INVARIANT CONDITIONAL DISTRIBUTIONS

By J. A. Bather

Stanford University

1. Introduction. Let $\{\theta_n ; n \geqq 0\}$ be a time-homogeneous Markov process whose behaviour is of interest, but which cannot be observed directly. Suppose instead, that at each stage an observation is taken from a distribution determined by the current state of the process. Let $\{x_n ; n \geqq 1\}$ be a sequence of observations, in which each x_n arises from a density $f(x_n \mid \theta_n)$ and is conditionally independent of its predecessors. Suppose now that any realisation of the Markov process must be accompanied by a sequence of decisions, each of which involves costs depending on the corresponding state but not explicitly on the time. We can imagine for example, that a prediction is required at each stage, or perhaps a decision whether or not to apply some external control in order to modify its state and renew the process.

However, the present investigation is not concerned with any particular sequential decision problem, but in general with the information on which such decisions must be based, and with the possibility of expressing this in a simple and useful form. For any specified initial conditions, the information available at time n consists of the distribution of θ_n, given the previous observations $x_1, x_2, \cdots x_n$. In general as n increases, the situation becomes more and more complicated, because of the number of quantities involved. It is a considerable advantage if this increase in the dimensions of the relevant information does not occur: in other words, if the conditional distribution always reduces to a form depending only on a fixed number of variables. Such a reduction occurs in the case of sampling from a fixed distribution, provided that the distribution is of exponential type. Then statistical inferences concerning the unknown parameter can be analysed in terms of sufficient statistics, whose dimensions are unaffected by the sample size. We might expect distributions of exponential type to possess similar advantages here, where the parameter changes stochastically. As we shall see, this is so; although further conditions must also be satisfied.

An example will serve to illustrate the possibilities. Let the process $\{\theta_n\}$ have independent normal increments, each with zero mean and variance $(1 - c)^2/c$, where c is a constant; $0 < c < 1$. Let the observations $\{x_n\}$ be normal, such that x_n has mean θ_n and unit variance. Then the following properties can be established inductively. If θ_0 is normally distributed, then every subsequent conditional distribution is also normal: the distribution of θ_n in terms of the observations $x_1, x_2, \cdots x_n$, has mean $(1 - c)u_n$ and variance $(1 - c)v_n$, where

$$1/v_n = \{c/[1 - c(1 - v_{n-1})]\} + 1 - c,$$

$$u_n/v_n = \{cu_{n-1}/[1 - c(1 - v_{n-1})]\} + x_n .$$

Received 11 June 1964.

We notice that $\{v_n\}$ and hence the sequence of variances is deterministic, so that the conditional distribution depends only on n and the value of u_n generated by the second relation. Further, as n increases, v_n converges to 1, which indicates that the form of distribution also converges. In particular, if $v_0 = 1$, then $v_n = 1$ always holds and the conditional distributions are invariant in the sense that each depends only on the corresponding quantity $u_n = x_n + cx_{n-1} + \cdots c^{n-1}x_1 + c^n u_0$. Thus we have an exponentially weighted moving average which is always sufficient.

The above example has been used as a simple control chart model [1]. The present work arises from a study of its special properties and makes some progress towards characterizing the class of models with similar properties. We consider first, the general conditions for the existence of a sequence of statistics $\{u_n\}$; such that a complete description of the current state is always possible, in terms of n and $u_n(x_1, x_2, \cdots x_n)$. It is shown that the conditional density functions as well as $f(x_n \mid \theta_n)$, must be of exponential type and that a certain integral recurrence equation must be satisfied. The discussion of invariant conditional distributions is based on this equation, and several examples are produced. The final topic is the question of convergence. It is established that, if a suitable invariant distribution exists, then as n increases, it will gradually dominate any initial effects and emerge as the limiting form of the distribution of θ_n relative to u_n. In this paper, attention is restricted to processes with continuous state spaces. A treatment of discrete spaces appears to raise special problems which will require further research. For convenience, it is assumed throughout the arguments that θ_n and u_n are 1-dimensional quantities. However the results can be extended to the case when both are vectors of the same dimension, without any fundamental changes.

2. Exponential forms. Consider the Markov process $\{\theta_n\}$ with state space Θ contained in the real line and suppose its transition law is known. For any given state θ_{n-1}, let $p(\theta_n \mid \theta_{n-1})$ be the probability density with respect to Lebesgue measure, which generates $\theta_n \varepsilon \Theta$. Suppose that for each $\theta \varepsilon \Theta, f(x \mid \theta)$ is a probability density with respect to a fixed measure μ on the subset X of the real line. Further let $f(x \mid \theta) > 0$ always, so that the sample space X does not depend on θ and suppose that the family of distributions parametrised by θ is completely identifiable: $f(x \mid \theta) = f(x \mid \theta')$ a.e. (μ), only if $\theta = \theta'$. Let the process begin at θ_0 prescribed by a density $g_0(\theta_0)$ and let the subsequent states $\theta_1, \theta_2, \cdots$, give rise to observations $x_1, x_2, \cdots$, where x_n is distributed according to $f(x_n \mid \theta_n)$ and is conditionally independent of all preceding states and observations.

We shall investigate the consequences of the following assumption. Associated with the above system is a sequence $\{u_n(x_1, x_2, \cdots, x_n)\}$ of real valued functions such that in general, the conditional distribution of the state θ_n depends on the observations $x_1, x_2, \cdots x_n$, only through the corresponding value of u_n.

Let $g_n(\theta_n \mid u_n(x_1, \cdots x_n))$ be the density of θ_n when $x_1, x_2, \cdots x_n$, are known, and let $h_n(\theta_n \mid u_{n-1}(x_1, \cdots x_{n-1}))$ be the one which applies just before x_n is observed. The relations between these functions are determined by the two basic operations of the system. The first is obtained by considering a typical transition and the second describes the effect of a new observation.

$$(2.1)\quad h_n(\theta_n \mid u_{n-1}) = \int_\Theta p(\theta_n \mid \theta_{n-1}) g_{n-1}(\theta_{n-1} \mid u_{n-1})\, d\theta_{n-1}.$$

$$(2.2)\quad g_n(\theta_n \mid u_n) = [h_n(\theta_n \mid u_{n-1}) f(x_n \mid \theta_n)]/[\textstyle\int_\Theta h_n(\theta_n \mid u_{n-1}) f(x_n \mid \theta_n)\, d\theta_n].$$

Under suitable regularity conditions, Equation (2.2) implies that each of the densities concerned has an exponential form. The differentiability requirements made in this derivation may seem inappropriate, but it is not easy to weaken them substantially. We suppose that the space X can be extended to an interval in such a way that each function u_n is continuous in its arguments and possesses partial derivatives $\partial u_n/\partial x_{n-1}$ and $\partial u_n/\partial x_n$, which do not vanish identically. The range of possible values of u_n is therefore an interval U_n. Suppose further that for some fixed θ' and any $\theta \varepsilon \Theta$, the derivatives

$$(\partial/\partial x)\{\log [f(x \mid \theta)/f(x \mid \theta')]\}, \qquad (\partial/\partial u_n)\{\log [g_n(\theta \mid u_n)/g_n(\theta' \mid u_n)]\},$$
$$(\partial/\partial u_{n-1})\{\log [h_n(\theta \mid u_{n-1})/h_n(\theta' \mid u_{n-1})]\},$$

exist and are continuous in x, u_n, u_{n-1}, respectively. Then Relation (2.2) can be treated as follows.

$$\log [g_n(\theta \mid u_n)/g_n(\theta' \mid u_n)]$$
$$= \log [h_n(\theta \mid u_{n-1})/h_n(\theta' \mid u_{n-1})] + \log [f(x_n \mid \theta)/f(x_n \mid \theta')],$$
$$(\partial/\partial u_n)\{\log [g_n(\theta \mid u_n)/g_n(\theta' \mid u_n)]\}(\partial u_n/\partial x_{n-1})$$
$$= (\partial/\partial u_{n-1})\{\log [h_n(\theta \mid u_{n-1})/h_n(\theta' \mid u_{n-1})]\}(\partial u_{n-1}/\partial x_{n-1}),$$
$$(\partial/\partial u_n)\{\log [g_n(\theta \mid u_n)/g_n(\theta' \mid u_n)]\}(\partial u_n/\partial x_n)$$
$$= (\partial/\partial x_n)\{\log [f(x_n \mid \theta)/f(x_n \mid \theta')]\}.$$

The first equation is a direct consequence of (2.2) and the others are obtained on differentiating this with respect to x_{n-1} and x_n. Now let us choose another fixed state $\theta'' \neq \theta'$ and compare the last two equations with their counterparts for $\theta = \theta''$. Hence

$$(\partial/\partial x_n)\{\log [f(x_n \mid \theta)/f(x_n \mid \theta')]\}/(\partial/\partial x_n)\{\log [f(x_n \mid \theta'')/f(x_n \mid \theta')]\}$$
$$= (\partial/\partial u_n)\{\log [g_n(\theta \mid u_n)/g_n(\theta' \mid u_n)]\}/(\partial/\partial u_n)\{\log [g_n(\theta'' \mid u_n)/g_n(\theta' \mid u_n)]\}$$
$$= (\partial/\partial u_{n-1})\{\log [h_n(\theta \mid u_{n-1})/h_n(\theta' \mid u_{n-1})]\}/$$
$$(\partial/\partial u_{n-1})\{\log [h_n(\theta'' \mid u_{n-1})/h_n(\theta' \mid u_{n-1})]\} = \phi(\theta).$$

Since the densities $f(x \mid \theta)$; $\theta \varepsilon \Theta$ are completely identifiable, the denominators here cannot vanish identically. We observe that the first ratio depends on θ and x_n, whereas the third does not involve x_n at all. Hence the common value of these ratios depends only on θ.

The result is that each density is determined formally as the solution of a differential equation. For example,

$$(\partial/\partial x) \log f(x \mid \theta) = (\partial/\partial x) \log f(x \mid \theta') + \phi(\theta)(\partial/\partial x)\{\log [f(x \mid \theta'')/f(x \mid \theta')]\}.$$

The general solution for $f(x \mid \theta)$ has the form $a(\theta)b(x) \exp \{\phi(\theta)t(x)\}$. However, this can be simplified without loss of generality. The function $\phi(\theta)$ defines a (1, 1) correspondence, since the family of distributions is completely identifiable. It follows that $\{\phi(\theta_n)\}$ is a Markov process and we can transfer our attention to this instead of $\{\theta_n\}$. Further, with the above density function, $t(x)$ is sufficient for the parameter θ and since we are not directly concerned with $\{x_n\}$, the observations $\{t(x_n)\}$ are equivalent. The original notation is preserved by substituting θ for $\phi(\theta)$ and x for $t(x)$. Similar remarks apply to the general forms obtained for the other densities. Then the results can be expressed as follows.

$$\begin{aligned} f(x \mid \theta) &= a(\theta)b(x) \exp (\theta x), \\ g_n(\theta \mid u) &= m_n(\theta) \exp (u\theta - \lambda_n(u)), \qquad (2.3) \\ h_n(\theta \mid u) &= s_n(\theta) \exp \{c_n(u)\theta - \rho_n(c_n(u))\}. \end{aligned}$$

Finally, by substituting these in Equation (2.2), we can deduce that

$$s_n(\theta) = m_n(\theta)/a(\theta), \qquad u_n = c_n(u_{n-1}) + x_n . \qquad (2.4)$$

Our conclusions so far, correspond closely to the criteria for the existence of sufficient statistics for a fixed parameter. But here, we must take into account transitions of the unobservable quantity. Equation (2.1) can now be written:

$$\begin{aligned} (2.5)\quad \int_\Theta p(\theta_n \mid \theta_{n-1})m_{n-1}(\theta_{n-1}) \exp (u_{n-1}\theta_{n-1} - \lambda_{n-1}(u_{n-1}))\, d\theta_{n-1} \\ = [m_n(\theta_n)/a(\theta_n)] \exp \{c_n(u_{n-1})\theta_n - \rho_n(c_n(u_{n-1}))\}. \end{aligned}$$

The rest of our investigation is mainly concerned with this relation which represents the effect of a typical transition. The exponential forms (2.3) are necessary if we are to obtain the required simplification of all the conditional distributions. The sequence of densities $\{g_n(\theta_n \mid u_n)\}$ must then satisfy (2.5), but the proper interpretation of this relation is by no means obvious. We remark first that, if it can be satisfied, all the regularity conditions which have been assumed are automatically valid. In fact, much stronger properties are implicit and these will be useful in what follows.

The set of values of θ for which $f(x \mid \theta)$ is properly defined by (2.3) is an open interval, possibly unbounded, which can now be identified with the state space Θ. Thus Θ is the set of all θ for which $\int_X b(x)e^{\theta x}\, d\mu(x)$ is finite. Similarly, let us define U_n by the existence of the transformation $\int_\Theta m_n(\theta)e^{u\theta}\, d\theta = \exp \lambda_n(u)$. By considering the variance of the associated distribution, it follows that U_n is an open convex set, which we suppose contains the range of possible values of $u_n(x_1, x_2, \cdots x_n)$. Further, since its definition involves a Laplace transformation, the function $\lambda_n(u)$ can be regarded as analytic in the strip of the complex plane whose real projection is U_n. Later, we shall consider the characteristic function which corresponds to the density $g_n(\theta \mid u)$, and this is simply

$$\int_\Theta m_n(\theta) \exp \{(u + i\xi)\theta - \lambda_n(u)\} \, d\theta = \exp \{\lambda_n(u + i\xi) - \lambda_n(u)\}.$$

Again, let $\int_\Theta s_n(\theta)e^{c\theta} \, d\theta = \exp \rho_n(c)$ be finite whenever c lies in the open interval C_n. Then $\rho_n(c)$ is analytic in the appropriate complex strip. In view of Equations (2.4), we suppose that $c_n(u_{n-1}) + x_n \varepsilon U_n$ for every $u_{n-1} \varepsilon U_{n-1}$ and $x_n \varepsilon X$. Finally, for each fixed $\theta_n \varepsilon \Theta$, Relation (2.5) defines a Laplace transformation and we can deduce that, when complex values of u_{n-1} are admitted, the function $c_n(u_{n-1})$ is analytic.

Suppose now that we can choose $g_0(\theta_0)$ in such a way that it generates a sequence of densities $g_n(\theta_n \mid u_n) = m_n(\theta_n) \exp \{u_n\theta_n - \lambda_n(u_n)\}$ with associated connecting functions $c_n(u_{n-1})$, by the repeated application of (2.5). Then, as we required, the distribution of θ_n given all previous observations, depends only on the single statistic u_n and the time index. The value of u_n can be calculated by using the relations $u_j = c_j(u_{j-1}) + x_j$; $j = 1, 2, \cdots n$, successively. However, a further simplification may be possible: the conditional densities $g_n(\theta_n \mid u_n)$ may not depend explicitly on n. It will be established later, under suitable restrictions, that if there is an invariant form $g(\theta_n \mid u_n) = m(\theta_n) \exp (u_n\theta_n - \lambda(u_n))$, and a fixed connecting function $c(u_{n-1})$, then the conditional distribution of θ_n, given $x_1, x_2, \cdots x_n$, is approximately described by $g(\theta_n \mid u_n)$ when n is large; no matter what density is prescribed for the initial state θ_0. This convergence property suggests that it is most important to examine whether an invariant conditional density can be found.

3. Invariant conditional densities. At this point, it is convenient to adopt a simpler notation. Consider a typical stage in the development of the process $\{\theta_n\}$, and suppose that the current state $\theta \varepsilon \Theta$ is described by the probability density

$$g(\theta \mid u) = m(\theta) \exp (u\theta - \lambda(u)), \tag{3.1}$$

where the appropriate value $u \varepsilon U$ is determined by the preceding observations. We next encounter a transition from θ to a new state $\phi \varepsilon \Theta$ and this is followed by observing $x \varepsilon X$ according to the density

$$f(x \mid \phi) = a(\phi)b(x) \exp (\phi x). \tag{3.2}$$

The information regarding ϕ immediately after the transition, will be less precise than that concerning θ, but we are supposing that the new observation x modifies this in such a way that after both operations, the density of ϕ is still in the same family $\{g(\phi \mid v); v \varepsilon U\}$ as (3.1). In view of the previous analysis, this can only happen if

$$\int_\Theta p(\phi \mid \theta)m(\theta) \exp (u\theta - \lambda(u)) \, d\theta = [m(\phi)/a(\phi)] \exp (c\phi - \rho(c))$$

for some $c = c(u)$, and then the particular density obtained finally, will be $g(\phi \mid c(u) + x)$. Let

$$q(\phi, \theta) = a(\phi)p(\phi \mid \theta). \tag{3.3}$$

Thus (3.1) is an invariant conditional density function if and only if the integral equation

$$\int_\Theta q(\phi, \theta)m(\theta)e^{u\theta}\, d\theta = m(\phi) \exp \{c(u)\phi + \lambda(u) - \rho(c(u))\} \tag{3.4}$$

can be solved for the functions $m(\theta)$ and $c(u)$. In this case, provided that there exists a suitable normalising factor exp $(-\lambda(u))$, whenever $u \varepsilon U$, there is no difficulty in constructing the function $\rho(c)$ from the ratio $m(\phi)/a(\phi)$. The kernel of Equation (3.4) is determined by the transition density and the sampling density (3.2). The central problem is to find an invariant function $m(\theta)$, together with a connecting function $c(u)$.

In general the solution of this integral equation is no easy matter and in what follows, we shall concentrate mainly on a special case for which explicit solutions can be obtained. We shall see by examining the joint distribution of state θ and its successor ϕ, that the problem has a relatively simple interpretation when $c(u)$ is linear in u. This is not the only possibility: Example (iv) discussed in Section 5, involves a connecting function $c(u) = u^{\frac{1}{2}}$. However, the linear case seems to be most promising from the practical point of view.

Consider, for an arbitrary $u \varepsilon U$, the distributions of θ and ϕ determined by (3.1) and (3.4). For the reasons mentioned in Section 2, we can extend the definitions of $\lambda(u)$, $\rho(c)$ and $c(u)$ to admit complex arguments and hence find the appropriate characteristic functions. In particular, for the distribution of θ conditional on ϕ we have

$$\begin{aligned} E(e^{i\xi\theta} \mid \phi, u) &= \int_\Theta p(\phi \mid \theta)m(\theta) \exp (u\theta + i\xi\theta)\, d\theta \\ &\quad \cdot [\int_\Theta p(\phi \mid \theta)m(\theta) \exp (u\theta)\, d\theta]^{-1} \\ &= \exp [\{\lambda(u + i\xi) - \lambda(u)\} - \{\rho(c(u + i\xi)) - \rho(c(u))\} \\ &\quad + \phi\{c(u + i\xi) - c(u)\}]. \end{aligned} \tag{3.5}$$

This is a direct consequence of (3.4). The joint characteristic function is then

$$\begin{aligned} E(e^{i(\xi\theta+\eta\phi)} \mid u) &= \exp [\{\lambda(u + i\xi) - \lambda(u)\} \\ &\quad + \{\rho(c(u + i\xi) + i\eta) - \rho(c(u + i\xi))\}]. \end{aligned}$$

However, a more interesting view of the distribution is provided by (3.5). When the natural order of the states is reversed and θ is considered conditional on ϕ, the cumulants of θ are linear in ϕ. Suppose now that

$$c(u) = cu + k, \tag{3.6}$$

where c and k are constants. Then we have

$$\begin{aligned} E(e^{i\xi(\theta-c\phi)} \mid \phi, u) &= \exp [\{\lambda(u + i\xi) - \lambda(u)\} \\ &\quad - \{\rho(c(u + i\xi) + k) - \rho(cu + k)\}], \end{aligned}$$

and since this does not depend on ϕ, it follows that the random variables ϕ and $\psi = \theta - c\phi$ are independent. This implies certain relationships between the density functions and hence leads to a solution of Equation (3.4).

4. Linear connecting functions. The joint density of the states θ and ϕ can be expressed in terms of either marginal density. We have seen that if the distribution of θ has the form (3.1) and this is an invariant conditional density for the process, and if the connecting function is linear as in (3.6), then $\theta = c\phi + \psi$, where ψ is independent of ϕ. Let the random variable ψ have probability density $w(\psi \mid u)$. The two alternative forms for the joint density of θ and ϕ are then

$$(4.1)\quad m(\theta) \exp (u\theta - \lambda(u))p(\phi \mid \theta) = [m(\phi)/a(\phi)] \exp \{(cu + k)\phi - \rho(cu + k)\}w(\theta - c\phi \mid u).$$

This identity holds almost everywhere and it will be shown that the exceptional null subset of the (θ, ϕ) plane does not depend on $u \varepsilon U$. Consider the necessary conditions on the function $q(\phi, \theta) = a(\phi)p(\phi \mid \theta)$.

Let $\Theta_m = \{\theta \varepsilon \Theta; m(\theta) > 0\}$ and notice that $a(\theta) > 0$ for every $\theta \varepsilon \Theta$, since this inequality defines the state space. Having fixed a particular value of u, we can select $\phi \varepsilon \Theta_m$ such that (4.1) is valid for almost all real values of θ. On replacing θ by $c\phi + \psi$, we have

$$w(\psi \mid u) = [m(c\phi + \psi)/m(\phi)]q(\phi, c\phi + \psi) \exp \{-k\phi + u\psi - \lambda(u) + \rho(cu + k)\}.$$

It follows that this density is of exponential type and can be written

$$(4.2)\qquad w(\psi \mid u) = v(\psi) \exp \{u\psi - \lambda(u) + \rho(cu + k)\}.$$

Then relation (4.1) reduces to a form not depending on u:

$$(4.3)\qquad m(\theta)q(\phi, \theta)e^{-k\phi} = m(\phi)v(\theta - c\phi)$$

and this holds except in a fixed null set N. Its interpretation as a condition on $q(\phi, \theta)$ differs according to whether $c = 1$ or not.

Case $c = 1$. The transition density $p(\phi \mid \theta)$ can be modified arbitrarily on any null subset of the (θ, ϕ) plane, without affecting the process $\{\theta_n\}$. In particular, it can be arranged that Condition (4.3) remains valid at those points $(\theta, \phi) \varepsilon N$, for which $\theta \varepsilon \Theta_m$. Then

$$(4.4)\qquad q(\phi, \theta)e^{-k\phi} = m(\phi)v(\theta - \phi)/m(\theta),$$

whenever $m(\theta) > 0$. This factorisation is necessary, but not sufficient for the existence of an invariant conditional density, since it does not imply (4.3).

Case $c \neq 1$. Here we can obtain a more explicit factorisation. In the first place, we may assume without loss of generality, that $k = 0$. Otherwise this can be arranged by absorbing a factor $\exp (k\theta/(1 - c))$ into $m(\theta)$ and $\exp (k\psi/(1 - c))$ into $v(\psi)$. Having set $k = 0$ in Equations (4.1)–(4.3), and $c(u) = cu$, we can then modify $p(\phi \mid \theta)$ as before, so that

$$(4.5)\qquad q(\phi, \theta) = m(\phi)v(\theta - c\phi)/m(\theta),$$

whenever $m(\theta) > 0$. Now set $\theta = \phi = \psi/(1 - c)$ and suppose that $\psi/(1 - c) \in \Theta_m$. Then

(4.6) $$v(\psi) = q(\psi/(1 - c), \psi/(1 - c)).$$

This may not provide a complete description of the function $v(\psi)$, but otherwise we can make a further modification of $q(\phi, \theta)$, along the line $\theta = \phi$. Thus, by redefining $p(\theta \mid \theta)$ where necessary and by suitably extending the definition of $q(\theta, \theta)$ for $\theta \notin \Theta$, Condition (4.5) can be replaced by the more convenient factorisation:

(4.7) $$q(\phi, \theta)/q[(\theta - c\phi)/(1 - c), (\theta - c\phi)/(1 - c)] = m(\phi)/m(\theta).$$

This must be satisified when $\theta \in \Theta_m$ for every ϕ, provided that we allow its validity at any point where the left hand side reduces to a ratio of zeros. Conversely suppose that, given $q(\phi, \theta)$, there is a constant $c \neq 1$ and a function $m(\theta) \geqq 0$ such that the ratio

(4.8) $$r_c(\phi, \theta) = q(\phi, \theta)/q[(\theta - c\phi)/(1 - c),(\theta - c\phi)/(1 - c)]$$

can be factorised according to Condition (4.7). Strictly, an invariant conditional density may not exist. However, if Equation (4.3) can also be satisfied, with $k = 0$, and if $\int_\Theta m(\theta)e^{u\theta}\, d\theta$ converges for suitable values of u, then a solution of the basic Integral Equation (3.4) is determined.

Several examples of this type of invariant conditional density will be discussed in Section 5, but first let us summarise the general characteristics of $\{\theta_n\}$ and the parallel process $\{u_n\}$ which can be constructed when the invariance property holds.

The initial state θ_0 is described by a density $g(\theta_0 \mid u_0)$, where u_0 is the starting value of the second process. Subsequently, for each state θ_n which occurs, an observation x_n is taken and $u_n = c(u_{n-1}) + x_n$ is calculated. We can imagine that the values $u_0, u_1, \cdots$, are plotted successively on a control chart so that the transitions between states $\theta_0, \theta_1, \cdots$, can be tracked as precisely as possible. Then at any time, the current state θ_n is described by the density $g(\theta_n \mid u_n)$ and this always retains the same form given by (3.1). Unlike the observations $\{x_n\}$, the sequence $\{u_n\}$ is a Markov process. It follows from the relations already established that its transition law is

(4.9) $$P(u_n \leqq v \mid u_{n-1} = u) = \int_{x \leqq v - c(u)} b(x) \exp\{\lambda(c(u) + x) - \rho(c(u))\}\, d\mu(x).$$

We might expect a similarity between this and the law of the original process $\{\theta_n\}$.

One reason for interest in linear connecting functions is that the sequence $\{u_n\}$ then has a very simple structure. If $c(u) = cu$, then in general $u_n = x_n + cx_{n-1} + \cdots c^{n-1}x_1 + c^n u_0$. Thus, the only quantity relevant to a decision about the state θ_n, is an exponentially weighted sum of the previous observations. If $|c| < 1$, the greatest weight is always attached to the most recent observations and as n increases, the effects of u_0 and any early observations diminish to zero. So far, we have assumed that the initial conditions on the system can be chosen in a mathematically convenient form. There would be little practical value in formulating invariant models if their desirable properties depended critically on this assumption. However, it will be shown that if in the linear case, $|c| < 1$, and in general under analogous conditions, then the particular initial distribution will gradually lose its effect and leave the invariant form of conditional distribution dominant in the long run. Section 6 contains a detailed study of this type of convergence.

5. Examples. The first three examples below, illustrate linear connecting functions. Example (i) is a more general version of the one mentioned in the introduction, now discussed in a wider context. All the distributions concerned are normal. For this model, the factor c, which roughly represents its memory, always lies in the range $|c| < 1$. In Example (ii) the observations are from a gamma distribution with a changing scale parameter. Here the magnitude of c is unrestricted. Thus if $c > 1$, the most recent observations are always given least weight in calculating the value of u_n. However, this can only occur when the process $\{\theta_n\}$ is automatically decreasing to zero, while $\{u_n\}$ is increasing at least as fast as the sequence $\{c^n\}$. Example (iii) is concerned with the mean of a Poisson distribution and illustrates the special case $c = 1$. The final example demonstrates an invariant conditional density for which the connecting function is non-linear, and suggests the need for further research in this direction, rather than any immediate practical applications.

Example (i). Let the observations $\{x_n\}$ arise from a normal distribution with known standard deviation which may be taken as the unit of measurement, and mean θ which changes according to the autoregressive scheme:

$$\theta_n = \alpha\theta_{n-1} + \epsilon_n, \qquad (n \geqq 1).$$

The errors $\{\epsilon_n\}$ are normal and independent with common variance σ^2: α and σ^2 are given constants, and we suppose for convenience that each error has mean zero. The process $\{\theta_n\}$ cannot have a stationary distribution unless $|\alpha| < 1$, but here we are interested in the conditional distribution of θ_n at each stage and such considerations are unnecessary. The variables x and θ range over the whole real line and the appropriate functions are as follows.

$$\begin{aligned}
f(x \mid \theta) &= (2\pi)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\theta^2 - \tfrac{1}{2}x^2 + \theta x\right), \qquad a(\theta) = e^{-\frac{1}{2}\theta^2},\\
p(\phi \mid \theta) &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[(-1/2\sigma^2)(\phi - \alpha\theta)^2\right],\\
q(\phi, \theta) &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[(-1/2\sigma^2)\{(1 + \sigma^2)\phi^2 - 2\alpha\phi\theta + \alpha^2\theta^2\}\right],\\
r_c(\phi, \theta) &= \exp\{(-1/2\sigma^2)[\{1 + \sigma^2 - [c^2/(1-c)^2]((1-\alpha)^2 + \sigma^2)\}\phi^2\\
&\quad + 2\{[c/(1-c)^2]((1-\alpha)^2 + \sigma^2) - \alpha\}\phi\theta\\
&\quad - \{[1/(1-c)^2]((1-\alpha)^2 + \sigma^2) - \alpha^2\}\theta^2]\}.
\end{aligned}$$

The condition for a factorisation according to (4.7) is obtained either by equating the coefficients of ϕ^2 and $-\theta^2$ in this expression, or by demanding that the coefficient of $\phi\theta$ should vanish. In both cases, the result is a quadratic equation for c:

$$\alpha c^2 + (1 + \alpha^2 + \sigma^2)c + \alpha = 0.$$

If this is satisfied, then $\sigma^2 = [(\alpha/c) - 1](1 - \alpha c)$ and we can take $m(\theta) = \exp\{-\frac{1}{2}\theta^2/[1 - (c/\alpha)]\}$. It is not difficult to verify that, except in the trivial case $\alpha = 0$, the equation for c always has two real roots c and c' with the properties: $|c| < 1$, $0 < c/\alpha < 1$; $|c'| > 1$, $c'/\alpha > 1$. Thus the function $m(\theta)e^{u\theta}$ is integrable for all values of u if c is used, but never in the case of c'. The invariant conditional density is given by

$$m(\theta)\exp(u\theta - \lambda(u)) = \{2\pi[1 - (c/\alpha)]\}^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}[\theta - [1 - (c/\alpha)]u]^2/[1 - (c/\alpha)]\},$$

and we have the following result. Given an initial distribution of this form, the conditional distribution of any subsequent state θ_n is normal with mean $[1 - (c/\alpha)]u_n$ and variance $[1 - (c/\alpha)]$, where $u_n = \sum_{j=0}^{n-1} c^j x_{n-j} + c^n u_0$. The process $\{u_n\}$ is similar to $\{\theta_n\}$ in that $\{[1 - (c/\alpha)]u_n\}$ follows exactly the same transition law. Then any sequential decision problem associated with the states $\{\theta_n\}$, can be formulated in terms of the derived sequence $\{u_n\}$, which can be observed.

Example (ii). Suppose that the observations $\{x_n\}$ are taken from a gamma distribution with fixed index v, and scale parameter θ which changes as follows. Given the state θ_{n-1}, the ratio $c\theta_n/\theta_{n-1}$ has a beta distribution with indices σ and v. The corresponding density functions are

$$f(x \mid \theta) = \{\Gamma(v)\}^{-1}\theta^v x^{v-1}e^{-\theta x}, \qquad a(\theta) = \theta^v,$$

$$p(\phi \mid \theta) = \{c\Gamma(\sigma + v)/[\Gamma(\sigma)\Gamma(v)]\}(1/\theta)(c\phi/\theta)^{\sigma-1}(1 - (c\phi/\theta))^{v-1} \quad (0 < \phi < \theta/c).$$

The constants c, σ and v must be strictly positive and by changing the sign in the above exponent, we have arranged that the variables have ranges $x > 0$ and $\theta > 0$.

Notice that if $c \geqq 1$, $p(\phi \mid \theta)$ vanishes along the line $\theta = \phi$, so that we cannot apply Condition (4.7) without modifying the definition. However, it is immediately clear that Condition (4.5) is satisfied.

$$q(\phi, \theta) = \{c^\sigma\Gamma(\sigma + v)/[\Gamma(\sigma)\Gamma(v)]\}\phi^{\sigma+v-1}(\theta - c\phi)^{v-1}/\theta^{\sigma+v-1}.$$

Hence $m(\theta) = \theta^{\sigma+v-1}$ is an invariant function with $c(u) = cu$; $u > 0$. The corresponding density

$$m(\theta)\exp(-u\theta - \lambda(u)) = \{\Gamma(\sigma + v)\}^{-1}u^{\sigma+v}\theta^{\sigma+v-1}e^{-u\theta}$$

represents a gamma distribution with index $\sigma + v$ and scale parameter u. The transition law for the process $\{u_n\}$ is easily determined and it follows that the sequence $\{u_n^{-1}\}$ behaves exactly like the original process $\{\theta_n\}$.

Example (iii). Let $\{x_n\}$ be Poisson variates, for which the corresponding means $\{\mu_n\}$ are generated by the rule

$$\mu_n = (1 + (1/\sigma))\mu_{n-1} + \delta_n \qquad (n \geqq 1).$$

Here, $\{\delta_n\}$ is a sequence of independent gamma variates with common index $k > 0$ and scale parameter $\sigma > 0$. Thus $\{\mu_n\}$ is necessarily an increasing sequence. If we set $\mu = e^\theta$ and consider the process so determined, then we have

$$\begin{aligned} f(x \mid \theta) &= \exp(-e^\theta)e^{\theta x}/x! \qquad (x = 0, 1, \cdots), \\ a(\theta) &= \exp(-e^\theta), \\ p(\phi \mid \theta) &= [\sigma/\Gamma(k)]e^\phi\{\sigma e^\phi - (1+\sigma)e^\theta\}^{k-1}\exp\{-(\sigma e^\phi - (1+\sigma)e^\theta)\}, \\ q(\phi, \theta)e^{-k\phi} &= [\sigma/\Gamma(k)]\exp\{-(1+\sigma)e^\phi\}\{\sigma - (1+\sigma)e^{\theta-\phi}\}^{k-1} \\ &\quad \cdot[\exp\{-(1+\sigma)e^\theta\}]^{-1}, \end{aligned}$$

whenever $\phi > \theta + \log[1 + (1/\sigma)]$. Condition (4.4) is satisfied for $m(\theta) = \exp\{-(1+\sigma)e^\theta\}$. The associated invariant conditional density determines the following density for μ, given $u > 0$:

$$(1+\sigma)^u\mu^{u-1}\exp\{-(1+\sigma)\mu\}/\Gamma(u) \qquad (\mu > 0).$$

Thus, if a suitable initial distribution is chosen, then each subsequent state μ_n is described by a gamma distribution with scale parameter $(1 + \sigma)$ and index $u_n = u_0 + nk + \sum_{j=1}^n x_j$.

Example (iv). Consider the observations $\{x_n\}$ from the gamma distribution with density function

$$\begin{aligned} f(x \mid \theta) &= \theta^{\frac{1}{4}}x^{-3/4}e^{-\theta x}/\Gamma(\tfrac{1}{4}); \\ a(\theta) &= \theta^{\frac{1}{4}} \qquad (x > 0, \theta > 0). \end{aligned}$$

Let the corresponding Markov process $\{\theta_n\}$ be generated by the transition density

$$p(\phi \mid \theta) = [2^{\frac{1}{2}}/\Gamma(\tfrac{1}{4})]\phi^{-\frac{1}{2}}\theta^{-\frac{1}{4}}\exp\{-\phi^2/4\theta\} \qquad (\phi > 0).$$

For this system, $m(\theta) = \theta^{-\frac{1}{4}}$ is an invariant function, with connecting function $c(u) = u^{\frac{1}{2}}$; $u > 0$. For each value of u, the associated conditional density is

$$m(\theta)\exp(-u\theta - \lambda(u)) = u^{3/4}\theta^{-\frac{1}{4}}e^{-u\theta}/\Gamma(\tfrac{3}{4}),$$

and the process $\{u_n\}$ is determined by $u_n = u_{n-1}^{\frac{1}{2}} + x_n$. All the above statements can be established by verifying Equation (3.4), which is an elementary but non-trivial exercise.

6. Convergence. We turn now to the general question of stability for invariant conditional distributions. Roughly speaking; if invariance is possible, then provided that the initial density prescribed for the system does not differ too greatly from the invariant form, the conditional distributions which describe its subsequent states must converge to this form. Several assumptions will be needed in order to make this assertion precise.

We begin by restating the essential properties of such a system. To each value θ in an open interval Θ, there corresponds a distribution on the subset X of the real line, having density function $a(\theta)b(x)e^{\theta x} > 0$, with respect to a fixed measure μ. $\{\theta_n\}$ is a Markov process with continuous state space Θ and transitions according to the conditional density $p(\theta_n \mid \theta_{n-1})$. $\{x_n\}$ is a sequence of observations with each $x_n \varepsilon X$, distributed according to the corresponding state θ_n. We assume that there is an invariant function $m(\theta) > 0$ within a subset Θ_m of the state space and otherwise zero, with a connecting function $c(u)$ defined for values of u in an open interval U, such that $\int_\Theta m(\theta)e^{u\theta}\, d\theta = \exp(\lambda(u)) < \infty$,

$$\int_\Theta [m(\phi)/a(\phi)]e^{c(u)\phi}\, d\phi = \exp(\rho(c(u))) < \infty$$

and $c(u) + x \varepsilon U$, whenever $x \varepsilon X$, $u \varepsilon U$. These functions must also satisfy the integral equation

$$(6.1) \quad \int_\Theta p(\phi \mid \theta)m(\theta) \exp(u\theta - \lambda(u))\, d\theta = [m(\phi)/a(\phi)] \exp\{c(u)\phi - \rho(c(u))\}.$$

The basic property which follows is that, if the specification of the processes $\{\theta_n\}$ and $\{x_n\}$ is completed by a density $m(\theta_0) \exp(u_0\theta_0 - \lambda(u_0))$ for the initial state θ_0, where $u_0 \varepsilon U$; then in general, the density for θ_n conditional on the previous observations $x_1, x_2, \cdots x_n$, is $m(\theta_n) \exp(u_n\theta_n - \lambda(u_n))$, where $u_j = c(u_{j-1}) + x_j$; $j = 1, 2, \cdots n$.

Convergence to this invariant form of conditional distribution depends on the way in which the sequence $\{u_n\}$ is constructed from the observations. It is clear that if the particular choice of u_0 affects every element of the sequence even when n becomes large, then the system cannot be stable under more radical modifications of the initial conditions. In view of this, we make the following assumption:

(a) Let the distribution of θ_0 correspond to a value $u_0 \varepsilon U$. Let u_0' be any complex number with real part εU and set $u_n' = c(u_{n-1}') + x_n$; $n \geqq 1$. Then for almost all sequences $\{x_n\}$,

$$\lim_{n\to\infty}(u_n' - u_n) = 0, \lim_{n\to\infty}(\lambda(u_n') - \lambda(u_n)) = 0.$$

From this, it follows immediately that the particular value of u_0 has only a transient effect. Suppose that $u_0' \varepsilon U$ defines another initial distribution and compare the corresponding densities $m(\theta_n) \exp(u_n'\theta_n - \lambda(u_n'))$ and $m(\theta_n)\cdot \exp(u_n\theta_n - \lambda(u_n))$, as n increases. For a proper comparison, these must be evaluated at a fixed point $\theta_n = \theta$, not depending on n, in which case their ratio is $\exp\{(u_n' - u_n)\theta - (\lambda(u_n') - \lambda(u_n))\}$ and this converges to 1. The aim in what follows, is to prove a similar result for a more general initial density $\pi(\theta_0)$ subject to the condition:

(b) $\pi(\theta)$ is a probability density function on the space Θ, such that for some $u_0 \varepsilon U$, the ratio $\pi(\theta)/m(\theta)e^{u_0\theta}$ is bounded and continuous in θ.

Finally, we need some condition which ensures that the initial densities compared, cannot lead to fundamentally different behaviour of the process $\{\theta_n\}$. Let $\{B_n\}$ be a sequence of Borel subsets of Θ and suppose that $\int_{B_n} P(\theta_n \mid \theta_{n-1})\, d\theta_n = 1$ whenever $\theta_{n-1} \varepsilon B_{n-1}$; $n \geqq 1$. If such a sequence exists, then we demand that one of the following conditions must be satisfied:

(c_1) For any $\theta_0 \varepsilon \Theta$, $P(\theta_n \varepsilon B_n$ for some $n \geqq 1 \mid \theta_0) = 1$; or

(c_2) If $\int_{B_0} m(\theta_0) \exp\, (u_0\theta_0 - \lambda(u_0))\, d\theta_0 < 1$, then $\int_{B_0} \pi(\theta_0)\, d\theta_0 < 1$.

We shall consider various distributions connected with the processes $\{\theta_n\}$ and $\{x_n\}$. In what follows, the symbols l and L represent densities and distribution functions respectively. The particular cases concerned will be clear from their arguments and the appropriate initial conditions will be indicated by a suffix π or u_0. For example, the result established in Theorem 6.3 is that for almost every sequence $\{x_n\}$, the ratio $l_\pi(\theta_n \mid x_1, \cdots x_n)/l_{u_0}(\theta_n \mid x_1, \cdots x_n)$, evaluated at a fixed point $\theta_n = \theta \varepsilon \Theta_m$, converges to 1 as n becomes infinite. Notice that the phrase *almost every sequence* can be used without ambiguity in the following sense. Let A be any class of sequences $\{x_n\}$ such that $P_{u_0}(\{x_n\} \varepsilon A) = 1$. Then

$$\int_\Theta P(A \mid \theta_0) m(\theta_0) \exp\, (u_0\theta_0 - \lambda(u_0))\, d\theta_0 = 1$$

$$\text{and hence} \quad P(A \mid \theta_0) = 1; \qquad \theta_0 \varepsilon \Theta_m$$

except perhaps in a subset of Lebesgue measure zero. Thus $P_{u_0'}(A) = 1$ holds for every $u_0' \varepsilon U$. Further, by Assumption (b), $\pi(\theta_0) = 0$ if $\theta_0 \not\varepsilon \Theta_m$, so that $P_\pi(A) = 1$.

The theorem depends on two preliminary results, the first of which concerns the existence of certain conditional distributions and does not involve our special assumptions.

LEMMA 6.1. *For almost every sequence* $\{x_n\}$, $\lim_{n\to\infty} L_{u_0}(\theta_j \mid x_1, x_2, \cdots x_n) = L_{u_0}(\theta_j \mid x_1, x_2, \cdots)$ *exists and is a distribution function in* θ_j ; $j = 0, 1, \cdots$.

For each fixed θ_j and any given observations $x_1, x_2, \cdots x_j$, the random variables $L_{u_0}(\theta_j \mid x_1, x_2, \cdots x_{j+k})$; $k = 1, 2, \cdots$, form a uniformly bounded martingale sequence. The lemma can be obtained from the properties of martingales established by Doob ([2], p. 319). We omit a formal proof, since it is not central to the present discussion.

Consider sequences of observations which are well behaved in the sense of Lemma 6.1. Let $B_0 = \{\theta \varepsilon \Theta;\ \pi(\theta) > 0\}$ and let A be the class of sequences $\{x_n\}$ for which $P_{u_0}(\theta_0 \varepsilon B_0 \mid x_1, x_2, \cdots) > 0$. The next lemma shows that almost every sequence possesses this property.

LEMMA 6.2. *Under Conditions* (b) *and* (c), $P_{u_0}(\{x_n\} \varepsilon A) = 1$.

PROOF. Let A' be the complement of A within the class of well behaved sequences $\{x_n\}$. Suppose that $P_{u_0}(A') > 0$. It will be shown that this contradicts Assumption (c), in both its alternative forms.

Consider any fixed $\{x_n\} \varepsilon A'$. We have

$$P_{u_0}(\theta_0 \varepsilon B_0 \mid x_1, x_2, \cdots) = \int_\Theta P_{u_0}(\theta_0 \varepsilon B_0 \mid \theta_1)\, dL_{u_1}(\theta_1 \mid x_2, x_3, \cdots) = 0.$$

This is obtained by considering the distribution of θ_1 and noting that

$$P_{u_0}(\theta_0 \varepsilon B_0 \mid \theta_1, x_1, x_2, \cdots) = P_{u_0}(\theta_0 \varepsilon B_0 \mid \theta_1)$$

since $x_1, x_2, \cdots$, are independent of θ_0 when θ_1 is given, and $L_{u_0}(\theta_1 \mid x_1, x_2, \cdots) = L_{u_1}(\theta_1 \mid x_2, x_3, \cdots)$, where $u_1 = c(u_0) + x_1$. Let $B_1' = \{\theta_1 \varepsilon \Theta; P_{u_0}(\theta_0 \varepsilon B_0 \mid \theta_1) = 0\}$, and set $B_1 = \Theta - B_1'$. Since the above integrand is strictly positive whenever $\theta_1 \varepsilon B_1$, it follows that $P_{u_1}(\theta_1 \varepsilon B_1 \mid x_2, x_3, \cdots) = 0$. Now consider

$$P_{u_0}(\theta_0 \varepsilon B_0 \mid \theta_1) = [\textstyle\int_{B_0} p(\theta_1 \mid \theta_0) m(\theta_0) \exp (u_0\theta_0 - \lambda(u_0))\, d\theta_0]$$
$$\cdot[\textstyle\int_\Theta p(\theta_1 \mid \theta_0) m(\theta_0) \exp (u_0\theta_0 - \lambda(u_0))\, d\theta_0]^{-1}.$$

Since $B_0 \subset \Theta_m$ by Condition (b), this ratio vanishes only at those points θ_1 for which $p(\theta_1 \mid \theta_0) = 0$ almost everywhere in B_0. In particular, the sets B_1, B_1' do not depend on the value of $u_0 \varepsilon U$. Further, by removing a null subset of B_0, it can be arranged that $\int_{B_1} p(\theta_1 \mid \theta_0)\, d\theta_1 = 1$ for every $\theta_0 \varepsilon B_0$. Such a modification of B_0 cannot affect the class A, nor can it alter the essential property that $\int_{B_0} \pi(\theta_0)\, d\theta_0 = 1$.

We can now repeat the above argument using the equation

$$P_{u_1}(\theta_1 \varepsilon B_1 \mid x_2, x_3, \cdots) = \int_\Theta P_{u_1}(\theta_1 \varepsilon B_1 \mid \theta_2)\, dL_{u_2}(\theta_2 \mid x_3, x_4, \cdots) = 0,$$

where $u_2 = c(u_1) + x_2$. Proceeding in this way, we determine a sequence of sets $\{B_n\}$ such that $\int_{B_n} p(\theta_n \mid \theta_{n-1})\, d\theta_n = 1$, whenever $\theta_{n-1} \varepsilon B_{n-1}$; $n \geqq 1$. This sequence does not depend on the particular choice of $\{x_n\} \varepsilon A'$ and hence, in every case, $P_{u_n}(\theta_n \varepsilon B_n \mid x_{n+1}, x_{n+2}, \cdots) = 0$. Consider the product set $\mathcal{B} = B_0' \times B_1' \times \cdots$. Then for every $\{x_n\} \varepsilon A'$, we have $P_{u_0}(\{\theta_n\} \varepsilon \mathcal{B} \mid x_1, x_2, \cdots) = 1$ and it follows that $P_{u_0}(\{\theta_n\} \varepsilon \mathcal{B}) \geqq P_{u_0}(\mathcal{B} \cap A') = P_{u_0}(A') > 0$. But this is inconsistent with Assumption (c). If (c_1) holds, then $P(\{\theta_n\} \varepsilon \mathcal{B} \mid \theta_0) = 0$ for every $\theta_0 \varepsilon \Theta$. Otherwise ($c_2$) is satisfied and $P_{u_0}(\theta_0 \varepsilon B_0) = 1$, since $\int_{B_0} \pi(\theta_0)\, d\theta_0 = 1$. Again we have a contradiction and the proof is complete.

We are now in a position to investigate the asymptotic form of the conditional density functions $l_\pi(\theta_n \mid x_1, x_2, \cdots x_n)$.

THEOREM 6.3. *Let the initial state θ_0 of the system be determined according to the density $\pi(\theta_0)$ and suppose that Conditions* (a), (b) *and* (c) *are valid. Then for almost every sequence of observations $\{x_n\}$ and any state $\theta \varepsilon \Theta_m$,*

$$\lim_{n\to\infty} [l_\pi(\theta_n \mid x_1, x_2, \cdots x_n)]/[m(\theta_n) \exp (u_n\theta_n - \lambda(u_n))] = 1,$$

where the densities are all evaluated at the fixed point $\theta_n = \theta$.

PROOF. We shall restrict attention to a particular sequence $\{x_n\}$. In view of the preceding lemmas, it may be assumed that the distribution function $L_{u_0}(\theta_0 \mid x_1, x_2, \cdots)$ is properly defined and that $P_{u_0}(\theta_0 \varepsilon B_0 \mid x_1, x_2, \cdots) > 0$. It follows from this and Assumption (b) that

$$(6.2) \quad \int_\Theta \pi(\theta_0)/[m(\theta_0) \exp (u_0\theta_0 - \lambda(u_0))]\, dL_{u_0}(\theta_0 \mid x_1, x_2, \cdots) > 0.$$

The argument is based on the fact that the distribution of θ_0, conditional on $x_1, x_2, \cdots x_n$, and any fixed value θ of θ_n, converges as n increases, to the distribution represented by $L_{u_0}(\theta_0 \mid x_1, x_2, \cdots)$. Although we are concerned with the same state $\theta \varepsilon \Theta_m$ at each stage, it is useful to retain the appropriate time index.

Consider the joint probability density $l_{u_0}(x_1, x_2, \cdots x_n, \theta_n)$ given by

$$\int_\Theta \cdots \int_\Theta m(\theta_0) \exp (u_0\theta_0 - \lambda(u_0)) \prod_{j=1}^n \{p(\theta_j \mid \theta_{j-1})a(\theta_j)b(x_j)e^{\theta_j x_j}\} \, d\theta_0 \cdots d\theta_{n-1} .$$

By applying the Invariance Property (6.1) repeatedly, this becomes

$$l_{u_0}(x_1, x_2, \cdots x_n, \theta_n) = \exp \{ \textstyle\sum_{j=1}^n \lambda(u_j)\} \exp \{ -\textstyle\sum_{j=0}^{n-1} \rho(c(u_j))\} \cdot \textstyle\prod_{j=1}^n b(x_j) m(\theta_n) \exp (u_n\theta_n - \lambda(u_n)). \tag{6.3}$$

Formula (6.1) also holds for complex values of u and by extending this argument, we can evaluate the characteristic function which corresponds to

$$\begin{aligned} l_{u_0}(\theta_0 \mid x_1, x_2, \cdots x_n, \theta_n) &= [l(x_1, x_2, \cdots x_n, \theta_n \mid \theta_0) m(\theta_0) \exp (u_0\theta_0 - \lambda(u_0))] \\ &\quad \cdot [l_{u_0}(x_1, x_2, \cdots x_n, \theta_n)]^{-1}. \end{aligned}$$

$$\begin{aligned} E_{u_0}(e^{i\xi\theta_0} \mid x_1, x_2, \cdots x_n, \theta_n) &= \exp \{ \textstyle\sum_{j=0}^n (\lambda(u_j') - \lambda(u_j))\} \exp \{ -\textstyle\sum_{j=0}^{n-1} (\rho(c(u_j')) - \rho(c(u_j)))\} \\ &\quad \cdot \exp \{(u_n' - u_n)\theta_n - (\lambda(u_n') - \lambda(u_n))\}, \end{aligned} \tag{6.4}$$

where $u_0' = u_0 + i\xi$, $u_j' = c(u_{j-1}') + x_j$; $j = 1, 2, \cdots n$. Further, the density $l_{u_0}(x_1, x_2, \cdots x_n)$ is given by integrating out the last term of (6.3) and the characteristic function associated with $l_{u_0}(\theta_0 \mid x_1, x_2, \cdots x_n)$ is obtained from (6.4) simply by omitting the final term there. We observe that since θ_n is fixed, this term of (6.4) depends only on $(u_n' - u_n)$ and $\lambda(u_n') - \lambda(u_n)$, and therefore converges to 1, by Assumption (a).

Now compare the two sequences of distributions with typical density functions $l_{u_0}(\theta_0 \mid x_1, x_2, \cdots x_n, \theta_n)$ and $l_{u_0}(\theta_0 \mid x_1, x_2, \cdots x_n)$, respectively. We can apply Levi's continuity theorem, which establishes that any sequence of distribution functions converges if and only if the corresponding characteristic functions converge. Here, we have two sequences for which the characteristic functions have the same limiting behaviour. Since the second sequence of distributions converges, it follows that both sequences converge to the same limit, represented by $L_{u_0}(\theta_0 \mid x_1, x_2, \cdots)$. Finally, let us apply the Helly-Bray theorem to the expectations of the function $\pi(\theta_0)/m(\theta_0) \exp (u_0\theta_0 - \lambda(u_0))$. Since this ratio is bounded and continuous, both sequences of expectations converge to the limit given by (6.2). On referring to the definitions of $l_{u_0}(\theta_0 \mid x_1, x_2, \cdots x_n, \theta_n)$ and $l_{u_0}(\theta_0 \mid x_1, x_2, \cdots x_n)$, we find that

$$\begin{aligned} &\lim_{n\to\infty} [l_\pi(x_1 . x_2 . \cdots x_n, \theta_n)]/[l_{u_0}(x_1, x_2, \cdots x_n, \theta_n)] \\ &\qquad = \lim_{n\to\infty} [l_\pi(x_1, x_2, \cdots x_n)]/[l_{u_0}(x_1, x_2, \cdots x_n)] > 0. \end{aligned}$$

Then by considering the ratio of corresponding terms, we have the required result:

$$\lim_{n\to\infty} [l_\pi(\theta_n \mid x_1, x_2, \cdots x_n)]/[l_{u_0}(\theta_n \mid x_1, x_2, \cdots x_n)] = 1.$$

One consequence of the theorem concerns the uniqueness of the invariant conditional density. Let $m^*(\theta)$ be any other invariant function and let $c^*(u)$, $\lambda^*(u)$ be the corresponding functions of u, satisfying Equation (6.1).

COROLLARY 6.4. *Suppose, that for some constant d, the ratio $m^*(\theta)e^{d\theta}/m(\theta)$ is bounded and continuous in θ. Then there exists a constant d^* such that $m^*(\theta)e^{d^*\theta}/m(\theta)$ does not depend on $\theta \varepsilon \Theta_m$ and $c^*(u + d^*) = c(u) + d^*$; $u \varepsilon U$.*

PROOF. Set $\pi(\theta) = m^*(\theta) \exp\{(u_0 + d)\theta\}$ for some $u_0 \varepsilon U$ and $u_0^* = u_0 + d$, $u_n^* = c^*(u_{n-1}^*) + x_n$; $n \geqq 1$. Then for each $\theta, \theta' \varepsilon \Theta_m$ and almost every $\{x_n\}$, we have

$$\lim_{n\to\infty} [m^*(\theta)/m(\theta)] \exp\{(u_n^* - u_n)\theta - (\lambda^*(u_n^*) - \lambda(u_n))\} = 1$$

and similarly when θ is replaced by θ'. Hence

$$\lim_{n\to\infty} \exp\{(u_n^* - u_n)(\theta' - \theta)\} = [m^*(\theta)/m(\theta)][m(\theta')/m^*(\theta')],$$

and $(u_n^* - u_n)$ converges to a limit d^* which does not depend on θ or θ'. The first assertion follows immediately and $c^*(u)$ is then determined by the properties of $m(\theta)$ and $c(u)$.

In a practical decision problem such as we have imagined underlying this investigation, although an approximate knowledge of the values of the conditional density function at each stage of the process $\{\theta_n\}$ is important, we may be more interested in a particular event $\theta_n \varepsilon B$, where B is a fixed subset of Θ_m. The convergence discussed in Theorem 6.3 is not established uniformly in θ, so that we cannot deduce the analogous result for the probability of a collection of states. However, the result can be obtained by repeating the same argument, at least when the set B is bounded.

THEOREM 6.5. *Under the conditions of the previous theorem, for any bounded event $B \subset \Theta_m$ and almost every sequence $\{x_n\}$:*

$$\lim_{n\to\infty} [P_\pi(\theta_n \varepsilon B \mid x_1, x_2, \cdots x_n)/P_{u_0}(\theta_n \varepsilon B \mid x_1, x_2, \cdots x_n)] = 1.$$

PROOF. This follows by considering the event $\theta_n \varepsilon B$ instead of the single state $\theta_n = \theta$, throughout the proof of Theorem 6.3. In Equation (6.4), the critical last term becomes

$$[\textstyle\int_B m(\theta_n) \exp(u_n'\theta_n - \lambda(u_n'))\, d\theta_n]/[\textstyle\int_B m(\theta_n) \exp(u_n\theta_n - \lambda(u_n))\, d\theta_n]$$

and provided that this converges to 1, the argument leads on to the required result. Let $\delta_n = u_n' - u_n$. Then $\lim_{n\to\infty} \delta_n = 0$ and $\lim_{n\to\infty} (\lambda(u_n') - \lambda(u_n)) = 0$. It is sufficient to verify that

$$\lim_{n\to\infty} [\int_B m(\theta)e^{u_n\theta}\{e^{\delta_n\theta} - 1\}\, d\theta/\int_B m(\theta)e^{u_n\theta}\, d\theta] = 0,$$

for every sequence $\{u_n\}$ with $u_n \varepsilon U$ and complex numbers $\{\delta_n\}$ such that $\lim_{n\to\infty} \delta_n = 0$. But since B is bounded, it follows that $\lim_{n\to\infty} (e^{\delta_n\theta} - 1) = 0$ uniformly in $\theta \varepsilon B$, and the proof is complete.

It would be preferable to extend this result to any Borel set within the effective state space. In fact, the theorem remains true in the case when B is a semi-infinite interval, but its proof is more awkward. However, it is not clear whether the corresponding general result can be established without some strengthening of Assumption (a).

In conclusion, we consider very briefly, how the theory developed here applies to the examples in Section 5.

Condition (a) is not difficult to verify, where it is valid. In the case of linear connecting functions, $u_n' - u_n = c^n(u_0' - u_0)$ and this tends to zero if and only if $|c| < 1$. It remains to see whether $\lim_{n\to\infty} (\lambda(u_n') - \lambda(u_n)) = 0$. In Example (i), $\lambda(u) \propto u^2$ and it is sufficient if $\lim_{n\to\infty} c^n u_n = 0$. But we can express $c^n u_n = c\alpha c^{n-1} u_{n-1} + c^n \eta_n$, where $\{\eta_n\}$ consists of independent and identically distributed normal variates, and $|c\alpha| < 1$ follows from the quadratic equation for c. For Example (ii), we have $\lambda(u) \propto \log u$ and hence it is sufficient to show that $\lim_{n\to\infty} (u_n'/u_n) = 1$ or $\lim_{n\to\infty} (c^n/u_n) = 0$. But when $c < 1$, this follows from the behaviour of the process $\{u_n^{-1}\}$. The condition is slightly more difficult to verify in Example (iv). In this case $u_n = u_{n-1}^{\frac{1}{2}} + x_n$, $u_n' = (u_{n-1}')^{\frac{1}{2}} + x_n$, where the square roots are taken with non-negative real parts. It is necessary to show that $\lim_{n\to\infty} (u_n' - u_n) = 0$ and $\lim_{n\to\infty} (u_n'/u_n) = 1$. However, since the observations are always positive, $\{u_n\}$ is bounded away from zero and only the first limit need be considered. To establish this, we observe that in general

$$|u_{n+1}' - u_{n+1}| = |((u_{n-1}')^{\frac{1}{2}} + x_n)^{\frac{1}{2}} - ((u_{n-1})^{\frac{1}{2}} + x_n)^{\frac{1}{2}}| \leqq |(u_{n-1}')^{\frac{1}{4}} - (u_{n-1})^{\frac{1}{4}}|.$$

Since u_{n-1} and the real part of u_{n-1}' are non-negative, the maximum occurs here when $x_n = 0$. Thus, at worst, we must account for a sequence of zero observations. But in this case $\lim_{n\to\infty} u_n' = \lim_{n\to\infty} u_n = 1$ and the differences obviously converge to zero.

In these examples, where Condition (a) is satisfied, it is clear that (c) also holds. The only case for which there is a non-trivial sequence of sets $\{B_n\}$, of the type envisaged, is Example (ii) and for this, (c_1) is valid when $c < 1$. Condition (b) however is much more restrictive, although it may be possible to relax its demands in specific applications. It means roughly, apart from continuity considerations, that the initial state θ_0 must not be defined with either too much or too little precision. The properties of the normal example mentioned in the introduction, suggest however, that neither demand is completely necessary. A normal initial distribution with variance $(1 - c)v_0$, always leads to normal conditional distributions whose variances converge, whether or not the inequality $0 < v_0 < 1$, required by Assumption (b) is satisfied. But it should be noted that here we have a very special type of convergence, involving a sequence of connecting functions all of which are linear. In fact this can only occur when the transition density is of exponential type, as well as all the other distributions.

REFERENCES

[1] BATHER, J. A. (1963). Control charts and the minimization of costs. *J. Roy. statist. Soc. Ser. B*, **25** 49–80.
[2] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.