple and subjective. Mechanisms with which through personal experience and imagination we feel familiar are "elementary" to us but, perhaps, not to others. Also, there are, undoubtedly, a great number of chance mechanisms familiar to physicists to the point that they are able to think in terms of these mechanisms, which are entirely unknown to us. These mechanisms, then, are "elementary" for the physicists but not elementary for us. In general, the greater the number of elementary chance mechanisms that one has at his disposal, the easier it is to construct stochastic models of phenomena.

Both authors are on the staff of the University of California, Berkeley. Currently, Professor Neyman is in the Miller Institute for Basic Research in Science. This article is based on a paper presented on 29 Dec. 1958 at the Washington meeting of the AAAS, when the authors were awarded the Newcomb Cleveland Prize. It was presented at the invitation of the Society for General Systems Research, the principal aim of which is to encourage the development of theoretical systems applicable to more than one of the traditional departments of knowledge.

Purpose of Models

We see two distinct purposes of stochastic models, one esthetic and the other utilitarian.

When it happens that a number of apparently disconnected manifestations of a class of phenomena are logical consequences of a single simple combination of a few "elementary" mechanisms then, to most of us, this circumstance is a cause of distinct esthetic satisfaction. This satisfaction is particularly strong if, after studying the conceptual model, we suddenly notice an observable detail which thus far escaped the attention of the experimenter and on which the predictions of the theory appear to agree with empirical fact. Upon observing a circumstance of this kind we generally feel that we "understand" the development of the phenomena.

The utilitarian considerations regarding models come in when we are concerned, for practical purposes, with such manifestations of a particular class of phenomena which, momentarily, are not directly observable. At that moment, the availability of a model, whose consequences agree with a number of manifestations of the same phenomena already observed, is most convenient because it provides suggestions about what might be expected in the domain where the observations are still lacking. A case in point is the present discussion, in the scientific literature and the daily press, whether the increase in radioactivity caused by the tests of nuclear weapons is or is not going to increase the incidence of cancer among human beings. Cancer develops slowly, and direct empirical evidence will require years to accumulate. Thus, in order to have a hint right now about what might be the effects observable in the future, our only source is a theoretical model of the phenomenon of carcinogenesis. If one is constructed and if it stands up in comparisons with a number of different manifestations of the phenomenon already observed, then, and until direct observations of the effects of increased irradiation are available, the predictions of the model will form a natural basis for practical action.

Subject of This Article

In the sections that follow we attempt to give a brief account of three distinct stochastic models of dynamic population phenomena with which we have been concerned in our work. Details of two of these models have already been published, and our present purpose is to analyze the solutions obtained in terms of the concepts of "elementary mechanisms" mentioned at the outset. These two models are: a model (*1*) of struggle for existence between two nonaggressive, noninterbreeding species (*2*) and a model of clustering which appears to have interesting applications in a number of domains, from the temporal-spatial development of biological populations (*3*), through physics (*4*) to cosmology (*5*). The third model (*6*) is concerned with the temporal development of populations of cancer cells (clones) conceived as generated by a two-step, mutation-like process. Due to limitations of space, publication of this model is postponed.

Struggle for Existence

Thomas Park's observations on the two species of flour beetles, *Tribolium confusum* and *Tribolium castaneum* (denoted by him *b* and *c*, for short), extend over a very considerable set of experimental conditions. We are concerned here with one set of experiments, in a particular combination of temperature and humidity, where the species *b* and *c*, kept separately from each other, were found to be capable of maintaining relatively stable populations indefinitely. The numbers of adults, counted once a month, varied about averages which, themselves, varied from one replicate of the experiment to the next. For species *b* the averages varied from about 70 to 130, the grand average being about 110. For species *c* this variation was from 80 to 220, with the grand mean about 140. Thus, on the average, species *c* appeared to be more "viable" than *b*.

The interesting feature of the experiment was that in all vials containing a combined culture of both species *b* and *c*, initially in equal numbers, one of the species invariably died out completely. The further interesting detail is (i) that the dying-out species was not always the same and, particularly, (ii) that the more frequently dying-out species was species *c* which, in separation from the other, is capable of maintaining a larger population!

The basic question here was whether this surprising result of the experiment is indicative of some *biological* changes in the beetles of the two species induced by cohabitation or whether it is a natural outcome of the interplay of their biological functions exhibited equally when the two species live in isolation and when they share the same habitat. When speaking of biological changes, we have in mind possible changes in the fecundity, fertility, and duration of the several phases in the life cycle that might conceivably occur in one (or both) species, perhaps due to the contamination of the flour by the excreta of the other species, and the like.

Our contribution to the study was very modest. We simply put in formulae the facts found by Thomas Park and recounted to us. These facts are:

1) While the beetles of the two noninterbreeding species have no racial prejudices, and do not fight, they do consume each other's eggs and pupae. Also, it appears that in this particular respect there are again no prejudices and each of the species likes the eggs and the pupae of the other just as well as its own.

2) In general, no two females lay their eggs with the same abundance but, in addition, there appears to be a difference between the two species.

3) There appear to be variations in the egg and pupae eating both within and between species.

4) Similarly, there appears to be a within- and between-species variation in the duration of inactive forms, egg and pupae.

These four characteristics, observed for the two species reared in isolation from each other, were taken as the basis of the stochastic model constructed. For the sake of simplicity (in fact, we were not able to do otherwise) the model constructed is of the type called a discrete time model. Also, we ignored the distinction between sexes. The basic idea was to compute for each egg the chance, depending upon the number of beetles present in the vial, of surviving to hatch. Naturally, given a fixed number of predator forms, the longer the egg stage, the smaller the computable probability that the egg will hatch and, eventually, produce an adult.

The consequences of these hypotheses proved to be in qualitative agreement with the observations both of development in pure cultures of the two species and in competition. Specifically, with reference to pure cultures, it was shown that, in order to explain the time fluctuations of the number of adults about a point of apparent equilibrium, there is no need to postulate any *biological* changes due to crowding. The phenomenon is a simple consequence of the assumptions that the fecundity and fertility are random variables with fixed distributions, that the duration τ in the inactive states (egg and pupa) is a constant, that for each egg the probability of being eaten by a given beetle during a given period of time has a fixed value, independent of crowding, and that the fates of particular eggs and of particular beetles are mutually independent.

With reference to competition of two species, it was also shown that the invariable extinction of one of them does not require any assumption of biological changes induced by cohabitation or by crowding. In fact, the necessity of extinction of one of the two species is a consequence of the assumption that the two species in competition have the same characteristics as in separate cultures and that they differ in the value of the criterion

$$C = (\log \nu)/\tau$$

where ν stands for the global-fecundity, meaning the average number of adults generated by a female during her lifetime in the absence of cannibalism. According to the model, of the two competing species the predominant winner of competition is the one with the larger value of the criterion C.

The question arises whether the above success of the model *proves* that crowding and cohabitation induce no biological changes in the beetles. The answer is no. All that is shown is that the observations described do not constitute evidence for the existence of such changes. In fact, some further, more delicate experiments may show that because of crowding or because of cohabitation with another species, the biological properties of the beetles do change.

Further developments regarding the model may be of interest, and we wish to report the following two items. First, a paper by Amelia Polnik (*7*) showed that the factors of unequal fecundity and unequal duration in inactive states certainly play an important role in the competition of species. Second, according to a personal communication of Thomas Park, the original model of competition described above appears incomplete. In fact, further experiments with the two species of *Tribolium* revealed two genetic strains of one of the species which have practically identical values of the criterion C. According to the original model, the fates of these two strains in competition with the one strain of the other species should be the same. Yet, in actual experiments, one of the strains is a consistent winner and the other a consistent loser in competition with the same opponent. Thus, the two strains must differ in some important characteristic not involved in the original model. A preliminary inspection of experimental data already obtained suggested to Thomas Park that this relevant characteristic is the total duration of the life cycle "from egg to egg." The shorter this period, then, *caeteris paribus,* the better the chance of winning the competition. We are looking forward to the completion of this novel series of Park's experiments, which we expect to yield further important information.

Stochastic Process of Clustering

The second stochastic model of population dynamics to which we wish to call attention is the so-called stochastic process of clustering. The first idea of this process occurred to one of us (*8*) in 1939 in connection with the distribution of the number of insects in equalsized plots on the ground. Subsequently, this same chance mechanism proved useful during the war in the theory of bombing by formations of planes. After the war, we found that the same process, unexpectedly, has interesting applications in indeterministic cosmology (*9, 10*), concerned with the phenomenon of clustering of galaxies. Also, one of our colleagues, Robert R. Read, found interesting applications of the clustering process in the study of ionization in a cloud chamber caused by cosmic rays. Still more recently (*3*), armed with the same stochastic process of clustering, we returned to the study of biological populations. Finally, we now have hopes of obtaining some interesting results in applying the clustering process to carcinogenesis.

In addition to this multiplicity of applications, the process of clustering is very interesting because many of its properties still remain unexplored and because there is a broad field for mathematical generalizations.

The particular novel elements in the stochastic approach to the study of populations introduced by the process of clustering are two.

1) Practically all classical attempts to study chance mechanisms of the development of a population treat all the individuals forming the population, in a sense, uniformly. Striking examples of this "uniformity" are the various theories of epidemics. Here it is usually assumed that an increase in the number of infectious individuals increases equally the probability of becoming infected for every single susceptible in the population. As the relevant authors themselves point out, this equality in the increase of the probability of infection may be realistic only in very small communities, such as families. On the other hand, while a case of bubonic plague in a community does indeed increase the danger of infection with this disease everywhere, the increase in the household of the individual concerned must be presumed to be greater than that for a typical person living in a distant city.

The first novel element introduced by the stochastic process of clustering is that here the populations studied are not treated as aggregates of individuals all characterized by the same probabilities but as aggregates of groups or aggregates of clusters.

2) Another element of the classical approach to population dynamics is that the two aspects observed in the development of populations, namely (i) migration and (ii) changes in the numbers of individuals forming the population, are treated separately. The usual mathematical model of migration is the so-called diffusion process. The subject of study is the aggregate of chance motions either of one particle starting from a given point or of a number of such particles. The mathematical instruments of study of numerical changes in a population are the so-called multiplicative stochastic processes, concerned with variation in time of the number of individuals, under certain hypotheses about immigration, birth, and death. In other words, the classical diffusion processes are concerned with the time and space coordinates of some existing particles but ignore the phenomena of their birth and death, while multiplicative processes deal with births, deaths, and immigration but ignore the spatial relationships of the particles.

The second novel element introduced by the process of clustering is that in its treatment of populations it provides a unified theory of spatial as well as of numerical changes in the population. Specifically, for any preassigned system of "regions" in the habitat, say

$$R_1, R_2, \ldots, R_s$$

and for any epoch t, the stochastic process of clustering studies the joint distribution of $2s \times n$ variables

$$\gamma_{jk}(R_i, t)$$

with $i = 1,2, \ldots, s$; $j = 1,2$; and $k = 1,2, \ldots, n$. Here

$$\gamma_{1k}(R_i, t)$$

represents the number of male members

of the kth generation of the population which at time t are alive and are located within the region R_i. Further,

$$\gamma_{2k}(R_i, t)$$

represents the number of females of the same kth generation alive at time t and, at that time, located in R_i.

Thus the process of clustering represents a particular combination of two mechanisms that one might consider "elementary" (although each of them is rather difficult): the diffusion mechanism and the birth and death mechanism, both stochastic mechanisms. Since these two "elementary" mechanisms are still the subjects of intensive mathematical research, it must be obvious that the possibility of their combination is limited to considerable simplifications and that whatever results are obtained now have been obtained at some cost. There are several items in this cost that we sincerely regret. One is our inability to treat the problem other than by assuming that multiplication processes occur only at fixed intervals of time.

Postulates Underlying Clustering

After this somewhat long introduction, we enumerate briefly the postulates underlying the process of clustering. These postulates may be stated in terms of abstract "particles" which later can be identified with more or less arbitrary concrete entities, such as cosmic ray particles, galaxies, beetles, cancer cells, or even ideas. However, we expect a gain in intuitiveness by speaking directly of consecutive generations of some living species.

We begin by contemplating a habitat, essentially a Euclidean space of an arbitrary number of dimensions, possibly a plane. In this habitat, we introduce a system of coordinates and consider arbitrary regions $R_1, R_2, \ldots$, postulated to be Borel-measurable. The datum of the problem is the distribution law of "ancestors" of a population. Given an arbitrary finite system of regions

$$R_1, R_2 \ldots, R_s$$

this law must determine the joint probability distribution of the random variables $\alpha(R_i)$ and $\beta(R_i)$, $i = 1,2, \ldots, s$, representing the numbers of male and female ancestors which at time $t = 0$ are located in R_i. For example, at time $t = 0$ there may be just one female ancestor located at the origin of coordinates. Then

$$P\{\alpha(R_i) = 0\} = 1$$

for all i, while

$$P\{\beta(R_i) = 1\} = 1$$

only if the region R_i contains the origin of coordinates. Alternatively, it may be given that at time $t = 0$ the ancestors, both males and females, are Poisson-wise distributed over the habitat (or over a specified part of it), with a constant or a variable density, and so forth.

Further postulates are as follows.

1) *Postulate regarding births.* It is assumed that at each moment kT, where T is a fixed positive number and $k = 0,1, \ldots$ every then living female, whether of the ancestral or of any other generation, gives birth to a "litter" of progeny. The individuals born at moment kT form the kth "generation."

2) *Postulate regarding the distribution of litters.* The point at which a litter is born coincides with the location, at that time, of the mother female.

3) *Postulate regarding the size of litters.* The number of members of a litter is a random variable $\nu = 0,1,2, \ldots$, with an unspecified probability generating function $G_\nu(u)$ which may be assumed to depend on the generation of the mother female or on the location of the litter, or on both. However, it is assumed that the variables ν corresponding to different litters are mutually independent and independent of all other variables involved in the model.

This is one of the assumptions that we were forced to adopt with great regret. One of the notable features in the development of real populations is the possible dependence of the birth rate on the degree of local congestion of the population so that an important generalization needed in the theory of clustering is a device, seemingly easy but one we found difficult, introducing "density dependent" distributions of the size of the litter.

4) *Randomness of sex.* It is assumed that the determination of sex of a litter member is random, with a fixed probability F that a given individual will be a female. The determinations of sex of any given individuals are mutually independent and independent of all other variables of the system.

5) *Random survival.* For any two positive numbers t_1 and $t_2 > t_1$, we postulate the existence of the conditional probability $\theta(t_1, t_2)$ that an individual of the population alive at age t_1 will survive up to age t_2. For

$$0 < t_1 < t_2 < t_3$$

this probability must satisfy the consistency relation

$$\theta(t_1, t_3) = \theta(t_1, t_2)\theta(t_2, t_3)$$

The survival of particular individuals of the population is assumed to be independent of the survival of other individuals and independent of all other events in the system.

The probability of survival $\theta(t_1, t_2)$ may be assumed to depend upon the sex of the individual concerned and also on its location at its age t_1. However, we did not succeed in developing a theory involving "density dependence" of survival.

6) *Random migration.* The essential contents of this postulate are that the individuals of the population migrate in the same manner as, and independently from, one another and that the speed of migration depends only on the age of the individuals concerned. Mathematically, the postulate refers to an individual aged t_1 located at a point u (we use just one letter to denote the several coordinates of a point in the habitat, whether it is a plane or a higher dimensional Euclidean space), and to a subsequent moment of time $t_2 > t_1$. Assuming that the individual concerned survived up to age t_2, we denote by X its position at that time. X is considered a random variable with probability density

$$f(x - u, t_1, t_2)$$

depending upon the distance $x - u$ between the location x at age t_2 and the location u at age t_1, and on the ages t_1 and t_2. Also it is assumed that the migrations of particular individuals are mutually independent and independent of all other variables of the system. The function f thus introduced may be called the "dispersal" function of the litters. It must satisfy the obvious consistency relation.

Starting with the above postulates, we were able to obtain an explicit formula determining the joint spatial distribution of males and females of each two consecutive generations that are alive at any preassigned moment t. Also, a recursive procedure was developed leading to a similar result for any number of generations. Unfortunately, the process of births, the determination of sex, migration, and deaths are complicated, so that the resulting formulae are also complicated and the limitations of time prevent us from writing the formulae. They have already been published (*3*).

Biological Populations

Out of the various domains in which the process of clustering has been applied thus far, the development of biological populations appears to be more complicated and more demanding than others. Here the main complicating factor is the multiplicity of generations contributing to the population of individuals alive at any time *t*. Also, there is the already mentioned demand for density-dependent probabilities. In other domains we frequently have to deal with only two generations, the "ancestors" and the "progeny." Furthermore, there is ordinarily, no embarrassment with the assumption of regularly spaced "births."

Bombing

In the study of bombing we were concerned with the so-called force requirement and the optimum method of releasing bombs. The criterion was the expected pattern of bombs over a target. Here the generation of "ancestors" is composed of imaginary points representing "centers of impact" of bombs released by single formations of planes. Their distribution is generated by the so-called "errors of aiming." The "progeny" are the single bombs, and their "migration" is the ballistic dispersion combined with the intentional spread achieved by the method of release of bombs by particular planes.

Cloud Chambers

In the problem of ionization within a cloud chamber, the "ancestors" are the so-called "primary events" consisting of encounters of the cosmic ray particle with the atoms of gas filling the chamber. These primary events are supposed to be Poisson-distributed along the path of the particle. Single primary events are not observable. On the other hand, the individual members of "progeny" "surviving" up to a certain time are observable, and the problem consists of deducing their distribution.

Each primary event creates a random number ν of pairs of ions. Also, most of the ions (again a random number) become centers of droplets. The ions created by a given primary event are its "progeny." Those of the ions that bear droplets are the "surviving" progeny. The "dispersal" of progeny corresponds to the displacements of droplets due to Brownian motion occurring between the time of passage of the cosmic ray particle and the time a photograph of the cloud chamber is taken. The observations give the numbers of droplets in unit lengths along the path of the cosmic ray particle, counted on the photographs.

Cosmology

The two problems just described are mentioned here not because of their intrinsic importance or particular interest but merely as illustrations, with the hope that their perusal may suggest other fields where the theory of the same process of clustering might be useful.

The domain in which the stochastic process of clustering appears to promise results more interesting than in others is the domain of cosmology. Here its applications, combined with unfortunately all-too-heavy numerical work, promise solutions of several problems of prime general interest which, as far as we can see, appear intractable in any other way. In fact, with a little luck in attracting the attention of more workers in the field, the process of clustering, with its further theoretical developments, may easily become the basis of a new theory which we like to call indeterministic cosmology. The presentation of the problems involved, in order to be intelligible, requires a substantial substantive astronomical background which may be found in a recent publication (*10*). Here we have to limit ourselves to a few brief remarks.

One of the most remarkable features of modern cosmology is a sharp divergence between the relevant mathematical work on the subject and the work of observational astronomers. The latter is concerned with single galaxies and with clusters of galaxies. The clusters appear as independent dynamical systems composed of a varying number of members. The variation in the number of cluster members is tremendous. There are galaxies which appear to be isolated, thus forming "clusters" with one single member. Also, there are known "double galaxies," or clusters with two members. Then there are groups of three, four five, and so forth, and, finally, there are clusters with memberships running into thousands. The member galaxies of particular clusters vary considerably in their structure, intrinsic brightness, and other characteristics. All this is the subject of intimate and intensive study by observational astronomers. Briefly, the subject may be described as the local concentrations of matter in our space, with their infinite variations of characteristics, mass, brightness, velocity, chemical composition, and so forth.

Paradoxically, while the observational astronomers concentrate their efforts on these local irregular concentrations of matter, the theoretical cosmologists begin with the hypothesis that no such local concentrations exist! In the words of one of the most distinguished contemporary theoretical cosmologists, all "the observed agglomeration of matter into stars and nebulae (galaxies), and even into clusters of nebulae, is to be replaced by a smeared-out substratum (a sort of fluid) which preserves only the uniformities common to all regions —and thereby robs it of most of those characteristics which make the skies a delight and a challenge to poet and astronomer alike!"

The reason for this divergence between observational and theoretical cosmology is that the latter is built on deterministic mathematics. In order to be able to treat the universe deterministically, it is necessary to deal with only a few hypothetical quantities characterizing, hypothetically, this universe, so as to be able to write down and to analyze the relevant differential equations. The results of this analysis are described as the "basic undergarment" of cosmology, and the hope is expressed that comparability with observations may be achieved by merely "tacking" onto the undergarment a few "frills and furbelows required to express individuality."

In addition to the concept of substratum, the classical cosmological theories adopt another hypothesis, described as the cosmological principle. The formulations of this principle are invariably extremely vague. Here is one of them: "Observations . . . have given rise to the notion that the distribution and motion of matter in any sufficiently large spatial region of this universe are, by and large, intrinsically much the same as those in any other similar region, regardless of its position and orientation. This presumed uniformity in the large implies a certain form of a principle of relativity, sometimes called, appropriately enough, the 'cosmological principle.' "

There is little doubt that the "cosmological principle" is an extension of the Copernican revolution against the dogma of the earth's being the center of the universe. After denying this privilege to our planet, a further step was taken leading to the assumption that the universe has no "center" at all and that, in fact, "essentially," whatever happens in one spot of the universe, is "essentially" equivalent to whatever happens in another. While the idea itself is attractive, the notorious "local irregularities" in the distribution of matter jump to the eye and create embarrassment.

As far as we can see, this embarassment persists only so long as one searches among the concepts of deterministic mathematics for a system that can account for the "essential" uniformity of the universe combined with the observed tremendous "local irregularities." On the other hand, it seems to us that the difficulty immediately disappears if one adopts the indeterministic point of view. This involves the adoption of the attitude that the universe, as we see it, is a single realization of a four-dimensional stochastic process (three spatial coordinates and one time coordinate). The vaguely expressed assumption that "the distribution and motion of matter in sufficiently large spatial regions are, by and large, intrinsically much the same . . ." corresponds to a conceptually rigorous hypothesis that the stochastic process in question is stationary in the three spatial coordinates. Operationally, the same assumption of stationarity may be expressed by postulating that in every region of space there exists a particular chance mechanism, the same for all regions, governing the distribution of matter and its motions.

It must be obvious that the adoption of the point of view just described removes the cleavage between observational and theoretical cosmology. The "local irregularities" are no longer in conflict with the theory postulating universal uniformity. The reason is that the sameness of chance mechanisms operating at two different points in space or time, or both, is, by virtue of their being *chance mechanisms,* perfectly consistent with even the most marked difference in observable manifestations.

Considerations of this kind led us to believe that the esthetic value of the cosmological theory will be markedly increased if, instead of starting deterministically and then "tacking on" the frills and furbelows of indeterministic elements, indeterminism is adopted from the very beginning. The work involved consists of the invention of a four-dimensional stochastic process with consequences which would agree with observations. In constructing this process it is natural to take into account both the ideas underlying the classical deterministic theories and the concepts and preoccupations of observational astronomy. As already mentioned, the latter involve clusters of galaxies with varying numbers of members. Thus, inspired by discussions and joint work with colleagues C. D. Shane and more recently N. U. Mayall, both of the Lick Observatory, our own efforts over the last decade have been directed toward the construction of a four-dimensional stochastic process of clustering to represent the universe, and the development of methods of its study using the available observations.

Before concluding we wish to mention an interesting aspect of the latter problem. Although the process of collecting observational data takes time, compared with the cosmic scale, the intervals of time involved are negligible and no numerical error is to be expected from the assumption that all the observations thus far accumulated were made simultaneously. However, it must be realized that the observations made on a single photographic plate, recording images of several galaxies, refer to happenings at those galaxies which occurred at widely different epochs in time. The reason is, of course, that, generally, no two galaxies are at the same distances from us, that the variation in these distances runs into millions of light years and that, therefore, the observation of one galaxy *A* may refer to what happened say 100 million years ago and the simultaneous observation of another galaxy *B* to the happenings some 200 million years earlier.

Because of this fact, the realization of the hypothetical four-dimensional stochastic process representing the universe is not directly observable. All the observations now available and also those that may be available in the future refer, essentially, to another process, with only two dimensions (symbolized by the angular coordinates of the galaxies), which represents sort of a reflection or an abbreviated projection of the basic four-dimensional process. Here the word *abbreviation* is used to refer to the fact that, owing to great distances, some of the far-away galaxies are too faint to be observed by our instruments. To our minds, the possibility and the intricacies of the study of the basic four-dimensional process, using only its "abbreviated" two-dimensional projection is "a delight and a challenge" to the mathematical statistician and astronomer alike!

In addition, the indeterministic study of cosmology has an aspect for which the word *utilitarian* may not be altogether inappropriate: it so happens that certain questions raised by the deterministic theory, namely the distinction between the various theories of expansion of the universe, cannot be solved through the deterministic approach yet promise a solution through a statistical study (*10, 11*). The point is that the nature of the two-dimensional stochastic process referring to observations—namely the properties of the distribution of images of galaxies on the photographic plates—depends upon whether or not the universe is expanding (strictly speaking, on whether the basic four-dimensional stochastic process is stationary in all four dimensions, including time) and this may be tested by comparisons with observations.

References and Notes

1. This work had the partial support of the Office of Ordnance Research, U.S. Army (contract DA-04-200-ORD-171).
2. J. Neyman, T. Park, E. L. Scott, "Struggle for existence, the *Tribolium* model: Biological and statistical aspects," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, Berkeley, 1956), vol. 4, pp. 41–79.
3. J. Neyman and E. L. Scott, "On a mathematical theory of populations conceived as conglomerations of clusters," *Cold Spring Harbor Symposia Quant. Biol.* 22, 109 (1957).
4. This work had the partial support of the Office of Naval Research and the Office of Ordnance Research.
5. This work had the partial support of the Office of Naval Research and the National Science Foundation.
6. This work had the partial support of the U.S. Public Health Service research grant RG-3666.
7. A. Polnik, paper presented before the Ecological Society of America, Washington, D.C., 28 Dec. 1958.

→ J. Neyman, "On a new class of 'contagious' distribution, applicable in entomology and bacteriology," *Ann. Math. Stat.* 10, 35 (1939).

9. J. Neyman, E. L. Scott, C. D. Shane, "Statistics of images with particular reference to clustering," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, Berkeley, 1956), vol. 3, pp. 75–111.
10. J. Neyman and E. L. Scott, "Statistical approach to problems of cosmology," *J. Roy. Statist. Soc.* B20, 1 (1958).
11. E. L. Scott, "Brightest galaxy in a cluster as a distance indicator," *Astron. J.* 62, 248 (1957).