QUARTERLY OF APPLIED MATHEMATICS
OCTOBER, 1970

WEATHER REGIMES IN STOCHASTIC METEOROLOGICAL MODELS*

BY

R. K. ADENSTEDT**

Brown University

1. Introduction. In the statistical analysis of stochastic processes such as arise in the observation of meteorological phenomena one often notices the *presence of very low frequency components*: the measured quantities behave in a statistically regular manner for long periods of time but display drastic changes between periods. One speaks here of the existence of *weather regimes* [1]. It is the intent of this paper to point out the feasibility and desirability of taking into account the presence of such regimes in statistical meteorology, to indicate how one might construct stochastic models based upon the presence of regimes and to mention a few of the inference problems that arise from such *regime models*.

In speaking of weather regimes, we are concerned less with natural periods such as the 24-hour or one-year periods that may be abstractable from geophysical data by straightforward statistical spectral analysis, but are rather more interested in periods, such as occur perhaps in the movement of high or low pressure areas, that *fluctuate in length and nature*. Even in a stationary model it would generally be difficult to estimate from data limited in time the low frequency part of the spectrum due to such periods, and for this reason we concentrate here mainly on a fundamental model-building approach rather than on statistical analysis of data.

A general approach that may be used in constructing a regime model is outlined in Sec. 2. Although any specific model may be formed in many ways, a definition based on the presence of regimes must essentially incorporate two distinct aspects: description of the *macroscopic* and the *microscopic* behavior of the model. The existence of regimes is introduced in the macroscopic description and is achieved by dividing the time scale into large regions classified according to some regularity conditions. The microscopic description then takes into account the local fluctuation of the final model on these regions. Some typical simplifying assumptions and model types are mentioned.

In Sec. 3 we illustrate the construction in Sec. 2 by defining a specific example of a regime model, which we call a *Markov regime process* and which essentially represents a randomly transformed Markov chain. Although limited in its scope, a Markov regime process is general enough to demonstrate the concept of and the problems associated with a regime process. A few simple properties and some possible generalizations of the example are also discussed in this section.

Some of the statistical inference problems associated with regime models are mentioned in Sec. 4. Typical problems, which arise from any stochastic model, are those of

* Received May 5, 1969. Supported by NSF Grant GP-7347.

** Present address: Center for the Application of Mathematics, Lehigh University.

prediction in terms of a specified model and *estimation* of unknowns in an incompletely specified model. An inference problem peculiar to regime models is that of *recognition* or *reconstruction* of regimes from observed data. Some approaches to these problems and results in terms of the Markov regime process example are presented.

The reader should note that this paper is not intended as a theoretical presentation. At present we are far from a complete theory even in specific cases, and a general complete theory is outrageously optimistic. We wish merely to present a new approach to what may be ancient problems. Accordingly, we have incorporated few actual results, and those without proof. Appropriate proofs and an abstract handling of a limited general case, plus a fairly complete discussion of Markov regime processes, may be found in [2].

2. Construction of a regime model.

2.1. Models in general. A stochastic model used to describe a phenomenon considered in some sense random may eventually be put in the form of a *stochastic process*

$$\{x(t), \quad t \in T\}, \tag{2.1}$$

that is, a collection of random variables taking values in some *state space* X, with *argument space* T.

In an abstract model, X and T may be quite general, but in a particular case we would govern our choice of state and argument spaces by the observations we make and how we make them. Typical examples of state space are a finite or denumerable set, the real line R^1 and Euclidean n-space R^n. Common types of observations for each example are obvious. The parameter t in (2.1) is most commonly regarded as time, so that the argument space T is the set of times at which observations are made. We might choose $T = \{0, 1, 2, \cdots\}$ when regularly spaced measurements are made or T as an interval $[a, b]$ when data is continuously recorded. The parameter t may also be vector-valued, representing more or other than time. An example particularly applicable in meteorology is $T = R^4$, where t represents position (in spherical coordinates) plus time.

The probabilistic structure of the stochastic process (2.1) may be defined in various conceptually different but fundamentally related ways. In our context it is not convenient to dwell on rigorous definitions, though if pressed to do so, we may think of the $x(t)$, $t \in T$, as measurable transformations on an abstract complete probability space (Ω, Σ, P) into the space (X, Σ_X), where Σ_X is a suitable σ-algebra of subsets of X. Standard works such as [3] and [4] may be consulted for alternate interpretations.

In a completely specified model, we would want to define all probabilities of the form

$$P\{x(t_1) \in B_1 , x(t_2) \in B_2 , \cdots , x(t_k) \in B_k\} \tag{2.2}$$

for any values of positive integer k, $t_1 , t_2 , \cdots , t_k \in T$ and $B_1 , B_2 , \cdots , B_k \in \Sigma_X$. In practice, we may be satisfied with much less. For example, knowledge of all the first and second moments

$$Ex(t), \quad Ex(s)x(t), \qquad s, t \in T \tag{2.3}$$

would suffice for purposes of linear least-squares prediction. We might also define the probabilistic structure indirectly, for instance by stipulating that random equations such as

$$Dx(t) = f(x(t), t) + y(t) \tag{2.4}$$

be satisfied. Here D could be a differential or integral operator, f could be a given (possibly random) function on $X \times T$ and $\{y(t), t \in T\}$ could be a simple stochastic process such as Brownian motion. An equation such as (2.4) might be based on knowledge of the dynamics of the meteorological phenomenon under study. Models of this type are considered in [1].

In the remainder of this section we are concerned with purely statistical models whose probabilistic definition is based on the presence of regimes. Extensions of our procedures to models incorporating dynamics will hopefully be straightforward.

2.2. The macroscopic description. The macroscopic description of a regime model $\{x(t), t \in T\}$ is essentially that part of the construction that introduces the presence of regimes, and is achieved by two often interrelated steps: *partition* of the argument space into regimes and *classification* of regimes.

We start by thinking of the argument space T as being *randomly* partitioned into countably[1] many subsets $R_0, R_1, R_2, \cdots$ called *regimes*. By a *partition* we mean that

$$R_n \cap R_m = \varnothing \ (n \neq m), \qquad \bigcup_{n=0}^{\infty} R_n = T. \tag{2.5}$$

The set R_n will be called the nth *regime*. We should point out immediately that the numbering of regimes is somewhat artificial and is introduced only for the purpose of labeling regimes. It should not play an essential role in the definition of the final regime process.

In many cases we would restrict those subsets of T that we admit as regimes. Typical restrictions that come to mind are that there be only finitely many regimes, that each regime be connected and that each regime have minimum "size".

A natural example of a regime partition may be defined in terms of a random *point sequence* when T is the nonnegative line $[0, \infty)$. We may set

$$R_n = [\tau_n, \tau_{n+1}), \qquad n = 0, 1, 2, \cdots, \tag{2.6}$$

where $\tau_0, \tau_1, \tau_2, \cdots$ is a random T-valued sequence satisfying

$$\tau_0 \overset{a.s.}{=} 0, \qquad \tau_n \overset{a.s.}{<} \tau_{n+1}, \qquad \lim_{n\to\infty} \tau_n \overset{a.s.}{=} \infty. \tag{2.7}$$

A regime process with this partition is defined in Sec. 3.

Defining what is meant by "randomness" of the regime partition, i.e. defining the probabilistic structure of the partition $\{R_n, n = 0, 1, 2, \cdots\}$, can in general be a complicated task. Theoretically, we could think of the R_n as abstract-valued random variables. However, we shall in practice only desire such probabilities as

$$P\{t_1 \in R_{n_1}, t_2 \in R_{n_2}, \cdots, t_k \in R_{n_k}\} \tag{2.8}$$

for fixed integers $k, n_1, n_2, \cdots, n_k$ and fixed $t_1, t_2, \cdots, t_k \in T$, or as

$$P\{t_1, t_2, \cdots, t_k \text{ are in the same regime}\} = \sum_{n=0}^{\infty} P\{t_1, t_2, \cdots, t_k \in R_n\}.$$

Such probabilities may often be defined by relatively simple means. For instance, in the example given by (2.6) it suffices to somehow define the joint distributions of the

[1] The denumerability condition simplifies concepts and in many practical cases is essential, but may in theory be easily removed.

τ_n, for then

$$P\{t_i \varepsilon R_{n_i}, 1 \leq j \leq k\} = P\{\tau_{n_i} \leq t_i < \tau_{n_i+1}; 1 \leq j \leq k\}.$$

The definition is greatly simplified when special conditions are placed on the τ-sequence such as that it has independent increments. For more complex argument spaces, such as the surface of the sphere, we might define the probabilities (2.8) via classical means by first dividing T into a large but finite number of *fixed* small regions and then allowing regimes to be only the unions of neighboring such regions.

In conjunction with the regime partition it is often convenient to define an auxilliary stochastic process $\{N(t), t \varepsilon T\}$ by

$$N(t) = n \qquad \text{whenever} \qquad t \varepsilon R_n .$$

This process will be called the *regime number process*. For the example (2.6), the definition reduces to

$$N(t) = \sup \{n \geq 0 : t \geq \tau_n\},$$

a formula familiar from renewal theory [5]. Probabilities such as (2.8) may now be written simply as

$$P\{t_i \varepsilon R_{n_i}, 1 \leq j \leq k\} = P\{N(t_i) = n_i, 1 \leq j \leq k\}.$$

We could, in fact, have *defined* the regime partition in terms of the regime number process, and this approach is taken in Sec. 3. We then must translate the conditions (2.5) as well as possible restrictions on regimes into statements about the N-process. For example, in the partition defined by (2.6) we must stipulate that

> almost every sample function of $\{N(t), t \varepsilon T\}$ is nondecreasing, right continuous, onto the nonnegative integers and zero at $t = 0$.

We could then define

$$\tau_n = \inf \{t \geq 0 : N(t) = n\},$$

and the conditions (2.7) are satisfied.

As a second step in the macroscopic description, we *classify* the regimes $R_0, R_1, R_2, \cdots$ by *randomly* assigning to each R_n a *regime index* J_n whose value is taken from a set S called the *regime index set*. The statistical behavior of the final model $\{x(t), t \varepsilon T\}$ on two distinct regimes R_n and R_m is considered to be in some sense similar if $J_n = J_m$. As extreme examples we might take $S = \{0, 1\}$, where $J_n = 0$ if there is no precipitation on R_n and $J_n = 1$ otherwise, or $S = R^1$, where J_n is mean temperature on R_n.

The regime indices J_n, $n = 0, 1, 2, \cdots$ may be regarded as random variables on the probability space (Ω, Σ, P) into (S, Σ_S), where Σ_S is a σ-algebra of subsets of S. In a completely specified model we would not only have to define all the probabilities

$$P\{J_{n_1} \varepsilon A_1, J_{n_2} \varepsilon A_2, \cdots, J_{n_k} \varepsilon A_k\} \tag{2.9}$$

for integers $k, n_1, n_2, \cdots, n_k$ and sets $A_1, A_2, \cdots, A_k \varepsilon \Sigma_S$, but must also define the joint probability structure of the regime indices and regime partition. Generally, we might do this by first defining the regime partition probabilistically and then defining the conditional probabilities

$$P\{J_{n_1} \in A_1 , J_{n_2} \in A_2 , \cdots , J_{n_k} \in A_k \mid R_n , n = 0, 1, 2, \cdots\}$$

or by first defining the probabilities (2.9) and then defining the conditional probabilistic structure of the partition given the indices.

In practical situations we could considerably decrease the complexity of our definitions by making suitable *simplifying assumptions* about dependence of the indices and partition on each other and about the form of each. For instance, we might assume that J_n depends conditionally only on R_n , so that

$$P\{J_{n_j} \in A_j , 1 \leq j \leq k \mid R_n , n \geq 0\} \overset{a.s.}{=} P\{J_{n_j} \in A_j , 1 \leq j \leq k \mid R_{n_j} , 1 \leq j \leq k\}.$$

In the example of Sec. 3, we simplify by assuming that the J-sequence is a Markov chain and that R_n depends conditionally only on J_n .

The probabilistic descriptions may be made somewhat easier by introducing a second auxilliary stochastic process as was done for the partition. We define the *regime index process* $\{K(t), t \in T\}$ by

$$K(t) = J_{N(t)} = s \quad \text{whenever} \quad t \in \bigcup_{n : J_n = s} R_n .$$

Since the J-sequence is regainable from this new process by

$$J_n = K(t) \qquad \text{whenever} \qquad t \in R_n ,$$

it suffices to define the probability distributions of the two-dimensional process $\{(K(t), N(t)), t \in T\}$.

We could also have originally *defined* the regime classifications in terms of the K-process. However, then we must stipulate that

$$P\{K(t_1) \neq K(t_2), N(t_1) = N(t_2)\} = 0, \quad t_1 , t_2 \in T,$$

so that only one index can be assigned to a particular regime.

2.3. The microscopic description. The second major step in constructing a regime process $\{x(t), t \in T\}$ consists of defining the *local behavior* of the process on individual regimes. This may be done in as many ways as a stochastic process itself may be defined. In a completely specified model we would essentially define all the probabilities (2.2). However, to take into account the influence of the regime partition and classification, it would be more natural to define the conditional probabilities

$$P\{x(t_j) \in B_j , 1 \leq j \leq k \mid R_n , J_n , n = 0, 1, \cdots\}$$

or equivalently

$$P\{x(t_j) \in B_j , 1 \leq j \leq k \mid K(t), N(t), t \in T\} \tag{2.10}$$

for all $t_1 , t_2 , \cdots , t_k \in T$ and $B_1 , B_2 , \cdots , B_k \in \Sigma_X$. By the very nature of the model we would expect these conditional probabilities to be strongly dependent on the regime partition and indices but, as previously mentioned, would expect the dependence on the regime numbering to be only superficial.

A complete prescription of all the probabilities (2.10) would generally be a hopeless task. But a model could be made quite manageable with the use of appropriate simplifying assumptions. It would, of course, not fit the purpose of the model to make the x-process independent of the partition and indices, but we might assume that $x(t)$ depends conditionally only on the regime number and index at time t, so that

$$P\{x(t_j) \in B_j\ ,\ 1 \le j \le k \mid K(t),\ N(t),\ t \in T\}$$
$$\overset{\text{a.s.}}{=} P\{x(t_j) \in B_j\ ,\ 1 \le j \le k \mid K(t_j),\ N(t_j),\ 1 \le j \le k\}.$$

This condition is equivalent to assuming that the model has the form

$$x(t) = \xi_{K(t),N(t)}(t) = \xi_{J_n,n}(t) \quad \text{for } t \in R_n\ , \tag{2.11}$$

where the $\xi_{s,n}(t)$ are X-valued random variables for each $s \in S$, each integer n and each $t \in T$ and are independent of the K- and N-processes. To downgrade the importance of regime numbering we might further assume that the distribution of $\xi_{s,n}(t)$ does not depend on n and use this subscript only to define independence on separate regimes. A model of the form (2.11) is used in Sec. 3, where it is given a diagrammatic interpretation. A very simple example of (2.11) is

$$x(t) = m_{K(t)} + y(t), \tag{2.12}$$

where the m_s , $s \in S$, are constants and $\{y(t),\ t \in T\}$ is a specified stochastic process such as white noise. Models such as (2.11) and (2.12) may easily be extended to dependence on a finite number of regimes and indices, and are seen to be related to *derived stochastic processes* as introduced in [6].

As mentioned above, we may be satisfied with much less than the full probability distributions (2.2). If we need only the first and second moments (2.3), we might define these conditionally, prescribing

$$E\{x(t) \mid K(t),\ N(t),\ t \in T\},\quad E\{x(s)x(t) \mid K(t),\ N(t),\ t \in T\}$$

for all values of s and t in T. Such moments are especially simple in a model of the form (2.12).

The final probabilistic structure of (2.1) may also be defined indirectly by means of random equations comparable to (2.4). Taking into account the regime indices, we might assume that

$$D_{t,\omega}x(t) = f_{K(t),N(t)}(x(t),\ t) + y(t), \tag{2.13}$$

so that the differential equation is different on different regimes. The interpretation of (2.13) is here identical to that of (2.4). Such an approach has an advantage over a model such as (2.11) as we can guarantee by it continuity of sample functions of the regime process. In conjunction with the regime partition (2.6), an interesting example of (2.13) for vector-valued processes is

$$\frac{d}{dt}x(t) = A_{J_n}x(t) + y(t), \qquad \tau_n < t < \tau_{n+1}\ , \tag{2.14}$$

where the A_s , $s \in S$, are fixed constant matrices. In (2.14) we may specify initial conditions for the various regimes so as to guarantee continuity of $x(t)$. Continuity of derivatives could be introduced by enlarging the degree of the equation.

We should remark that in any theoretical definition of $\{x(t),\ t \in T\}$ we must guarantee that our result is indeed a stochastic process, i.e., that the $x(t)$ are really random variables. We shall not ponder such questions here. A discussion of measurability conditions for the model (2.11) is found in [2]. Additionally, if we use a definition such as (2.13) or (2.14), we must be careful that the x-process is differentiable and must define the sense of differentiability.

3. An example. Markov regime processes. To construct an example of a regime process, we shall use a discrete time scale, for simplicity $T = \{0, 1, 2, \cdots\}$, and assume that the regime index set S is *finite*. We may as well let S be the first r positive integers.

It is easiest to define the regime partition in terms of the index and number processes of Sec. 2. We let $\{K(t), t = 0, 1, 2, \cdots\}$ be a first order Markov chain with state space S, initial probabilities

$$p_i = P\{K(0) = i\}, \quad i \in S, \tag{3.1}$$

and stationary transition probabilities

$$p_{ij} = P\{K(t+1) = j \mid K(t) = i\}, \quad i, j \in S. \tag{3.2}$$

We then define $\{N(t), t = 0, 1, 2, \cdots\}$ by

$$N(t) = \begin{cases} 0 & \text{if } t = 0, \\ t - \sum_{n=1}^{t} \delta_{K(n-1),K(n)} & \text{if } t \geq 1, \end{cases} \tag{3.3}$$

where δ_{ij} is Kronecker's delta. $N(t)$ is then the number of transitions to *different* indices (or "states") made by time t.

We would have arrived at an identical model had we defined the sequence $J_0, J_1, J_2, \cdots$ of regime indices as a Markov chain with *transitions inside a state disallowed* and then let the sequence $\tau_0, \tau_1, \tau_2, \cdots$ of regime change times in (2.6) have increments $\tau_{n+1} - \tau_n$ that are conditionally independent given the J-sequence and are *geometrically distributed* with parameters depending only on J_n. By (3.3) the length of a regime is the *sojourn time* of the Markov chain $\{K(t), t \in T\}$ at the index of the regime. To guarantee that (2.7) holds, we must assume that $p_{ii} < 1$ for $1 \leq i \leq r$, but in practice the p_{ii} would be taken close to unity so that expected regime lengths would be large.

To describe the local behavior we introduce a family $\{\xi_{i,n}(t), i \in S, t, n = 0, 1, 2, \cdots\}$ of real-valued random variables with properties:

(1) the family is independent of the K-process,

(2) the distributions of each process $\{\xi_{i,n}(t), t \in T\}$ do not depend on n and

(3) the processes $\{\xi_{i,n}(t), t \in T\}$ and $\{\xi_{j,m}(t), t \in T\}$ are independent when $n \neq m$.

Then each $\{\xi_{i,n}(t), t \in T\}$ is an independent copy of $\{\xi_{i,0}(t), t \in T\}$, which we call the *observed process for regime index i*.

The regime model defined by (2.11) in terms of the above quantities will be called a (discrete parameter) *Markov regime process*. It will be said to be of *type* 1 if each observed process $\{\xi_{i,0}(t), t \in T\}$ consists of independent random variables and of *type* 2 if this is not the case. The model essentially represents randomly transformed observations of the Markov chain $\{K(t), t \in T\}$. Its structure may be diagrammatically represented as follows: if we arrange the processes $\xi_{i,n} = \{\xi_{i,n}(t), t \in T\}$ in an array

$$\begin{array}{ccccccc} \xi_{10} & \xi_{20} & \xi_{30} & \xi_{40} & \cdots & \xi_{r0} \\ \xi_{11} & \xi_{21} & \xi_{31} & \xi_{41} & \cdots & \xi_{r1} \\ \xi_{12} & \xi_{22} & \xi_{32} & \xi_{42} & \cdots & \xi_{r2} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots \end{array}$$

then we start at time $t = 0$ and follow one of the processes in the first row until we change regimes. We then jump randomly to a process in the second row but a different column and follow it until we again change regimes. Then we jump to the third row, etc. As shown in the array, processes in different rows are independent while those in any column are identically distributed.

The probabilities (2.2) for the type 1 process are nicely expressible in terms of products of r-dimensional matrices and vectors. If we define $p = (p_1, p_2, \cdots, p_r)$ as the row vector of initial probabilities and $M = (p_{ij}, i, j \in S)$ as the transition matrix of the K-process, and if for $B \in \Sigma_X$ we let $\Lambda_t(B)$ be the diagonal matrix

$$\Lambda_t(B) = \operatorname{diag}(P\{\xi_{1,0}(t) \in B\}, \cdots, P\{\xi_{r,0}(t) \in B\}), \tag{3.4}$$

then for $0 \leq t_1 < t_2 < \cdots < t_k$ we have

$$P\{x(t_j) \in B_j, 1 \leq j \leq k\} = p\left[\prod_{j=1}^{k} M^{t_j - t_{j-1}} \Lambda_{t_j}(B_j)\right] u, \tag{3.5}$$

where $t_0 = 0$ and u is the $r \times 1$ column vector with unit entries. Proof of (3.5) is based on elementary properties of Markov chains and may be found in [2]. A similar formula for the type 2 process is much more complex and is omitted.

Further properties of Markov regime processes are mentioned in Sec. 4. Some obvious generalizations of the above model might also be noted. We could make $\{K(t), t \in T\}$ a higher order Markov chain or a *semi-Markov chain* [7]. We might also let S be countably or uncountably infinite, but versions of (3.5) will still be valid. Another type of generalization is a *continuous parameter* Markov regime process, which is essentially defined for $T = [0, \infty)$ by letting $\{K(t), t \in T\}$ be a continuous parameter Markov chain. See [2] for details.

4. Some inference problems in regime models. Many classical inference problems are adaptable to regime processes. We consider here approaches to a few such problems and show some typical results in terms of Markov regime processes.

4.1. Prediction from a specified model. A problem of particular interest to the meteorologist is *least-squares prediction*. We wish to estimate an unknown value $x(t)$ of our stochastic process by a function $x^*(t)$ of observed values $x(t_1), x(t_2), \cdots, x(t_k)$ so as to minimize the expected square error $E|x(t) - x^*(t)|^2$. The general solution to this problem is given [8] by the conditional expectation

$$x^*(t) = E\{x(t) \mid x(t_j), 1 \leq j \leq k\}. \tag{4.1}$$

The predictor (4.1) is generally difficult to calculate, requiring knowledge of the probabilities (2.2), but in some cases may take a manageable form. For instance, in the type 1 Markov regime process of Sec. 3, if we assume that the random variables $\xi_{i,0}(t)$ have density functions $f_{i,t}$ and finite means, then (4.1) is given for $0 \leq t_1 < t_2 < \cdots < t_k < t$ by

$$x^*(t) = \frac{p\left[\prod_{j=1}^{k} M^{t_j - t_{j-1}} C_{t_j}(x(t_j))\right] M^{t - t_k} A_t u}{p\left[\prod_{j=1}^{k} M^{t_j - t_{j-1}} C_{t_j}(x(t_j))\right] u} \tag{4.2}$$

in the notation of (3.5), where $t_0 = 0$, $C_t(x) = \operatorname{diag}(f_{1,t}(x), f_{2,t}(x), \cdots, f_{r,t}(x))$ and A_t is the matrix of means

$$A_t = \text{diag}\,(E\xi_{1,0}(t), E\xi_{2,0}(t), \cdots, E\xi_{r,0}(t)). \tag{4.3}$$

We might reduce the computational complexity of (4.2) considerably by approximations when M is close to the identity matrix.

The prediction problem is greatly simplified when only *linear predictors*

$$x^*(t) = a_1x(t_1) + a_2x(t_2) + \cdots + a_kx(t_k) \tag{4.4}$$

are considered. The *regression coefficients* $a_1, a_2, \cdots, a_k$ are determined by k linear equations in terms of the first and second moments (2.3). For the type 1 Markov regime process these moments may be written via (3.5) as

$$Ex(t) = pA_tu$$

and

$$\begin{aligned} Ex(s)x(t) &= pV_tu \quad \text{if } s = t, \\ &= pA_sM^{t-s}A_tu \quad \text{if } s < t, \end{aligned}$$

where $V_t = \text{diag}\,(E\xi_{1,0}^2(t), \cdots, E\xi_{r,0}^2(t))$ and A_t is given by (4.3). When k is large in the linear problem, we might need to compress our data by means of a *screening procedure* (see, e.g., the stepwise regression in [9]).

When our stochastic process $\{x(t), t \in T\}$ is *stationary* (perhaps only in the wide sense) it is often easier to consider linear estimators (4.4) based on the entire past and to solve for the regression coefficients by factorization of the *spectral density* of the process. Such an approach is well covered in [8]. A Markov regime process is stationary if the initial probabilities (3.1) are stationary for the transition probabilities (3.2) and if each observed process $\{\xi_{1,0}(t), t \in T\}$ is stationary. If, in addition, the transition matrix M is *ergodic, i.e.*, irreducible and aperiodic, and is diagonalizable, than the type 1 process has spectral density function

$$f(\lambda) = \frac{1}{2\pi}\left(\sigma^2 + 2\sum_{j=1}^{m} c_je_j\frac{\cos\lambda - c_j}{1 + c_j^2 - 2c_j\cos\lambda}\right), \qquad |\lambda| \leq \pi,$$

where the c_j are constant, $\sigma^2 = \text{Var}\,x(0)$ and $e_1, e_2, \cdots, e_m$ $(m < r)$ are the distinct nonzero eigenvalues (not counting multiplicity) of the matrix $M - \lim_{n\to\infty} M^n$. The prediction problem for this kind of rational (in $e^{i\lambda}$) spectral density is particularly well solved.

Instead of extrapolating for $x(t)$, one might instead wish to predict $I_B(x(t))$ for a fixed $B \in \Sigma_X$ in terms of observed values. Here I_B is the characteristic or indicator function of B. Such a prediction problem would be of interest in *probability forecasts*.

4.2. Inference about the model. A general regime model depends on many *parameters*, some of which may be unknown. We then wish to *estimate* the values of these unknown parameters from observed data. In any estimation problem we must first answer questions of *identifiability*: how much of the model may be ascertained from observations? We will usually be forced to assume a priori a functional form for the model, with only a limited amount of unspecified information. Secondly, we will in most cases want to establish some sort of *ergodic theorem* so that our estimators are "good" in the sense of being *consistent*.

A typical problem for Markov regime process is that of estimating the transition probabilities (3.2) when the probability laws of the observed processes $\{\xi_{i,0}(t), t \in T\}$

are known. A suitable ergodic property occurs here if the transition matrix M is ergodic with (unique) stationary probabilities $\pi_1, \pi_2, \cdots, \pi_r$ and if the observed processes are stationary. For then, as shown in [2],

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \phi(x(t)) \overset{a.s.}{=} \sum_{i=1}^{r} \pi_i E\phi(\xi_{i,0}(0))$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \psi(x(t), x(t+1)) \overset{a.s.}{=} \sum_{i,j=1}^{r} \pi_i p_{ij} E\psi(\xi_{i,0}(0), \xi_{j,1}(1))$$

for any real-valued functions ϕ on X and ψ on X^2 for which the right-hand sides exist. To estimate p_{kl} we must now only search for functions ϕ_k and ψ_{kl} satisfying

$$E\phi_k(\xi_{i,0}(0)) = \delta_{ik}, \quad E\psi_{kl}(\xi_{i,0}(0), \quad \xi_{j,1}(1)) = \delta_{ik}\,\delta_{jl}. \tag{4.5}$$

A suitable ϕ_k may always be found when the only solution of

$$\sum_{i=1}^{r} a_i P\{\xi_{i,0}(0) \in B\} = 0 \quad \text{for all} \quad B \in \Sigma_X$$

for constants $a_1, a_2, \cdots, a_r$ is $a_1 = a_2 = \cdots = a_r = 0$. When the random variables $\xi_{i,0}(0)$ have density functions f_i, then ϕ_k may be taken as the function $f_k - g_k$ normalized, where g_k is the projection of f_k on the linear span of the f_i, $i \neq k$. Such a construction was introduced in [2]. When the process is of type 1, we may define ψ_{kl} in (4.5) by $\psi_{kl}(x, y) = \phi_k(x)\phi_l(y)$.

Another approach to estimation problems is *hypothesis testing*. We would hypothesize that our unknown parameters take specific values and then test this hypothesis in the usual ways with observed data. Such an approach might be particularly valuable for models of the form (2.14) or even in testing for the presence of regimes. We could also combine methods of estimation and prediction, forming estimates of unknowns while using these estimates in predictors.

4.3. Reconstruction of the regime pattern. An interesting inference problem somewhat peculiar to regime models can be formulated as follows: given a completely or partially specified model, what procedure can we adopt to approximate the actual regime pattern (partition and indices) realized by a full or partial sample function of the regime process? In some ways this problem can be considered one in estimation. For example, if t is time and we observe $x(t)$ for $0 \leq t \leq a$, then for a partition of the form (2.6) we wish to estimate $(N, \tau_1, \tau_2, \cdots, \tau_n, J_1, J_2, \cdots, J_N)$, a somewhat unusual parameter. Here N represents the number of regimes represented in the sample. In other instances, we can consider the problem to be one of *filtering*: We wish to apply a filter to the "noise" represented by the microscopic behavior to regain the macroscopic behavior.

A possible general approach to the reconstruction problem might be from the standpoint of *statistical decision functions*. Suppose that our process takes the form $\{x(t), t = 0, 1, 2, \cdots\}$ and that we are satisfied in knowing what regime indices $K(0)$, $K(1)$, $K(2)$, $\cdots$ are realized by a particular sample function. If a *loss function* $L \geq 0$ has been defined as S^2, then we would wish to find a sequence $d_0, d_1, d_2, \cdots$ of S-valued *decision functions* such that the *risk*

$$r_t = EL(d_t(x(0), x(1), \cdots, x(t)), \quad K(t))$$

is minimized for each t. When the regime model is completely specified an optimal solution might be obtainable by standard minimax techniques. When the model is incompletely specified we must simultaneously estimate unknown parameters, hoping to construct decision functions that are optimal in some asymptotic sense. Some special cases of this approach for Markov regime processes are handled in [2].

A less general reconstruction solution could be arrived at by *hypothesis testing*. We would take the view point that nature has fixed the regime structure and that we will test various hypotheses about its form. Similarly, we might adopt a *Bayesian* viewpoint. We would assume an a priori probabilistic structure for the regime partition and indices to obtain a posteriori laws from an observed sample.

References

[1] W. F. Freiberger and U. Grenander, *On the formulation of statistical meteorology*, Rev. Inst. Internat. Statist. **33**, 59–86 (1965)

[2] R. K. Adenstedt, *Stochastic regime processes*, Brown University Tech. Rpt. No. 1, U. S. Army Electronics Command, March 1967

[3] M. Loeve, *Probability theory*, 3rd ed., Van Nostrand, Princeton, N. J., 1963

[4] J. L. Doob, *Stochastic processes*, Wiley, New York, and Chapman & Hall, London, 1953

[5] D. R. Cox, *Renewal theory*, Methuen, London and Wiley, New York, 1962

[6] A. J. Stam, *Derived stochastic processes*, Compositio Math. **17**, 102–140 (1965)

[7] R. Pike, a) *Markov renewal processes: definitions and preliminary properties*, Ann. Math. Statist. **32**, 1231–1242 (1961) b) *Markov renewal processes with finitely many states*, Ann. Math. Statist. **32**, 1243–1259 (1961)

[8] A. M. Yaglom, *Stationary random functions*, Prentice-Hall, Englewood Cliffs, N. J., 1962

[9] R. G. Miller, *Statistical prediction by discriminant analysis*, Meteorological Monographs, **vol 4**, no. 25, American Meteorological Society, 1962