STOCHASTIC PROCESSES OR THE STATISTICS OF CHANGE*

MAURICE S. BARTLETT
University of Manchester

The term 'stochastic processes' is used to denote statistical phenomena that develop in time and the theoretical models that arise in their treatment, and as these are encountered in many fields, this article has a wide practical interest. Professor Bartlett indicates how stochastic processes arise in physics and communication engineering, industry, economics, biology, and medicine. As an example he describes the construction of a mock epidemic series simulating successive measles outbreaks in a boarding school.

What are Stochastic Processes?

A few years ago not many people in this country knew what was meant by a stochastic process; today the situation is perhaps no different for the general public, though professional statisticians are becoming more familiar with the phrase. Like other new phrases or words, its use 'catches on,' and it gradually spreads through the community of statisticians or other receptive agents much in the way an actual infection will spread. Such a process is itself a *stochastic* process, by which is meant that it does not proceed according to any immutable law but is at least partly dependent on random and chance factors. We therefore call it a *random* or *stochastic* process, usually preferring the second adjective because *random* might convey the idea that every stochastic process appeared purely haphazard (like the emissions from a radioactive substance or the so-called Brownian motion of small dust particles on the surface of a liquid), whereas in many stochastic processes, such as the spread of epidemics or the growth of populations, any random fluctuations may be apparently eliminated by the large statistical groups involved, so that the development of the process appears comparatively smooth and even predetermined.

Anyone previously unfamiliar with the idea of a stochastic process will by now be beginning to see what is meant. He may, even if his mathematics is a relic from his schooldays, remember the distinction in mechanics between statics and dynamics. If he is an economist he will know that much of classical economic theory is erected on the same kind of static or equilibrium structure as classical statistical mechanics, and will know that modern economic theorists, like their physicist colleagues, are busy trying to formulate their theories to

* This article is based on a paper given at the Joint Conference of the Royal and Manchester Statistical Societies held in Manchester in September 1952.

represent a little more closely our dynamic changing world. If he is a statistician he will know that mathematical statisticians long ago began to study the statistical populations and frequency distributions arising in nature and how far they may be represented by theoretical models which assist in their interpretation. *The theory of stochastic processes is, roughly speaking, concerned with the corresponding wider theory of the statistics of change.*

Interpreting the subject in this way, we can be either excited by its generality or disappointed by its lack of novelty. Of course, stochastic processes have always been there in nature, and the industrial statistician studying his control charts or the commercial statistician his firm's fluctuating sales figures will not automatically solve his problems by calling them by a new name. However, while the statistician should never be hidebound by the standard techniques available to him, there is a limit to the extent to which even the best statistician can make *ad hoc* improvisations on current methods, and there is no doubt that recent systematic study of the theory of stochastic processes has greatly broadened his possible approach to actual statistical problems. I will cite two or three examples of how the older 'static' outlook tended to be a barrier to improved technique until it was broken down.

The first was in the statistical analysis of time-series.* The 'static' procedure of considering a given sample of *independent* observations had of course been adapted as far as possible to the study of time-series even in classical methods, in which the stochastic process was represented by a trend or a harmonic curve to which independent random fluctuations were supposed added; but even this assumption proved too narrow to cover many cases met with in practice. In particular, the impossibility of such an assumption always being feasible became apparent from the case of continuous time-records. In this case the assumption of independence implied unlimited statistical information if the discrete observations taken over a fixed period of time were increased indefinitely by reducing the interval between successive observations. Historically, the first attacks on this important statistical problem using a more general approach were made independently by the Russian mathematician and econometrician E. Slutsky[23] and by the English statistician Udny Yule[25] in the year 1927.

The second example was in the practice of industrial sampling. The 'single sampling' schemes were supplemented by 'double sampling,' 'inverse sampling,' and finally by the 'sequential sampling' methods, whose theory was mainly developed during the last world war by the American mathematical statistician Abraham Wald.[7] With these sequential methods the new feature is the *continuation* of sampling until enough information has been acquired for a decision to be taken with a specified risk. Sequential sampling is thus much less 'static' than the classical practice of taking a sample of predetermined size, and its

* This general heading strictly includes such topics as control charts, especially if the successive entries in the latter turn out to be correlated.

distributional theory (e.g. determining the average size of sample required in any application) is essentially one falling within the general field of stochastic process theory and, in particular, is related to the 'random walks' and diffusion processes referred to in the next section.

As a third example consider the manner in which discrete frequency distributions arise—in particular the well-known Poisson distribution for small numbers. In text-books this is usually derived from the binomial distribution, but a more direct and in many cases a more natural way is to obtain it as the fundamental distribution associated with events occurring randomly and independently of each other in time, such as the emission of alpha-particles by a radioactive substance or, in suitable cases, the occurrence of accidents to a particular individual or at a particular locality. The theoretical derivation is comparatively simple. If the total number of events occurring in the time t is $N(t)$, then the distribution of $N(t)$ can be specified by its 'probability-generating function'

$$G(z;\, t) \equiv p_0(t) + p_1(t)z + p_2(t)z^2 + \ldots,$$

in which the coefficient $p_r(t)$ of z^r is the probability that $N(t) = r$ after a time t. We suppose that in a small time-interval δt the chance of one extra event occurring is $a\ \delta t$ (for simplicity we assume a constant in time, though this is not essential), and the chance of none, $1 - a\ \delta t$. Then

$$G(z;\, t + \delta t) = a\ \delta t\, zG(z;\, t) + (1 - a\ \delta t)G(z;\, t)$$

or

$$dG/dt = a(z - 1)G,$$

whence

$$G = e^{at(z-1)}$$

if $N(0) = 0$. This is the probability-generating function of the Poisson law $p_r(t) = e^{-m}m^r/r!$, with a mean $m = at$. The theory of stochastic processes thus gives the Poisson distribution a basic role in statistical theory not less than that of any other distribution.

Of course, if the events (e.g. accidents) are not independent the distribution may be modified. An important case is that where the chance of an event in the small time-interval δt is not constant, but depends on the number of events that have already occurred, being of the form $[a + bN(t)]\ \delta t$. This 'contagion' hypothesis may be shown by an extension of the above method to lead to the 'negative binomial distribution,' a result first established as long ago as 1914 by A. G. McKendrick,[4] who was interested in its medical applications. Another way in which this same distribution can arise was discovered in 1920 by Greenwood and Yule, who were investigating the numbers of accidents experienced by a group of munition workers. They found that if these workers were variable in their proneness to accidents then the frequency distribution of the numbers of accidents per individual, obtained from the statistics of the whole group, may be of the 'negative binomial' type (see, for example, Lundberg[3]). If we wish to discriminate between the first hypothesis, that any individual may suffer more accidents than the average because initial accidents contracted by bad luck render him more liable to others, and the second hypothesis, that one individual will differ from another in his accident proneness right from the start, it is necessary to analyse the accidents per individual over more than one time-period. This has recently been done, for example, on statistics collected for South African shunters.[1]

The fact that more than one causal mechanism can generate the same statistical distribution is an obvious warning to the statistician who is hoping to learn something of the way an observed distribution may have arisen. Before embarking on such a task he should ideally be familiar with *all* the theoretical possibilities. Even so, without some further limitation of the possible hypotheses, the extent to which he can unravel data presented to him may be severely limited. This difficult but vital problem of what the statistician is entitled to ask before undertaking a statistical analysis, particularly in connection with stochastic processes, is returned to again later.

The Monte Carlo Method

While systematic study of stochastic processes is recent, it is evident that in various guises they have appeared since the concepts of probability and chance were first formulated. It is in fact remarkable how the early mathematicians in their attacks on probability problems raised by gamblers included studies of game sequences closely related to many modern stochastic process problems, such as sequential analysis or the use of artificial stochastic processes to solve differential equations and other theoretical problems (the so-called 'Monte Carlo method'). For example, in 1657 the famous Dutch mathematician C. Huyghens propounded the following problem (quoted from a paper by Professor G. A. Barnard[17]). 'A and B each take twelve counters and play with three dice on this condition, that if eleven is thrown, A gives a counter to B, and if fourteen is thrown, B gives a counter to A; and he wins the game who first obtains all the counters. Show that A's chance is to B's as 244 140 625 is to 282 429 536 481.' The mathematical equation for this problem is readily set up, for if the chances of obtaining fourteen or eleven at any trial are as $p : q$ (actually 15 : 27 in this case), the probability $P(x)$ of A winning when he has x counters must satisfy the 'difference equation'

$$P(x) = \frac{p}{p+q} P(x+1) + \frac{q}{p+q} P(x-1), \quad (0 < x < 24) \quad \ldots.(1)$$

and also $P(0) = 0$, $P(24) = 1$. The relevant solution is

$$P(x) = [(q/p)^x - 1]/[(q/p)^{24} - 1] \qquad \ldots.(2)$$

or in particular

$$P(12) = 1/[(q/p)^{12} + 1] \qquad \ldots.(3)$$

agreeing with Huyghens's answer.

The interesting point about this is that, if the theoretical solution of the problem, which has been formulated in mathematical terms in equation (1), had not been precisely known, an approximate solution could be obtained not by direct numerical methods, but by repeated simulation of the gambling problem. As a simple illustration, one hundred repetitions were made, all of which resulted in A losing, consistently with the true value of P being as low as 0·000 864. To expedite these repetitions they were made for convenience with the aid of four-figure random numbers rather than of dice. To obtain odds of 15 : 27 we may classify any random number into one of the two groups 0000–3570 and 3571–9999, giving the practically equivalent odds 3571 : 6429. Notice how the artificial games, which simulate real ones, can give us at the same time all possible information we may wish to know. For example, we know the number of trials required before a game is terminated, and so accumulate information on the statistical distribution of the 'length' of a game. This problem can also be solved theoretically, but the mathematical solution is quite complicated.

This 'artificial sampling' or 'Monte Carlo method' is well known to statisticians, so much so that tables of random numbers are a familiar item in their libraries. In recent years, however, it has also been seriously considered by mathematicians as an aid to the solution of differential or other mathematical equations (see, for example, the USA publication on the 'Monte Carlo method'[10]). Thus the above gambling problem is an example of what is called a 'random walk' process in which each 'step' (in this case the transfer of a counter) occurs independently of previous steps; if we further consider the individual steps of this 'walk' to be small compared with the total distances to be traversed (in the gambling problem the number of counters originally held must be comparatively large), it may be shown that the density f of 'paths' when a large number of repetitions of the process is envisaged satisfies the partial differential equation

$$\partial f/\partial t + a\,\partial f/\partial x = b\,\partial^2 f/\partial x^2 \qquad \ldots.(4)$$

where x is the net distance traversed in the 'time' t (the number of steps), $a = (p - q)/(p + q)$, and $b = 2pq/(p + q)^2$. This equation is well known as the 'equation of diffusion' in physics, and conversely, if we met this equation directly and wished to obtain a solution of it by the Monte Carlo method, we could choose an appropriate $p : q$ (altering if necessary the scale of t and hence of a and b), and proceed as already indicated. Of course, we should be unlikely in this rather simple case to use this method in practice unless the boundary conditions were more complicated or we required to accumulate a lot of information simultaneously about the underlying process. However, to illustrate the connection with equation (4), the frequency distribution of the *lengths* of the 'games' in the hundred repetitions already referred to is compared in Fig. 1 with the theoretical distribution from (4) of the time required to reach a boundary at distance 12 units from the starting point. The agreement, with so few stages as 12 to reach the boundary in the artificial games, is surprisingly good.

It might be added that the reason a stochastic process may so often be found corresponding to equations arising in physics is the obvious one that the equation has really arisen in the first place, as in the case of equation (4), from a stochastic process occurring in nature.

Stochastic Processes in Physics and Communication Engineering

The reader will, however, appreciate that this more fundamental role of stochastic processes in physical problems cannot be adequately

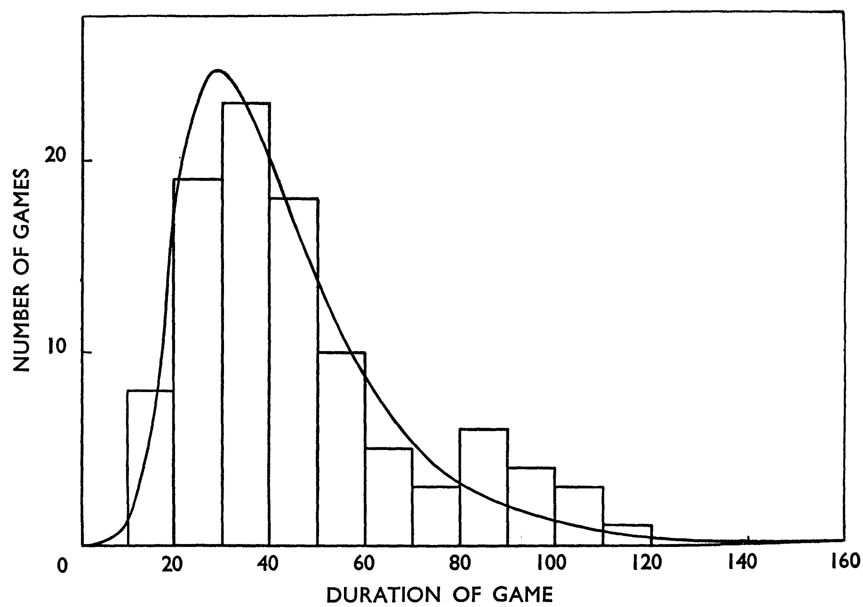

FIG. 1. Frequency distribution of the duration of play in 100 'games,' yielding an approximate solution of a diffusion problem whose correct solution is given by the continuous curve.

indicated here. It must suffice to remind him that the explicit use of the theory of stochastic processes for such physical phenomena as the showers of particles created by cosmic rays or other 'chain reactions,' as Brownian motion or other 'noise' phenomena, or as turbulence in gases and liquids, is merely one indication of the increasing necessity with all physical processes to allow fully for the role which chance and statistical concepts play in them.

Two topics may perhaps be singled out as of particular interest in industrial physics and communication engineering. The first is concerned with the stray disturbances in electrical and other delicate apparatus and has already been referred to above by its usual title, 'noise.' The problem of correcting and filtering a long train of signals to reduce the effect of noise to a minimum is one that comes under the general theory of what are called 'stationary' time-series, and has been so treated by Norbert Wiener, the American mathematician.[15] By a *stationary* time-series is meant a stochastic process in time in which the variable is fluctuating or oscillating but not otherwise changing as time goes on (an example is shown in Fig. 2). From the point of view of the communication engineer a continuing sequence of messages or signals may often be regarded as a stationary time-series, whether or not they are affected by external random disturbances. For example, in a long passage in English the way in which the various letters or even words happen to follow each other has a definite and constant statistical structure which can be studied. If someone sends such a passage by teleprinter or communicates it verbally by telephone, the resulting electrical signals will also constitute a stationary time-series.

The second topic is concerned with the *communication* and *coding* aspect of these series of messages or signals. It is clear that while the sequence of electrical signals should, apart from the effect of disturbances, represent the original sequences of messages, there is a considerable choice in how the representation is made; and one method may be better than another. Here again the concept of stationary time-series is used in the construction of a general theory of communication, in which are studied and made precise such questions as: what is the maximum rate of information that may be passed along a given channel, or equivalently, how 'big' has a channel to be to pass a required rate of information? To give some idea, to those familiar with the technical jargon of the electrical engineer, of the kind of results that can be reached, I will merely quote one important result. The maximum 'capacity' of a channel under certain conditions is given by the formula

$$W \log (1 + P/N)$$

where W is the band-width of frequencies employed, N is the average power in this band-width of the noise in the channel, and P the corresponding average power allowed in the signals. The efficiency of actual communication systems, for example those making use of frequency modulation in radio communication, can then be compared with this optimum.

This communication theory has been largely developed by workers at the Bell Telephone Laboratories, especially by C. Shannon.[13] It is quite general and is not confined to electrical methods of communication. A symposium[14] was held in London in 1950 to discuss its numerous ramifications and several statisticians attended who were interested in its important relation with other branches of statistical theory (see also Barnard[11]).

Stochastic Processes in Industry

An older problem in communication engineering associated with the theory of stochastic processes arises, say, in the design of telephone switchboards and is the one of determining 'waiting times' for any given capacity and density of 'traffic.' But this has so many guises that it is better thought of in the more general terminology of the problem of 'queues,' which the inimitable periodical *Punch* evidently considered (from its review of Mr D. G. Kendall's paper[2]) to be one of the universal problems of our time. Whether the wait is for a disengaged line, or a disengaged shop assistant, or a vacant landing strip at an aerodrome, or a vacant gap in the road traffic, or an available operative to attend to a machine, the wait loses time and money, not to mention our patience. A theoretical and practical study of the stochastic processes involved may help to reduce the amount of time lost.

As an example consider the problem of servicing machines which break down at random times at an average rate r per machine. It is evident that a single group of n operatives to nN machines will be, at least in the absence of any other practical considerations, more efficient than n separate operatives each servicing a separate group of N machines, because the possibility of an operative to one of the latter groups being idle while a machine in another group requires attention is excluded when all operatives are pooled. To illustrate the gain in more detail in a particular case, suppose that the time taken to finish servicing any machine once it has received attention is also random with an average S. Also for simplicity in this example we shall suppose that N is large, but that the average rate of breakdown for all machines remains at a reasonable figure R, say, $(= Nr)$. It might be noticed that our problem is now theoretically equivalent to a queue problem with customers coming in at random to be served, with the operatives representing servers.

The average number of machines (customers) waiting to be served comes out in the case $n = 1$ (one server) as $\alpha^2/(1 - \alpha)$ per server, where $\alpha = RS$, whereas for $n = 2$ it is $2\alpha^3/(1 - \alpha^2)$. Thus the ratio of the average number waiting in the second case to twice the average number for one server is $\alpha/(1 + \alpha)$, indicating the gain in efficiency already referred to. For α, which must be less than 1 if stable conditions are to be maintained, equal to $\frac{1}{2}$ we have $2\alpha^2/(1 - \alpha) = 1$, $2\alpha^3/(1 - \alpha^2) = \frac{1}{3}$, and the gain ratio is as much as $3 : 1$. As n further increases the ratio of the expected number of waiting machines to the total expected number in n individual groups with one operative per group tends steadily to zero. This example has, of course, been rather drastically simplified for illustrative purposes, and statisticians interested in this problem should consult a more comprehensive discussion by F. Benson and D. R. Cox.[18]

Apart from the connection already noted of sequential sampling theory with stochastic processes, many sampling problems have been reconsidered in recent years from the stochastic process viewpoint. Thus the problem of sampling from a continuous 'flow' of material has been discussed by G. H. Jowett[20] and the problem of sampling a two-dimensional area (using the idea of a stochastic process over two *spatial* dimensions) by M. H. Quenouille.[5]

It is clearly not only in the *sampling* of a continuous output of some material that a knowledge of stochastic processes may be useful; it will be needed in the statistical analysis and quality control of the material. For example, in the textile industry, whether for cotton, wool, flax, or other fibre, considerable attention has been given to maintaining uniform quality of the yarn. In particular, certain tendencies to periodicity in the thickness of cotton 'slivers' before they are spun into yarn were discussed by G. A. R. Foster at a symposium[24] on time-series held in 1946. It has already been mentioned that a new approach

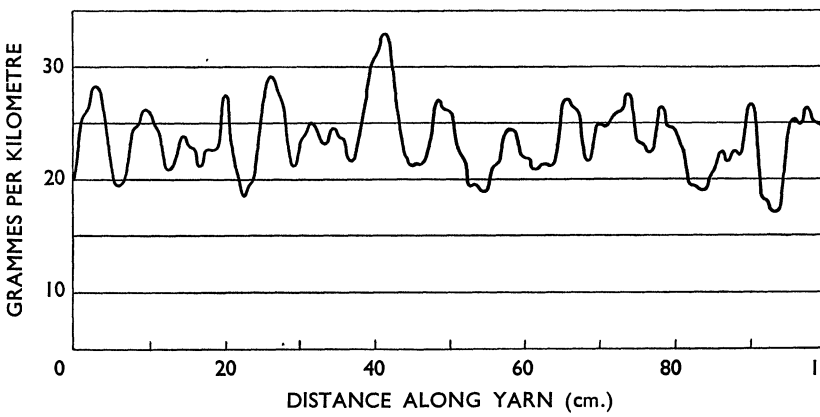

FIG. 2. This chart shows the variation in mass per unit length along a cotton yarn. It was kindly supplied by Mr G. A. R. Foster, who has dealt elsewhere[19,24] with the analysis of such variation. The form of the variation is typical of that shown by a stationary time-series.

to time-series in the last twenty-five years has enabled us to attack such problems statistically with a much greater understanding of what methods to employ.

The Analysis of Economic and Other Time-Series

I have given a technical survey of the statistical analysis of time-series elsewhere[14] and do not want to attempt it here; I do, however, without going into technicalities want to emphasise that the extent to which the statistician can hope to analyse time-series by purely empirical means is now realised to be severely limited. This is because any analysis depends on a theoretical specification or hypothesis of how the data have arisen, and the less detailed the specification available the fewer the conclusions that can be drawn. This is sometimes forgotten in classical statistical problems, where the assumption of a sample of *independent* and *homogeneous* observations is so common that it is often not mentioned explicitly. As an exception, in many experimental situations the lack of independence was not only recognised, but it was effectively eliminated by the randomisation devices introduced by Sir Ronald Fisher. But in time-series, and indeed in stochastic processes in general, the dependence between the successive observations is usually their most important feature. This has thrown up many new and difficult problems in the theory of statistical inference. However, even when these purely technical problems have been solved, it is necessary to realise that the nature of the dependence has so many possibilities *a priori* that these need first to be drastically restricted in any particular context by theoretical or other sources of information before any analysis is likely to be profitable. Possible exceptions are time-series of the stationary type occurring in some physical or meteorological applications, where the length of series available for study may be more or less unlimited.

In economic and social studies it is rare to have homogeneous series of any length, and any statistical analysis must be closely knit with as full a theoretical specification as possible. Referring especially to this field, Norbert Wiener[16] has made much the same point (p. 35): '. . . the modern apparatus of the theory of small samples, once it goes beyond the determination of its own specially defined parameters and becomes a method for positive statistical inference in new cases, does not inspire one with any confidence, unless it is applied by a statistician by whom the main elements of the dynamics of the situation are either explicitly known or implicitly felt.'

The statistical analysis of economic time-series has thus in recent years been based on rather well-defined hypothetical models of how the variables under study interact with one another. When Udny Yule introduced in 1927 methods of analysis for oscillatory time-series in which the random or stochastic element became incorporated with the future movement of the series, he used the vivid illustration of a swinging pendulum which was being bombarded by boys armed with peashooters. It is not unreasonable to suppose that any random or unpredictable disturbances affecting economic series similarly influence their future movement, and so the theoretical economic models used in the specification automatically become stochastic processes of the kind considered by Yule. One detailed (though somewhat indigestible !) exposition[21] of the statistical methods developed for analysing such types of stochastic series has been published by the Cowles Commission Economic Research Group at Chicago.

Stochastic Processes and the Statistician's Role

The essentially close relation between theory and statistical analysis in the case of economic time-series rather brings to a head the problem of what responsibility devolves on anyone who undertakes statistical analysis. It has always been stressed that the statistician must be fully cognisant of how the figures he is analysing were collected and of any other relevant information, but in the case of stochastic processes it is clear that this ancillary information should also include a very thorough theoretical knowledge of the possible mechanism and structure of the process before any analysis can proceed. We have seen that stochastic processes may arise in any field of application, and any statistician who finds himself responsible for their analysis must be prepared either to acquire such knowledge himself or to co-operate with someone who has it.

Let us consider an example in the actuarial field, the prediction of population trends. Here it is true that any random or stochastic element affecting separate individuals is practically eliminated if it is the total population size that is of interest. But while the process will appear smooth, its *detailed structure* is a determining factor in its evolution, and it is well known that any empirical extrapolation based on the census figures for the total population is quite inadequate for any but short-term purposes. This is most evident if we consider as an extreme case a fictitious population of young married emigrants who have founded an island colony. The island's birth-rate would at first show large but gradually damped oscillations with a period of about one generation, until the successive generations had had time to merge into each other. Thus more extended extrapolations before the war took full account of (i) the distribution by age and sex of the total population and (ii) fertility and death rates for individual ages. Even this, which it will be noticed takes account of the *instantaneous* detailed structure of the population of individuals, has been recently shown to be insufficient, for it is necessary to recognise the growing custom of planning family size, and hence to try to follow family histories.[22]

Even when random fluctuations are neglected it is worth remembering that they are still there, and that with smaller groups such as some animal populations they may become a crucial factor. Populations are examples of what are called multiplicative stochastic processes, for which fluctuations are cumulative in time. *In appraising the possible size of random fluctuations the statistician must therefore use the theory of fluctuations appropriate to the relevant stochastic process, and this may sometimes be quite different from the 'classical' theory of fluctuations.* Thus for populations with an expected balance of births and deaths relative fluctuations will theoretically tend to be of order $\sqrt{(t/n)}$, in contrast with the classical formula $\sqrt{(1/n)}$, where t is the time in generations and n the mean size of population. For a human population of 50 000 000 and a generation time of, say, 30 years, this is still only of order 1/5 000 after 60 years; but for an animal population of 100 with a generation time of one year it is of order 1/3 after 10 years.

Stochastic Processes in Biology

This theory of population fluctuations is linked with a problem first raised by Francis Galton at the end of the last century in connection with the extinction of family surnames: if each male individual in a population independently has a family containing n sons, where n is a random number following some given distribution, and the sons in turn each have a number of sons following the same distribution, what is the chance of any particular male line becoming extinct? The complete solution, first obtained by J. F. Steffensen, is a peculiar one. Suppose the probability-generating function of the distribution of n is

$$G(z) = p_0 + p_1 z + p_2 z^2 + \ldots,$$

the probability that $n = r$ being the coefficient p_r of z^r in $G(z)$. Then provided that $p_0 \neq 0$ (otherwise it is obvious that extinction could not occur), the chance of ultimate extinction is the smallest root z_0 of the equation

$$G(z) = z.$$

Moreover, this root z_0 is unity unless the average n is *greater* than unity; even for the United States population in 1920, when the average value of n was 1·145, it was shown by the American actuary Lotka from the statistics of family sizes that the chance of extinction was still nearly 0·9. For example, if for $G(z)$ he substituted the approximate expression $(0{\cdot}482 - 0{\cdot}041z)/(1 - 0{\cdot}559z)$ the above equation became the quadratic equation

$$0{\cdot}482 - 1{\cdot}041z + 0{\cdot}559z^2 = (1 - z)(0{\cdot}482 - 0{\cdot}559z) = 0,$$

giving $z_0 = 0{\cdot}482/0{\cdot}559 = 0{\cdot}86$.

This extinction problem* is important also in the theory of natural selection, since n may alternatively be interpreted as the number of mutant genes in one generation stemming from a mutant gene in the last; as it is unlikely that even a favourable mutation will give an average n much above unity, it will require many occurrences of any such mutation before it is likely to become firmly established in the population.[27]

This last application reminds us that, since the theory of fluctuations in populations is specially important for small populations, it is particularly relevant to all biological population problems involving occasional mutations—the mutant individuals form at first a small population however large the rest of the population may be.

A recent application of the theory of stochastic processes has been to the study of fluctuations in bacterial populations. When the normal bacterial type is placed in an unfavourable environment (for example, a nutrient medium impregnated with bacteriophage) it is possible that a mutant type will arise resistant to this environment. An alternative hypothesis advanced by Sir Cyril Hinshelwood to account for the survival of the bacteria in the new environment is that an actual adaptation of the organism to the environment may lead to survival. Whichever theory may be the correct one (and the latest evidence suggests that neither theory alone is likely to be applicable to all

* For further historical references to this problem see D. G. Kendall's paper 'Stochastic Processes and Population Growth.'[6] The ubiquity of the extinction problem is also indicated (i) by its identification with the gambler's 'ruin' in the historical problems referred to early in this article, and (ii) by its relevance for the epidemiological model treated in the next section.

situations), a study and comparison with observation of the postulated mechanism of growth is evidently necessary. This has been attempted for the mutation theory, and a recent survey is given by Dr P. Armitage.[26]

Stochastic Processes in Medicine

So far I have been taking very much a 'bird's-eye view' of stochastic processes in relation to statistics, and the reader can justifiably complain that the references to applications have been too brief to be other than tantalising. Stochastic processes appear in many branches of medicine also, for example in the study of nervous and cerebral activity or in the study of possible mechanisms of carcinogenesis (see papers by McCulloch & Pitts,[31] and Iverson & Arley[30]). It may, however, be more helpful if I conclude with a single example from the medical field treated in somewhat greater detail. In the discussion following an admirable survey of statistical problems in medicine given at the 1951 Cambridge Conference of the Royal Statistical Society by N. T. J. Bailey[28] I mentioned the probable value of the 'Monte Carlo method' in epidemiological theory. We have seen that it is possible in the growth of very large populations to neglect the stochastic element, but if we do so for epidemics of infectious diseases, especially those whose incidence in the population exhibits a quasi-periodic character, I believe we are in danger of omitting an important factor in their theoretical mechanism.

As an illustration I shall describe the results obtained in a series of fictitious 'measles epidemics,' generated with the aid of random numbers, under conditions simulating a partially isolated group such as a boarding-school. Measles, although not usually a serious complaint, is a favourite infectious disease for study among epidemiologists owing to its relatively simple epidemic character; it confers permanent immunity among almost all those attacked, who are mostly children under the age of 15 (a useful summary of the epidemiology of measles may be found in the late Professor Greenwood's study of epidemics[29]). Notifications are known to exhibit two comparatively stable statistical features. The first is a tendency to biennial periodicity; for example, for Manchester for the years 1917–51 this tendency is quite marked. The second is a seasonal variation; in the case of Manchester for the same years this ranged from 60 per cent. above the average at the beginning of a calendar year to 60 per cent. below in the late summer. It is, however, not at all easy to reconcile these two statistical facts. One may investigate the theoretical consequences of a simple model in which 'susceptibles,' i.e. children who have not yet contracted measles, come in at a given rate and run a risk of infection proportional to the number of infected children present. This model was first shown by Sir William Hamer and later by H. E. Soper[29,33] to be sufficient to produce epidemic waves with a period in time of the right order of magnitude. Unfortunately, in contradiction to what is observed, these waves damp down until an endemic steady state of infection is reached. If we introduce a seasonal variation in infectivity, a 10 per cent. variation is amply sufficient to produce the observed seasonal variation in notifications; it may be shown that the corresponding theoretical variation in numbers of infected comes out at about 80 per cent. But the seasonal variation forces its own annual period on the waves, and the longer natural period, which corresponds to the observed period, still disappears. Dr Soper believed that the introduction of a definite incubation period of a fortnight counteracted the damping effect but, as first pointed out by E. B. Wilson and J. Worcester[35], he was misled by an inaccurate numerical method. From a study of the transmission of infection in individual households Dr P. Stocks and Miss Mary Karn were led to the hypothesis that some of the children exposed to risk acquire a temporary immunity for about a year without visible contraction of the disease, and suggested that this could contribute to the biennial periodic tendency by protecting unattacked children during the danger period of the following year. But this amended model does not appear on theoretical examination to eliminate the difficulty, and I have begun to suspect that a way out from the dilemma must introduce rather different ideas.

The common feature of the calculations on all these models has up to recently been their straightforward actuarial basis, with no attempt to incorporate the random or stochastic element. I shall refer to such models as 'deterministic' models. But it may prove necessary to recognise the essentially local and hence stochastic nature of infection (random overall variation in infectivity due to weather, etc., will of course also contribute), and on such a new basis continual 'extinction' and replenishment of infected individuals within small groups may create a statistical balance with the damping tendency. In other words, the endemic steady state cannot be attained because it is *stochastically* unstable. To investigate this hypothesis for, say, an entire city is a tall order, for it means studying the stochastic vicissitudes of infections in our theoretical model, which must first be expanded to cover adequately the geographical grouping. It is hoped in due course to carry out such an investigation with the aid of the electronic computer, but in the meantime I have compromised by investigating the similar but simpler problem for a fictitious boarding-school, which is treated as an effectively isolated group of children apart from the influx of children at the beginning of each term.

The precise conditions assumed for the stochastic model are as follows:

(i) *Influx of Susceptibles.*

'Lent term' (1st week)	7
'Summer term' (18th week)	7
'Christmas term' (36th week) ..	23

The numbers are intended to represent a typical case, with much the greatest influx at the beginning of the school year. It should be noted that immune children are to be ignored.

(ii) *Influx of Infectives.* The entry of infection is assumed to occur by an occasional one or two incoming susceptibles being already infected; the actual number for any term is random, following a Poisson frequency law, for convenience cut off at the value 3. The mean of this distribution is taken proportional to the number of new entrants, but some provisional assumption is also necessary about the seasonal proportion of infectives in the population from which these new entrants are drawn. It seems reasonable, in the absence of a complete stochastic theory for the population outside the school, to base this on the annual oscillations in numbers which follow, as already noted, from an assumed seasonal variation in infectivity on the *deterministic* model. However, it will be seen that the precise assumptions made about the entry of infection are not very crucial provided that some new infection is present from time to time. In the real situation the children would of course not be completely isolated from the rest of the community during term. Moreover, the actual dispersal of the children in the vacations may not only introduce infection through these children but will effectively terminate school epidemics at the end of each term. In the model the vacations are entirely ignored.

(iii) *Infectivity.* The average infection rate per susceptible is assumed to be $\lambda_r = 0{\cdot}01[1 + 0{\cdot}1 \cos(2\pi r/52)]$ per infected person per week, where r is the number of the week in the calendar year, 1st, 2nd, . . ., 52nd. This gives a maximum seasonal infectivity of 10 per cent. above the average at the beginning of the calendar year, and a minimum of 10 per cent. below the average in the middle of the year. The coefficient 0·01 corresponds roughly to the value 1/300 000 originally adopted by Hamer and Soper, based on rates for the increase in numbers of susceptibles and incidence of infection for the whole of London. Its value has been scaled by a factor of 3 000 to be consistent with the very much smaller group represented by a single school, the average number of susceptibles being taken of the order 50 instead of 150 000. This leaves unaltered the approximate period of one and a half years for the gradually damped-out oscillations which follow an initial major epidemic in the deterministic model.

In the stochastic model the chance of infection in a small time-interval δt is taken to be $\lambda_r N_t S_t \, \delta t$, where N_t is the number of infectious children at any time t (measured in weeks), and S_t the number of children susceptible to attack.

(iv) *Recovery.* The chance of 'recovery,' by which is meant rather non-infectivity, corresponds to an average infectivity period of a fortnight. For simplicity it is assumed that the chance of recovery per interval δt is $\frac{1}{2}\delta t$, and that an infected child is infectious all the time until such recovery. The actual situation is more complicated; for instance, part at least of the incubation period of about a fortnight is non-infectious and this appears to lead in some local measles epidemics to a fortnightly periodic structure at the beginning of the epidemic; but we cannot expect any such 'fine-structure' phenomena to appear in our rather crude model. However, it is emphasised that the primary object of the investigation is to examine the self-consistency and possible broad appropriateness of the model; the above oversimplified pattern of infection and recovery actually exaggerates, by a factor of about 2, the damping of the epidemic waves in the deterministic model, but this will not prejudice its relevance if, as in fact is the case, the damping is eliminated in the stochastic model.

(v) *Initial Conditions.* An initial 'epidemic' was started with 100 susceptibles and 5 infectives, these rather arbitrary numbers ensuring, with the high chance of about 0·97, that the artificial series began with a major epidemic.

Complete details of the calculations will not be given, but the way in which they were made can perhaps be indicated. With a process or model developing continuously in time two methods of obtaining artificial realisations of it are possible. With the first, the time is divided up into small intervals and an approximate model proceeding in terms of these small steps is used. This was the method by which the random walk process referred to under the 'Monte Carlo method' could be used to obtain an approximate solution of the diffusion equation, although this equation represents a *continuous* diffusion of particles. This method is not very convenient for the epidemiological problem, for events such as infections multiply rapidly when an epidemic is under way and may be few and far between in the quiescent periods between epidemics. We want in effect our time-scale much finer during an epidemic than at other times. It is therefore more convenient to adopt the second method, in which we determine the random interval between two consecutive events rather than the number of events in a given interval (cf. reference[9]).

Thus the chance of an 'event' in a small interval δt on the above assumptions is

$$a\,\delta t \equiv \lambda_r N_t S_t\,\delta t + \tfrac{1}{2} N_t\,\delta t$$

where the first term on the right represents the chance of a new infection and the second the chance that the infectives drop by one. The coefficient λ_r changes only slowly with time and may be treated as temporally constant. N_t and S_t, apart from the influx at the beginning of terms, cannot by definition change until the next 'event' has occurred, and hence the coefficient a is effectively constant. It follows from the theory developed earlier in this article that the chance that no event occurs in the interval t is the initial probability in a Poisson distribution with mean at, viz. e^{-at}. This defines the so-called 'exponential distribution' for the interval before a further event occurs, and if we write $T = at$, it will be seen that the distribution of the standardised interval T on this adjusted time-scale is independent of a. We may thus choose a random interval T by some convenient method, and then convert back to the real time scale t by writing $t = T/a$. The method actually used to obtain random intervals was to take $T = \frac{1}{2}(X^2 + Y^2)$, where X and Y are random standardised normal or Gaussian variables available in published tables, but any other method of obtaining a random quantity following the exponential distribution will of course do equally well. At the end of the interval $t = T/a$ we still have to decide *which* event has occurred, but since the relative odds of a new infection and a recovery are $\lambda_r S_t N_t : \frac{1}{2}N_t = 2\lambda_r S_t : 1$, we can write $p = (2\lambda_r S_t)/(1 + 2\lambda_r S_t)$ and from reference to a table of ordinary random numbers decide whether the event that has occurred is a new infection. If it is, N_t goes up by one and S_t down by one; if it is not, N_t goes down by one and S_t up by one.

A further slight approximation is made at the beginning of each term when, if the last interval goes past this date when conditions change, it is ignored and calculations made afresh from this date with the new conditions. If at any stage N_t drops to zero, no change can of course occur until a new influx of children occurs.

The start of the calculations is shown in Table I covering the first week, for which $\lambda_r = 0{\cdot}011\,00$. The coefficient λ_r used is that pertaining at the beginning of each week, and it is not changed until the next week is reached. The number of new cases during the week is the decrease in the number of susceptibles, viz. 7.

TABLE I

Random Interval T	a	T/a	p	Random Number	t (weeks)	N_t	S_t
	8·000				0	5	100
1·291 25		0·161	0·688	2952			
	9·534				0·161	6	99
0·077 45		0·008	0·685	4167			
	11·046				0·169	7	98
6·223 40		0·563	0·683	2730			
	12·536				0·732	8	97
0·547 85		0·044	0·681	0560			
	14·004				0·776	9	96
0·210 85		0·015	0·679	2754			
	15·450				0·791	10	95
0·370 00		0·024	0·676	5870			
	16·874				0·815	11	94
0·000 50		0·000	0·674	9268			
	15·340				0·815	10	94
0·276 25		0·018	0·674	2002			
	16·753				0·833	11	93
0·648 65		0·039	0·672	9568			
	15·230				0·872	10	93
2·645 45		0·174	0·672	8243			
					1·046	9	93

By such calculations the mock epidemic series was generated continuously for a total 'time' of 13 years, during which six major outbreaks were obtained with an average period between these epidemics of 125 weeks. The series was evidently 'steady' in the stochastic sense, with the last epidemics comparable in magnitude with the first. In addition there were four minor outbreaks, in which the entry of infection led to fresh cases but did not lead to a serious epidemic. The results are summarised in Table II, and the course of the largest epidemic (the fifth major epidemic) is shown in Fig. 3. It is stressed that these results are not intended in any sense as a 'fit' to observed data, but statistics for actual measles epidemics in boarding-schools[32] appear to have many similar features if we bear in mind all the further possible complications we have ignored in our model.

TABLE II

Mock epidemic series—summary table

Outbreak	Beginning in Week:	Total Notifications	Percentage Susceptibles Attacked
Major Epidemic			
1	1	98	87
2	105	72	83
3	244	88	85
4	365	67	76
5	504	117	91
6	625	74	87
Minor Outbreak			
1	140	7	18
2	226	7	8
3	556	6	14
4	660	5	15
(On eight other occasions infection entered the 'school' but did not lead to any fresh cases.)			

The observed attack rate of presumed susceptibles in all the schools ranged up to 77 per cent. with, however, not such a clear distinction between major and minor outbreaks. (The actual seasonal incidence in the five years covered by the MRC Report[32] seems specially anomalous, with a higher incidence in the summer term than expected from the average in the population at large or in the corresponding mock series. But here it is again stressed that the extent to which the observed statistics can deviate from the average depends on the mechanism producing them,and for only a five-year period, especially as the incidence in different schools is likely to be correlated, considerable fluctuations would be possible.)

To conclude, while it is certainly not claimed that the mechanism of this simple model is sufficient to explain all the observed facts, it is hoped that this investigation will at least indicate the value of stochastic process models in epidemiological theory. In particular its results, which may be supported by theoretical argument, clearly demonstrate the possibility of a statistical balance between 'extinction' of infectives in a local group and fresh infection from outside. The period will partly depend on the 'isolation' of the local group, but in any case *a major epidemic cannot occur until the concentration of susceptibles has passed*

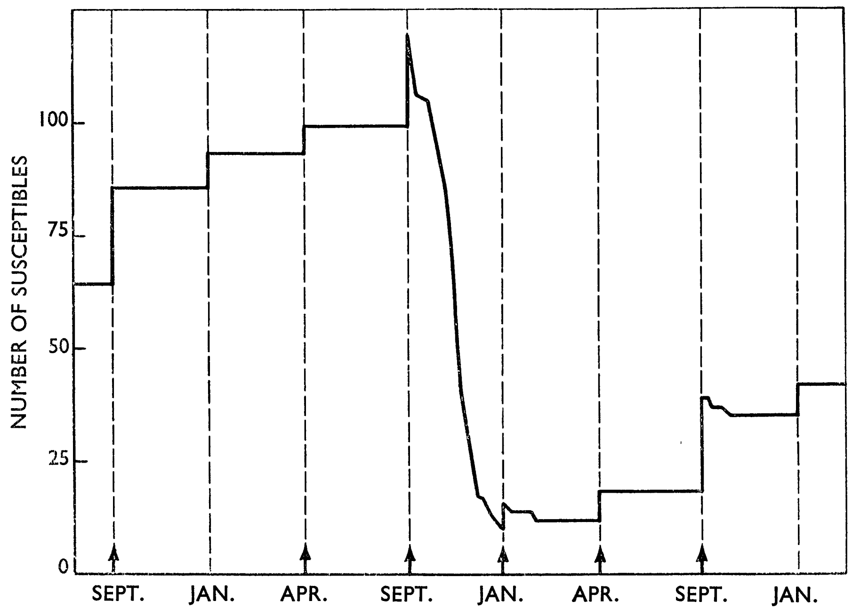

FIG. 3. Extract from mock epidemic series. This graph shows a major epidemic beginning in the autumn of the second 'school' year shown and ending in the New Year. A minor outbreak also occurs the following autumn. The dotted lines indicate the beginning of terms, when there are a number of new entrants to the school; the arrows indicate dates when infection also entered.

its critical threshold value. This critical density of susceptibles is not altogether a new notion to epidemiologists but it is suggested that only with the stochastic approach, in which a smaller density ensures local extinction, can it acquire proper theoretical justification.

I am greatly indebted to Mrs A. Linnert for carrying out the detailed computation for the artificial epidemic series.

REFERENCES AND BIBLIOGRAPHY

This short list of papers and books is confined mainly to applications referred to in the text. For the convenience of readers it has been classified according to the field of application, though this necessitates some cross-referencing. Many further references will of course be found in these publications.

I. *General*

[1] ARBOUS, A. G., & KERRICH, J. E. (1951). 'Accident Statistics and the Concept of Accident Proneness,' *Biometrics*, **7**, 340.

[2] KENDALL, D. G. (1951). 'Some Problems in the Theory of Queues,' *J. Roy. Statist. Soc.* B, **13**, 151.

[3] LUNDBERG, O. (1940). *On Random Processes and their Application to Sickness and Accident Statistics.* Almqvist & Wiksells, Uppsala.

[4] MCKENDRICK, A. G. (1914). 'Studies on the Theory of Continuous Probabilities, with Special Reference to its Bearing on Natural Phenomena of a Progressive Nature,' *Proc. Lond. Math. Soc.* (2), **13**, 401.

[5] QUENOUILLE, M. H. (1949). 'Problems in Plane Sampling,' *Ann. Math. Statist.*, **20**, 355.

[6] SYMPOSIUM ON STOCHASTIC PROCESSES (1949). *J. Roy. Statist. Soc.* B, **11**, 150.

[7] WALD, A. (1947). *Sequential Analysis.* John Wiley, New York.

II. *Monte Carlo Method*

[8] HAMMERSLEY, J. M. (1950). 'Electronic Computers and the Analysis of Stochastic Processes,' *Math. Tables and Other Aids to Comp.*, **4**, No. 29, 56.

[9] KENDALL, D. G. (1950). 'An Artificial Realisation of a Simple "Birth-and-Death" Process,' *J. Roy. Statist. Soc.* B, **12**, 116.

[10] NATIONAL BUREAU OF STANDARDS (1951). *Monte Carlo Method*, Applied Mathematics Series, **12**. Washington, DC.

III. *Physics and Communication Engineering*

[11] BARNARD, G. A. (1951). 'The Theory of Information,' *J. Roy. Statist. Soc.* B, **13**, 66.

[12] CHANDRASEKHAR, S. (1943). 'Stochastic Problems in Physics and Astronomy,' *Reviews of Modern Physics*, **15**, 1.

[13] SHANNON, C. E., & WEAVER, W. (1949). *A Mathematical Theory of Communication*, University of Illinois Press, Urbana.

[14] SYMPOSIUM ON INFORMATION THEORY (1950). Ministry of Supply, London.

[15] WIENER, N. (1949). *The Extrapolation, Interpolation and Smoothing of Stationary Time-series, with Engineering Applications.* John Wiley, New York.

[16] *Idem.* (1948). *Cybernetics.* John Wiley, New York.

IV. *Industry* (See also I[1,2,3,5,7])

[17] BARNARD, G. A. (1946). 'Sequential Tests in Industrial Statistics,' *J. Roy. Statist. Soc.* B, **8**, 1.

[18] BENSON, F., & COX, D. R. (1951). 'The Productivity of Machines requiring Attention at Random Intervals,' *J. Roy. Statist. Soc.* B, **13**, 65.

[19] FOSTER, G. A. R. (1951). 'Fibre Motion in Roller Drafting,' *Shirley Inst. Mem.*, **25**, 51, and *J. Text. Inst.*, **42**, T335.

[20] JOWETT, G. H. (1952). 'The Accuracy of Systematic Sampling from Conveyor Belts,' *Applied Statistics*, **1**, 50.

V. *Analysis of Economic and Other Time-Series* (See also III[14,15,16] and IV[19])

[21] KOOPMANS, T. C. (editor) (1950). *Statistical Inference in Dynamic Economic Models.* John Wiley, New York.

[22] *Report of the Royal Commission on Population* (1949). HMSO, London.

[23] SLUTSKY, E. (1927). 'The Summation of Random Causes as the Source of Cyclic Processes,' *Problems of Economic Conditions*, ed. by the Conjuncture Institute, Moscow, **3**, No. 1 (later reprinted in *Econometrica*, **5** (1937), 105).

[24] SYMPOSIUM ON AUTOCORRELATION IN TIME-SERIES (1946). *J. Roy. Statist. Soc.* B, **8**, 27.

[25] YULE, G. U. (1927). 'On a Method of Investigating Periodicities in Disturbed Series, with special reference to Wolfer's Sun-spot Numbers,' *Philos. Trans.* A, **226**, 267.

VI. *Biology* (See also I[4,5,6] and VII)

[26] ARMITAGE, P. (1952). 'The Statistical Theory of Bacterial Populations subject to Mutation,' *J. Roy. Statist. Soc.* B, **14**, 1.

[27] FISHER, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford University Press.

VII. *Medicine*

[28] BAILEY, N. T. J. (1952). 'The Scope of Medical Statistics,' *Applied Statistics*, **1**, 149.

[29] GREENWOOD, MAJOR (1935). *Epidemics and Crowd Diseases.* Williams & Norgate, London.

[30] IVERSEN, S., & ARLEY, N. (1950). 'On the Mechanism of Experimental Carcinogenesis,' *Act. Path. and Microbiol. Scand.*, **27**, Fasc. 5, 773.

[31] MCCULLOCH, W. S., & PITTS, W. (1948). 'The Statistical Organisation of Nervous Activity,' *Biometrics*, **4**, 91.

[32] MEDICAL RESEARCH COUNCIL (1950). *Epidemics in Schools*, Special Report No. 271, HMSO, London.

[33] SOPER, H. E. (1929). 'The Interpretation of Periodicity in Disease Prevalence,' *J. Roy. Statist. Soc.*, **92**, 34.

[34] WELCH, J. D., & BAILEY, N. T. J. (1952). 'Appointment Systems in Hospital Outpatient Departments,' *Lancet*, **1**, 1105.

[35] WILSON, E. B., & WORCESTER, J. (1945). 'Damping of Epidemic Waves' and 'The Spread of an Epidemic,' *Proc. Nat. Acad. Sci. Wash.*, **31**, 294 and 327.