

INVERSE PROBABILITY AND EVOLUTIONARY MARKOFF STOCHASTIC PROCESSES*

BY ALLADI RAMAKRISHNAN

(*Department of Physics, University of Madras*)

Received February 19, 1955

INTRODUCTION

THE subject of inverse probability has unfortunately been regarded with general suspicion and so not many attempts have been made to apply the theory even to cases where unambiguous answers can be given, on a proper mathematical formulation of the problems. It is the object of this paper to emphasise that in the case of evolutionary stochastic processes, the application of Bayes' theorem yields results as meaningful and rigorous as could be obtained by the application of the concepts of ordinary (as contrasted with inverse) probability theory.

We shall first formulate the basic problem of inverse probability in relation to univariate stochastic processes and point out why it was erroneously supposed to be a difficult problem. We shall then show how the problem can be solved rigorously by reconsidering the Chapman-Kolmogoroff equation and consequently the stochastic differential equations derived therefrom, in a new light.

2. FORMULATION OF THE PROBLEM

Let us consider a stochastic process representing the 'evolution' in a Markovian manner with respect to the one-dimensional parameter t of the probability distribution of a continuous stochastic variate $x(t)$. We are thus interested in the probability frequency function (p.f.f.) $\pi(x_1 | x_0; t_1, t_0) dx_1$ representing the probability that $x(t_1)$ lies between x_1 and $x_1 + dx_1$ given that $x(t_0) = x_0$, ($t_1 > t_0$).

It shall be assumed that the process is homogeneous in t and so π is a function only of the interval $t_1 - t_0$ and not t_1, t_0 severally. Though it is customary to set $t_0 = 0$ and write $t_1 - t_0$ as t_1 , we shall deliberately refrain from doing so, for reasons which will be clear presently. On the contrary we shall use the notation $\pi(x; t)$ to represent the p.f.f. of $x(t)$ when we do

* Read at the Annual Meeting of the Indian Academy of Sciences at Belgaum, December, 1954.

not specify the "initial" conditions, *i.e.*, the p.f.f. at some t' , $(t' < t)$ †. Thus by definition we immediately write

$$\pi(x_1; t_1) = \int_{x_0} \pi(x_0; t_0) \pi(x_1 | x_0; t_1, t_0) dx_0, \quad (t_0 < t_1). \quad (1)$$

When we deal with distributions at two different values of t , say t_1 and t_0 , $(t_1 > t_0)$, we shall call the distribution at t_1 the emergent spectrum with respect to the injected spectrum at t_0 , and $\pi(x_1 | x_0; t_1, t_0)$ the fundamental transition probability for the finite interval $t_1 - t_0$. It is quite clear that $\pi(x_1 | x_0; t_1, t_0)$ is only a particular case of $\pi(x_1; t_1)$ where the injection spectrum at t_0 is given by $\pi(x; t_0) = \delta(x - x_0)$ where δ is the Dirac delta function. π satisfies the Chapman-Kolmogoroff equation,

$$\pi(x; t_1) = \int_{x'} \pi(x'; t_0 + \tau) \pi(x | x'; t_1, t_0 + \tau) dx' \quad (2)$$

for all values of τ such that $t_0 < t_0 + \tau - t_1$ and in particular

$$\pi(x_1 | x_0; t_1, t_0) = \int_{x'} \pi(x' | x_0; t_0 + \tau, t_0) \pi(x_1 | x'; t_1, t_0 + \tau) dx'. \quad (3)$$

By making $t_0 + \tau \rightarrow t_1$, we obtain the well-known forward differential equation of Kolmogoroff for π provided we know the limit of $\pi(x_1 | x'; t_1, t)$ as $t_1 - t \rightarrow +0$, *i.e.*, to speak in physical terms we know what 'happens' to the stochastic variable in an infinitesimal interval Δ .

From equation (1) we find that the emergent p.f.f. $\pi(x_1; t_1)$ can be computed if we know the injected p.f.f. $\pi(x_0; t_0)$ and the fundamental transition probability $\pi(x_1 | x_0; t_1, t_0)$. But the computation of $\pi(x_0; t_0)$ given $\pi(x_1; t_1)$ seems at first a formidable task since it involves the solution of equation (1) which is a Volterra integral equation of the first kind. It is the principal object of the present paper to show that no complex procedure is necessary to solve for the injected spectrum *in view of the peculiar properties of the fundamental transition probability*. We shall prove by very simple arguments that the inversion of (1) is given by

$$\pi(x_0; t_0) = \int_{x_1} \pi(x_1; t_1) \hat{\pi}(x_0 | x_1; t_1, t_0) dx_1, \quad (t_1 > t_0), \quad (4)$$

where $\hat{\pi}(x_0 | x_1; t_1, t_0)$ is a function obtained by replacing $t_1 - t_0$ by the negative quantity $t_0 - t_1$, in $\pi(x_0 | x_1; t_0, t_1)$ which we remember is only a function of $t_1 - t_0$ and not t_1, t_0 severally. $\hat{\pi}$ unlike π has no probability significance and is only a functional operator.

† Throughout this paper π denotes a p.f.f. Distinction between different distributions will be obvious from the variable used to define π .

In equation (1) since $\pi(x; t)$ and $\pi(x_1 | x_0; t_1, t_0)$ are both probability densities, they are positive and so $\pi(x; t_1)$ exists for any $t_1 > t_0$, i.e., $t_1 - t_0$ is positive and finite and can be chosen as large as we please. If $\pi(x | x_0; t_1, t_0)$ tends to a stationary distribution as $t_1 - t_0 \rightarrow +\infty$, so also does $\pi(x_1; t_1)$.

On the contrary in (4) since $\hat{\pi}$ is not a p.f.f. $\pi(x; t_0)$ does not remain a p.f.f. for all values of $t_0 < t_1$. It shall be proved that there exists a t_p such that $\pi(x_0; t_0)$ is a p.f.f. in the domain $t_p \leq t_0 \leq t_1$ and the process cannot be traced back to a point $t < t_p$.

3. SOLUTION OF THE PROBLEM

It is necessary to recall the method of derivation of the forward differential equation for $\pi(x; t)$ from the Chapman-Kolmogoroff equation. In particular, we first consider the simple case when

$$\pi(x_1 | x_0; t_1, t_0) \rightarrow R(x_1 | x_0) \Delta + \delta(x_1 - x_0) (1 - \Delta \int_{x'} R(x' | x_0) dx') \quad (5)$$

as $t_1 - t_0 \rightarrow \Delta \rightarrow +0$.

In such a case, (1) assumes the form

$$\begin{aligned} \pi(x; t_0 + \Delta) &= \Delta \int_{x_0} \pi(x_0; t_0) R(x | x_0) dx_0 \\ &+ \pi(x; t_0) \{1 - \Delta \int_{x'} R(x' | x) dx'\} + O(\Delta^2). \end{aligned} \quad (6)$$

(6) can be reduced to a differential equation in t by making $\Delta \rightarrow +0$. From (6) $\pi(x; t_0 + \Delta)$ can be computed if $\pi(x; t_0)$ is known, and thus the emergent p.f.f. $\pi(x_1; t_1)$ is obtained from $\pi(x_0; t_0)$ by a process of integration, i.e., by solving the differential equation in t .

We notice we can write (6) in the following form:

$$\begin{aligned} \pi(x; t_0 + \Delta) &= \Delta \int_{x_0} \pi(x_0; t_0 + \Delta) R(x | x_0) dx_0 \\ &+ \pi(x; t_0) - \pi(x; t_0 + \Delta) \cdot \Delta \int_{x'} R(x' | x) dx' + O(\Delta^2) \end{aligned} \quad (7)$$

which can be re-written in the equivalent form (8) by using the device of writing $t_0 - \Delta$ and t_0 for t_0 and $t_0 + \Delta$ respectively.

$$\begin{aligned} \pi(x; t_0 - \Delta) &= \pi(x; t_0) + \Delta \pi(x; t_0) \int_{x'} R(x' | x) dx' \\ &- \Delta \int_{x_0} \pi(x_0; t_0) R(x | x_0) dx_0. \end{aligned} \quad (8)$$

In this form we immediately realise $\pi(x; t_0 - \Delta)$ can be obtained from $\pi(x; t_0)$ by merely replacing Δ by $-\Delta$ in $\pi(x_0; t + \Delta)$. Thus $\pi(x_0; t_0)$ can be obtained by a process of integration from $\pi(x_1; t_1)$, ($t_1 > t_0$).

Considering processes more general than that defined by (1), Bartlett (see Ramakrishnan, 1952) has observed that the forward differential equation for π can be written in the *generalised symbolic form*,

$$\frac{\partial \pi(t_0 + \tau)}{\partial \tau} = F\{\pi(t_0 + \tau)\}, \quad (9)$$

(τ positive) omitting for the moment the argument x to indicate that the right-hand side may involve several values of the argument. If $\pi(t_0)$ is a p.f.f. $\pi(t_0 + \tau)$ is a p.f.f. for all τ from 0 to $+\infty$. Noting that the above equation is derived from the following

$$\pi(t_0 + \tau + \Delta) = \pi(t_0 + \tau) + F\{\pi(t_0 + \tau)\} \cdot \Delta + O(\Delta^2), \quad (10)$$

we re-write it in the form

$$\pi(t_0 + \tau + \Delta) = \pi(t_0 + \tau) + F\{\pi(t_0 + \tau + \Delta)\} \cdot \Delta + O(\Delta^2). \quad (11)$$

Therefore using the same device as before of writing $t_0 - \overline{\tau + \Delta}$ and $t_0 - \tau$ in the place of $t_0 + \tau$ and $t_0 + \tau + \Delta$,

$$\pi(t_0 - \overline{\tau + \Delta}) - \pi(t_0 - \tau) = -\Delta F\{\pi(t_0 - \tau)\}$$

or

$$\frac{\partial \pi(t_0 - \tau)}{\partial \tau} = -F\{\pi(t_0 - \tau)\}. \quad (12)$$

In other words $\pi(t_0 - \tau)$ is obtained from $\pi(t_0)$ by merely replacing τ by $-\tau$ in $\pi(t_0 + \tau)$.

It now remains to show that this process of inversion, *i.e.*, determining the p.f.f. at $t_0 - \tau$ from the p.f.f. at t_0 cannot be carried on for all values of τ except in the trivial case when the emergent distribution at t_0 is a stationary distribution.

To do this we first consider the simpler case defined by (5) and examine the essential difference between (6) and (7). Noting that $R(x|x_0)$ and $\pi(x; t_0)$ are non-negative in (6), we find that $\pi(x; t_0 + \Delta)$ is also non-negative. On the contrary $\pi(x; t_0 - \Delta)$ will be negative for such values of x for which $\pi(x; t_0) = 0$ if $R(x|x')$ is positive for some value of x' . In such a case $\pi(x; t_0)$ cannot be an 'emergent' distribution, *i.e.*, it cannot be traced back. *Thus there exists a smallest value τ_p for τ such that $\pi(x; t_0 - \tau_p)$*

is zero for some value of x . Thus $t_p = t_0 - \tau_p$ is the earliest point to which we can trace back the process. Identical arguments apply to a π function satisfying Bartlett's generalised symbolic equation. We call t_p the 'primitive origin' of the process and τ_p the 'age' of the process at t_0 . The distribution at any point $t_0 - \tau$, $0 < \tau < \tau_p$ is given by equation (12), i.e., if we inject $\pi(x; t_0 - \tau)$ satisfying (12), then it will emerge as $\pi(x; t_0)$ at t_0 . It also follows that $\pi(x; t_0 - \tau_1)$, the injection spectrum at $t_0 - \tau_1$, is the emergent spectrum with respect to $\pi(x; t_0 - \tau_2)$ the corresponding injection spectrum at $t_0 - \tau_2$, ($\tau_2 > \tau_1$). $\pi(x; t - \tau)$ satisfies the modified Chapman-Kolmogoroff equation

$$\pi(x; t_0 - \tau) = \int_{x'} \pi(x'; t_0 - \tau') \pi(x|x'; t_0 - \tau, t_0 - \tau') dx', \quad (13)$$

where

$$t_0 - \tau_p < t_0 - \tau' < t_0 - \tau$$

i.e.,

$$\tau_p > \tau' > \tau.$$

The distribution at $t_0 - \tau_p$ can only be an injected distribution and cannot be the emergent distribution of a process started 'earlier'. Therefore we call it the primitive *a priori* spectrum.

Illustration in the case when $x(t)$ is discrete.—We shall now consider the case when $x(t)$ is discrete and can assume mutually exclusive values $x_1, x_2, \dots, x_j, \dots$ and define $\pi(j; t)$ as the probability (not probability density) that $x(t) = x_j$. If $\pi(j|i; t_1, t_0) \rightarrow R(j|i) \cdot \Delta$ as $t_1 - t_0 \rightarrow \Delta \rightarrow +0$ it is well known that π satisfies the matrix differential equation [cf. integro-differential equation when $x(t)$ is continuous]

$$\frac{d\vec{\pi}(j; t)}{dt} = [R] \vec{\pi}(j; t), \quad (14)$$

where $\vec{\pi}(j; t)$ is a column vector with components corresponding to different values of j and $[R]$ is the matrix with non-diagonal elements $R(j|i)$ and the diagonal elements defined by $R(i|i) = - \sum_j R(j|i)$. The solution of the above equation is given by

$$\vec{\pi}(j; t_1) = e^{[R](t_1 - t_0)} \vec{\pi}(j; t_0). \quad (15)$$

By considering the infinitesimal transformation $\pi(j; t)$ to $\pi(j; t + \Delta)$ we note that if, $\pi(j; t) \geq 0$ for all j so is $\pi(j; t + \Delta)$. Thus $\pi(j; t_1) \geq 0$ for all j for all $t_1 > t_0$ provided $\pi(j; t_0) \geq 0$ for all j .

By arguments now familiar from the previous section $\pi(j; t - \Delta)$ can be obtained from $\pi(j; t)$ and so

$$\vec{\pi}(j; t_0) = e^{-[R](t_1 - t_0)} \vec{\pi}(j; t_1), \quad (t_1 > t_0) \quad (16)$$

i.e., given the emergent distribution $\vec{\pi}(j; t_1)$ we compute the injected distribution at t_0 . This computation can be done for t_0 in the domain $t_1 - \tau_p \leq t_0 \leq t_1$. At $t_0 = t_1 - \tau_p$, $\pi(j; t_1 - \tau_p) = 0$ at least for one value of j .

To illustrate this we shall take a system capable of assuming one of four states, i.e., $j = 1, 2, 3, 4$. We shall take the following values for R :

$$\begin{pmatrix} R(1|1) & R(1|2) & R(1|3) & R(1|4) \\ R(2|1) & R(2|2) & R(2|3) & R(2|4) \\ R(3|1) & R(3|2) & R(3|3) & R(3|4) \\ R(4|1) & R(4|2) & R(4|3) & R(4|4) \end{pmatrix} = \begin{pmatrix} -12 & 0 & 0 & 0 \\ 1 & -8 & 0 & 0 \\ 7 & 6 & -4 & 0 \\ 4 & 2 & 4 & 0 \end{pmatrix}$$

We assume that at some t_1 , the values of $\pi(j; t_1)$ are as given below: then the primitive origin $t_1 - \tau_p$ is determined.

$$\left. \begin{array}{l} \pi(1; t_1) = \cdot 1, \pi(1; t_1 - \tau_p) = \cdot 3729 \\ \pi(2; t_1) = \cdot 1, \pi(2; t_1 - \tau_p) = \cdot 2074 \\ \pi(3; t_1) = \cdot 2, \pi(3; t_1 - \tau_p) = 0 \\ \pi(4; t_1) = \cdot 6, \pi(4; t_1 - \tau_p) = \cdot 4197 \end{array} \right\} \text{for } \tau_p = \cdot 1097$$

Conditional Inverse Probabilities.—We have observed that the p.f.f. $\pi(x_1 | x_0; t_1, t_0)$, $(t_1 > t_0)$ is a particular case of $\pi(x_1; t_1)$ with $\pi(x; t_0) = \delta(x - x_0)$. In this case the process cannot be traced back further than t_0 , i.e., to a point $t < t_0$, since $\pi(x; t_0) = 0$ for $x \neq x_0$. Hence $\pi(x; t_0 - \Delta)$ is negative (Δ positive) for $x \neq x_0$. Therefore a delta function can only be an injected spectrum and cannot be the emergent spectrum of a stochastic process. Thus $\hat{\pi}(x | x_0; t, t_0)$, $(t > t_0)$ has no probability interpretation.

On the other hand, in a previous paper, the author (1954) defined $P(x_0 | x_1; t_0, t_1)$ as the conditional inverse probability that $x(t_0)$ lies between x_0 and $x_0 + dx_0$ given that $x(t_1) = x_1$, $(t_0 < t_1)$. This has to be distinguished from $\hat{\pi}$. When in defining P we state that $x(t_1) = x_1$, we mean that the observed value of the emergent spectrum at t_1 is x_1 and not that the emergent spectrum is a delta function. In $\hat{\pi}$ we require the emergent spectrum to be a delta function and this is not possible. Hence $\hat{\pi}$ is not a probability magnitude.

In his previous contribution, the author showed that $P(x_0|x_1; t_0, t_1)$ cannot be obtained unless the injected spectrum at a certain $t < t_0$ is known. If that is known to be $\pi(x; t)$ then by considering the joint probability of obtaining $x(t) = x$, $x(t_0) = x_0$ and $x(t_1) = x_1$ and applying Bayes' theorem, it was shown that

$$\begin{aligned} P(x_0|x_1; t_0, t_1) &= \frac{\int \pi(x; t) \pi(x_0|x; t_0, t) \pi(x_1|x_0; t_1, t_0) dx}{\int \pi(x; t) \pi(x_1|x; t_1, t) dx} \\ &= \frac{\pi(x_0; t_0) \pi(x_1|x_0; t_1, t_0)}{\pi(x_1; t_1)}, \quad (t < t_0 < t_1) \end{aligned} \quad (17)$$

If $\pi(x; t)$ is known $\pi(x_0; t_0)$ and $\pi(x_1; t_1)$ can be computed from the fundamental transition probabilities corresponding to the intervals $t_0 - t$ and $t_1 - t_0$. Hence the method of computation of P described in that paper is equivalent to the present one.

In this paper we have shown that if $\pi(x; t)$ is known at some point, it can be extended to the entire permissible domain of t and so P can be computed if the distribution at some t , and of course the fundamental transition probability for a finite interval are known.

PROBLEMS INVOLVING A PROBABILITY FREQUENCY FUNCTION FOR THE PRIMITIVE ORIGIN

Till now we have assumed that the distribution at t is emergent with respect to that at $t - \Delta$ (except when t is the primitive origin) and injected with respect to that at $t + \Delta$. Or, in other words, given the distribution at t_0 , the injected distribution at $t < t_0$ is a function of $t_0 - t$. The primitive origin therefore is a definite point and is uniquely determined if we know the probability frequency function at t_0 .

We shall now consider the following class of problems. We are given the probability frequency function at t_0 . The primitive origin of the process lies between $t_0 - \tau$ and $t_0 - (\tau + d\tau)$ with probability $\eta(\tau) d\tau$. At this origin a p.f.f. $\phi(x)$ independent of τ is injected. Given $\eta(\tau)$ our problem is to determine the injection spectrum $\phi(x)$.

Using simple probability arguments we write

$$\pi(x; t_0) = \int_0^\infty d\tau \int_{x'} \phi(x') \pi(x|x'; t_0, \tau) \eta(\tau) dx' \quad (18)$$

Since $\pi(x; t_0)$ and $\pi(x|x'; t_0, \tau)$ are known, $\eta(\tau)$ or $\phi(x')$ can be determined if the other is known, by the classical theory of integral equations;

but the explicit solution of $\phi(x')$ or $\eta(t)$ amenable to numerical computation is a very difficult problem. On the contrary, the computation of π given ϕ is a much easier problem. Considering the very simple special case when x is a discrete variable representing a Poisson distribution

$$\pi(x|x'; t, t - \tau) = \frac{e^{-\lambda\tau} (\lambda\tau)^{x-x'}}{(x-x')!}$$

and is independent of t ,

$$\pi(x; t_0) = \int_0^\infty e^{-\lambda\tau} \frac{(\lambda\tau)^x}{x!} \eta(\tau) d\tau \quad (19)$$

As expected $\pi(x; t_0)$ is independent of t_0 since $\pi(x|x'; t, t - \tau)$ is only a function of τ . The computation of η given π is still a difficult problem.

Consael has discussed a class of distributions defined by

$$\pi(x; t) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^x}{x!} \eta(\lambda) d\lambda \quad (20)$$

The physical meaning of this class of distributions is obvious. The parameter λ of the Poisson distribution corresponding to an interval t is itself a stochastic variable with a p.f.f. $\eta(\lambda)$. Comparing (20) with (19), we find that all the results of Consael apply to processes defined by (19) provided we replace λ by t , since the Poisson distribution is symmetrical with respect to t and λ .

Problems involving a "switch-off" of the process.—Let t_0 be the primitive origin of a process where we inject the primitive p.f.f. $\phi(x)$. Let us assume that we "switch-off" the process between $t_0 + \tau$ and $t_0 + \tau + d\tau$ with probability $\eta(\tau) d\tau$. After "switch-off" the stochastic variable retains the value it had at τ till infinity. Thus the p.f.f. at $t_0 + \tau$ as $\tau \rightarrow \infty$ is given by

$$\pi(x'; \infty) = \int_x^\infty dx \int_0^\infty \phi(x) \pi(x'|x; t_0 + \tau, t_0) \eta(\tau) d\tau \quad (21)$$

Comparing this equation with (18) we find $\pi(x'; \infty)$ is identical with $\pi(x; t_0)$ of the previous problem. It is interesting to note that in the previous case the point of origin has a p.f.f. while here the point of switch-off has a p.f.f. and therefore the duration of the process has the same p.f.f. in both cases !

Incidentally the author wishes to mention that the process of acceleration of cosmic ray particles in the galactic magnetic field proposed by Fermi is a process which can be interpreted to belong to either of the two equivalent types discussed above.

‡ We assume, without loss of generality, that $x' = 0$.

Process involving "absorption".—Finally, we advert to an interesting problem in inverse probability occurring in physics (relating to the range of a fast particle passing through matter) which for some unknown reason has not been investigated in a rigorous manner.

A fast particle of initial energy E_0 passes through matter and loses energy through ionisation or radiation. When it reaches an energy zero or some critical value E_c we say the particle is 'absorbed' or stopped, and the distance t of penetration (considering the passage to be one-dimensional) till it is absorbed is called its range. The author has, in a separate contribution to this journal, derived an expression for $P(E_0; t) dt$, the probability that a particle of initial energy E_0 is absorbed between t and $t + dt$ if we know the fundamental transition probability $\pi(E|E_0; t, 0)$.

If we now assume that the initial energy of the particle is a stochastic variable defined by the p.f.f. $\phi(E_0)$, i.e., if $\phi(E_0)$ is the primitive injected spectrum, the p.f.f. of the range is given by

$$P(t) dt = \int_{E_0} \phi(E_0) P(E_0; t) dE_0 \quad (22)$$

Given $P(t)$ and $P(E_0; t)$ it is theoretically possible to compute the primitive injected spectrum $\phi(E_0)$ but this is not as easy as the computation of $\phi(E_0)$ if we are given the emergent spectrum at t .

There still remains the further question of conditional inverse probability. Given that the observed range was t_0 to compute the probability $\phi_{t_0}(E_0) dE_0$ that the initial energy of the particle lies between E_0 and $E_0 + dE_0$. By the simple application of Bayes' theorem we find $\phi_{t_0}(E_0)$ is given by

$$\phi_{t_0}(E_0) = \frac{\phi(E_0) P(E_0; t_0)}{P(t)} = \frac{\phi(E_0) P(E_0; t_0)}{\int \phi(E_0) P(E_0; t) dE_0} \quad (23)$$

It is clear that the conditional inverse probability cannot be obtained unless we know $P(t)$ and $P(E_0; t)$ or $\phi(E_0)$ and $P(E_0; t)$. In the opinion of the author, the function $\phi_{t_0}(E_0)$ is physically more important than $P(E_0; t)$ since in many experiments the energy of the particle is estimated from the range and somehow this aspect seems to have attracted little attention till now.

REFERENCES

Alladi Ramakrishnan	.. <i>Proc. Camb. Phil. Soc.</i> , 1952, 48 , 451.
Consael, R.	.. <i>Bulletin de la class. des Sciences; Acad. Roy. de Belg.</i> , 1952, 38 , 442.
Alladi Ramakrishnan and Mathews, P. M.	<i>Proc. Ind. Acad. Sci.</i> , 1953, 38A , 450,
Alladi Ramakrishnan	.. <i>Ap. J.</i> , 1954, 119 , 443.