Information Processing for Observed Jump Processes*

DONALD L. SNYDER

Biomedical Computer Laboratory, Washington University School of Medicine, St. Louis, Missouri 63110

An equation is derived for the posterior statistics of a Markov process that modulates the transition rates of an observed jump process. This equation forms the basis of a nonlinear filtering theory for observed jump processes that is the counterpart of the Stratonovich–Kushner filtering theory for nonlinear observations in additive white Gaussian noise.

I. INTRODUCTION

In Snyder (1970a, b) we have given an equation for the posterior statistics of an information signal that modulates the intensity function of an observed doubly stochastic Poisson counting process. This equation forms the basis of a nonlinear filtering theory for Poisson observations that is remarkably parallel to that of Stratonovich (1960) and Kushner (1964) for nonlinear observations of the information signal in an additive white Gaussian noise.

The results in Snyder (1970a, b) have been generalized in two ways. Frost (1971a, b) takes the observation process to have independent increments conditioned on the information signal; thus, the Stratonovich–Kushner filtering equation and the equation in Snyder (1970b) obtain as special cases of Frost's equation as both the Wiener and Poisson processes have independent increments. Rubin (1971) and later Snyder (1971) take the observation process to be the counting process associated with a self-exciting point process conditioned on the information signal (i.e., the intensity is a function of the counting history as well as the information signal.) The equation of Snyder (1970b) obtains from Rubin's equation when the point process is not self-exciting.

In this paper, we offer a third generalization. Here we take the observation

* This work was supported by the National Institutes of Health Research Grant No. RR 00396 from the Division of Research Resources and Grant No. GK-32239 from the National Science Foundation.

Copyright © 1973 by Academic Press, Inc.
All rights of reproduction in any form reserved.
process to be a self-exciting jump process conditioned on the information signal. This process can change by other than the unit positive jumps of the counting processes in Snyder (1970b, 1971) and Rubin (1971) which are included here as special cases; in addition, Markov jump processes, semi Markov jump processes (see Cox and Lewis, 1966, p. 194), birth-death processes, age-dependent branching processes (see Bharucha-Reid, 1960, p. 94), and many other processes commonly used as models for jump phenomena are also included, where each is conditioned on the information signal. Yashin (1970) takes the observation process to be a Markov jump process conditioned on an information signal that is also a Markov jump process. This is generalized here to include jump observation processes that are not conditionally Markov and to include continuous-state Markov processes as the information signal.

II. Model Formulation and Problem Statement

Denote by $\mathcal{S}$ the denumerable state space associated with the continuous-time jump process $\{z_t ; t > t_0\}$ and by $z_{t_0,t}$, paths of $z$ on $(t_0, t]$. The paths $z_{t_0,t} = \{(x_\sigma, \sigma); t_0 < \sigma \leq t\}$ are piecewise constant with randomly occurring jumps between the states $\xi_i$ of $\mathcal{S}$. We assume that the following limits exist (a.s.):

$$\lim_{\Delta t \to 0} (\Delta t)^{-1} \Pr(z_{t+\Delta t} = \xi_j | z_t = \xi_i, z_{t_0,t}, x_0 ; t < \sigma \leq t + \Delta t) = \lambda_{i,j}(t, z_{t_0,t}, x_t), \quad i \neq j$$

(1)

$$\lim_{\Delta t \to 0} (\Delta t)^{-1} [1 - \Pr(z_{t+\Delta t} = \xi_i | z_t = \xi_\zeta, z_{t_0,t}, x_\sigma ; t < \sigma \leq t + \Delta t)]$$

$$= \sum_{\xi_j \neq \xi_i} \lambda_{i,j}(t, z_{t_0,t}, x_t)$$

(2)

for $t > t_0$, where $\{x_t ; t > t_0\}$ is an information signal to be defined. The function $\lambda$ in (1) is the instantaneous rate, or intensity, of jumps from state $\xi_i$ to state $\xi_j$. The sum on $j$ of $\lambda$ for $i \neq j$ in (2) is simply the total instantaneous rate of jumps from state $\xi_i$. We have assumed that these intensities are known, nonnegative functions of time $t$, of the previous path $z_{t_0,t}$, and of an information signal $x_t$; thus, $z$ can be termed an inhomogeneous ($t$ dependent), self-exciting ($z_{t_0,t}$ dependent), doubly stochastic ($x_t$ dependent) jump process.
It is convenient to associate with \( \{z_t ; t > t_0\} \) an integer-valued counting process \( \{N_t ; t > t_0\} \) with unit jumps at each jump of \( z \). This can be introduced in the following way. A *marked point process* is a point process with an auxiliary variable, called a mark, associated with each point [see Konig and Matthes (1963), and Daley and Vere-Jones (1971).] Thus, an equivalent way to think of \( z \) is as a marked point process in which the marks indicate the state transitions in \( \mathcal{Z} \). Assign \( N \) to count these points regardless of their mark. The point processes of Snyder (1970b, 1971) and Rubin (1971) are here generalized to include these marks.

Some examples of the type of observation processes included in our model are the following:

**Example 1.** The doubly stochastic Poisson processes of Snyder (1970a, b) occur when

\[
\lambda_{ij}(t, z_{t_0, t}, x_t) = \begin{cases} 
\lambda_i(x_t), & j = i + 1, \\
0, & \text{otherwise,}
\end{cases}
\]

(3)

and \( \mathcal{Z} \) is the set of nonnegative integers.

**Example 2.** The self-exciting point processes of Rubin (1971) and Snyder (1971) occur when

\[
\lambda_{ij}(t, z_{t_0, t}, x_t) = \begin{cases} 
\lambda_i(z_{t_0, t}, x_t), & j = i + 1, \\
0, & \text{otherwise}
\end{cases}
\]

(4)

and \( \mathcal{Z} \) is the set of nonnegative integers.

**Example 3.** Self-exciting birth–death processes occur when

\[
\lambda_{ij}(t, z_{t_0, t}, x_t) = \begin{cases} 
\lambda_i(z_{t_0, t}, x_t), & j = i + 1 \\
\mu_i(z_{t_0, t}, x_t), & j = i - 1, \\
0, & \text{otherwise}
\end{cases}
\]

(5)

and \( \mathcal{Z} \) is the set of integers.

**Example 4.** Markov jump processes occur when

\[
\lambda_{i, i'}(t, z_{t_0, t}, x_t) = \mu_{i, i'}(t, x_t),
\]

that is, when transitions depend only on \( z_t \) and not the entire path \( z_{t_0, t} \).
Example 5. A semi-Markov process is a generalized form of renewal process. Let the time \( u \) between a transition of \( z \) from state \( \zeta_i \) to the succeeding state \( \zeta_j \) have distribution \( P_{ij}(U \mid \mathbf{x}) \) conditioned on information variables \( \mathbf{x} \). Let \( p_{ij}(U \mid \mathbf{x}) = \partial P_{ij}(U \mid \mathbf{x})/\partial U \) be the density of \( u \) given \( \mathbf{x} \). Then an easy calculation shows

\[
\lambda_{t, e}(t, z_{t_0, t}, \mathbf{x}) = p_{ij}(t - W_{N_i} \mid \mathbf{x})[1 - P_{ij}(t - W_{N_i} \mid \mathbf{x})]^{-1},
\]

where \( N \) is the counting process introduced above and \( W_{N_i} \) is the observed waiting time to the \( N_i \)-th point of the marked point process.

We assume the information signal \( \{x_t; t > t_0\} \) is an \( n \)-vector Markov process exactly as in Snyder (1970b). See the discussion there for details.

The problem is to estimate \( x_t \) at time \( t \) in terms of an observed path \( z_{t_0, t} \) and to investigate how the estimate evolves with increasing \( t \). This is done in the next section where a representation is given for the conditional characteristic function of \( x_t \) given \( z_{t_0, t} \).

III. Equations of Filtering

Let \( c_t(v \mid z_{t_0, t}) \) be the posterior characteristic function for \( x_t \) given an observed path \( z_{t_0, t} \),

\[
c_t(v \mid z_{t_0, t}) = E\{\exp(j\langle v, x_t \rangle) \mid z_{t_0, t}\},
\]

where \( \langle \cdot, \cdot \rangle \) denotes the vector inner, or dot, product. Assume that

(a) \( (At)^{-1} E\{\exp[j\langle v, Ax_t \rangle] - 1 \mid z_{t_0, t}, x_t\} \leq g(v; t, z_{t_0, t}, x_t) \),

where \( E(||g||) < \infty \);

(b) \( p \lim_{At \to 0} (At)^{-1} E\{\exp[j\langle v, Ax_t \rangle] - 1 \mid z_{t_0, t}, x_t\} = \psi_t(v \mid z_{t_0, t}, x_t) \).

Then we have the following representation for \( c_t(v \mid z_{t_0, t}) \):

\[
dc_t(v \mid z_{t_0, t}) = E\{\exp(j\langle v, x_t \rangle) \psi_t(v \mid z_{t_0, t}, x_t) \mid z_{t_0, t}\} dt
\]

\[
- E\left\{\exp(j\langle v, x_t \rangle) \sum_{\zeta_k \neq \zeta_l} (\lambda_{z_l, z_{l+dt}} - \hat{\lambda}_{z_l, z_{l+dt}}) \mid z_{t_0, t}\right\} dt
\]

\[
+ E\{\exp(j\langle v, x_t \rangle)(\lambda_{z_l, z_{l+dt}} - \hat{\lambda}_{z_l, z_{l+dt}}) \mid z_{t_0, t}\} \hat{\lambda}_{z_l, z_{l+dt}}^{-1} dN_t
\]

\[
c_{t_0}(v \mid z_{t_0, t_0}) = E\{\exp(j\langle v, x_0 \rangle)\},
\]
OBSERVED JUMP PROCESSES

where

(a) The left side has the interpretation

\[ dc_i(v \mid z_{t_0,i}) = \lim_{\Delta t \to 0} \left[ c_{t+\Delta t}(v \mid z_{t_0,i}, \Delta z_i) - c_i(v \mid z_{t_0,i}) \right] \]

where \( \Delta z_i = z_{t+\Delta t} - z_i \);

(b) \( \hat{\lambda}_{z_t,z_{t+\Delta t}} = E\{\lambda_{z_t,z_{t+\Delta t}}(t, z_{t_0,i}, x_t) \mid z_{t_0,i}\} \) is the causal MMSE estimate of the transition rate from the state \( z_i \) at \( t \) to state \( z_{t+\Delta t} \) at \( t + \Delta t \);

(c) \( \sum_{k \neq i} (\lambda_{z_i,z_k} - \hat{\lambda}_{z_i,z_k}) \) is the error in estimating the rate of transitions from state \( z_i \) at \( t \) in terms of \( z_{t_0,i} \);

(d) \( dN_t = \lim_{\Delta t \downarrow 0} [N_{t+\Delta t} - N_t] \) is one or zero corresponding to whether or not a transition in \( z \) occurs in \( (t, t + \Delta t] \) as \( \Delta t \downarrow 0 \);

(e) The last term is zero when \( dN_t = 0 \);

(f) \( x_0 = x_{t_0} \) is the initial state of \( x \).

Equation (8) is our basic equation for filtering of jump processes of the type described. Its use in deriving various other filtering equations, in deriving approximate filtering equations, and in solving decision problems parallels that in Snyder (1970b) for what is now the special case of doubly stochastic Poisson processes.

The procedure for reducing (8) to the filtering equations in Snyder (1970b, 1971) and Rubin (1971) is clear. For illustration, we develop Yashin’s (1970) equation in somewhat more general form. For this purpose, suppose \( x_t \) is a real valued Markov jump process \( x_t \) with future independent of past \( z \) and with a denumerable state space with states \( \{\theta_i\} \). Following the notation of Bharucha-Reid (1960, Section 2.3), we let the matrix of infinitesimal transition rates have elements \( a_{ik}(t) \), where \( a_{ii}(t) = -\sum_{k \neq i} a_{ik}(t) \), and where

\[ \Pr(x_{t+\Delta t} = \theta_k \mid x_t = \theta_i) = a_{ik}(t) \Delta t + o(\Delta t), \quad i \neq k \]

in which \( o(\Delta t)/\Delta t \to 0 \) as \( \Delta t \to 0 \). Straightforward calculation shows

\[ W_t(v \mid z_{t_0,t}, x_t = \theta_i) = \exp(-jv\theta_i) \sum_k a_{ik}(t) \exp(jv\theta_k). \quad (9) \]

Let \( \pi_t(t) = \Pr(x_t = \theta_i \mid z_{t_0,t}) \) be the posterior probability that \( x_t \) is in state
\[ d\pi_i(t) = \sum_k \pi_k(t) a_{ki}(t) dt + \pi_i(t) \left[ \lambda_{\pi_i, \pi_i+1} - \hat{\lambda}_{\pi_i, \pi_i+1} \right] dN_i \]

This representation for the time evolution of \( \pi_i \) reduces to Yashin's (1970, note 2) equation when \( \lambda \) has the form in Example 4 above, so that \( z \) is a conditional Markov jump process. The left side with the first term on the right side of (10) describes the time evolution of the unconditional state probabilities of \( x \). These terms correspond to the forward Kolmogorov differential equation [see Bharucha-Reid, 1960, Eq. (2.21)] for

\[ \Pr(x_t = \theta_i) = \sum_j \Pr(x_t = \theta_i | x_r = \theta_j) \Pr(x_r = \theta_j). \]

As a further illustration, suppose that \( z_t \) is a two-state process with states \( \xi_0 = 0 \) and \( \xi_1 = 1 \) and with transition rates \( \lambda_{0,1}(t, z_{t_0}, t, x_t) \) and \( \lambda_{1,0}(t, z_{t_0}, t, x_t) \). Let \( x_t \) be a Markov diffusion with differential equation

\[ dx_t = f_t(x_t) dt + G_t(x_t) dB_t, \]

where \( B_t \) is a standardized Wiener process with future independent of past \( z \). Then

\[ \Psi_t(v | z_{t_0}, x_t) = J(v, f_t(x_t)) - \frac{1}{2} \langle v, G_t(x_t) G_t'(x_t) v \rangle. \]

Using this expression for \( \Psi \) in (8) and inverse Fourier transforming results in the following equation for the posterior probability density \( p_t(x | z_{t_0}, t) \) of \( x_t \) given \( z_{t_0}, t \)

\[ dp_t(x | z_{t_0}, t) = \left. L[p_t(x | z_{t_0}, t)] \right| dt \]

\[ + p_t(x | z_{t_0}, t) \left[ \lambda_{\pi_i,1-\pi_i} - \hat{\lambda}_{\pi_i,1-\pi_i} \right] \frac{1}{2} \left[ dN_t - \hat{\lambda}_{\pi_i,1-\pi_i} dt \right], \]

where \( L[\cdot] \) is the forward Kolmogorov–Fokker–Plank differential operator associated with \( x \).

Equation (8) is established as follows. Corresponding to an infinitesimal time increment \((t, t + \Delta t)\) there is an increment \( \Delta x_i = x_{t+\Delta t} - x_t \) in the observed path and an increment

\[ \Delta c_i(v | z_{t_0}, t) = c_{t+\Delta t}(v | z_{t_0}, t, x_{t+\Delta t}) - c_t(v | z_{t_0}, t) \]
in the conditional characteristic function of \(x\), where

\[ c_{t+\Delta t}(v \mid x_{t_0, t}, \Delta x_t) = E[\exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) \mid x_{t_0, t}, \Delta x_t] \]

\[ = \int_{R^{2n}} \exp(j\langle v, X \rangle) \exp(j\langle v, Y \rangle) d^{2n}P(X, Y \mid x_{t_0, t}, \Delta x_t), \]

where \(\Delta x_t = x_{t+\Delta t} - x_t\) is the increment in the information signal \(x\) and \(P\) is the joint conditional probability distribution function for \(x_t\) and \(\Delta x_t\) given \(x_{t_0, t}\) and \(\Delta x_t\).

We evaluate \(c_{t+\Delta t}\) for fixed \(z_t\) and \(\Delta z_t\); namely, for \((z_t = \zeta_i, \Delta z_t = 0)\) and \((z_t = \zeta_i, \Delta z_t = \zeta_i - \zeta_j\) with \(i \neq j\). For \(z_t = \zeta_i\) and \(\Delta z_t = 0\), we have

\[ c_{t+\Delta t}(v \mid x_{t_0, t}, x_t = \zeta_i, \Delta x_t = 0) \]

\[ = \int_{R^{2n}} \exp(j\langle v, X \rangle) \exp(j\langle v, Y \rangle) d^{2n}P(X, Y \mid x_{t_0, t}, x_t = \zeta_i, \Delta x_t = 0). \]

The finite difference

\[ \delta^{2n}P(X, Y \mid x_{t_0, t}, x_t = \zeta_i, \Delta x_t = 0) \]

\[ = \Pr(x_t \in (X - \delta X, X], \Delta x_t \in (Y - \delta Y, Y] \mid x_{t_0, t}, x_t = \zeta_i, \Delta x_t = 0) \]

can be evaluated as

\[ \delta^{2n}P(X, Y \mid x_{t_0, t}, x_t = \zeta_i, \Delta x_t = 0) \]

\[ = \Pr(\Delta z_t = 0 \mid x_t \in (X - \delta X, X], \Delta x_t \in (Y - \delta Y, Y], x_{t_0, t}, x_t = \zeta_i) \]

\[ \times \delta^{2n}P(X, Y \mid x_{t_0, t}, x_t = \zeta_i)/\Pr(\Delta z_t = 0 \mid x_{t_0, t}, x_t = \zeta_i), \]

where, as \(\delta X \to 0\) and \(\delta Y \to 0\),

\[ \Pr(\Delta z_t = 0 \mid x_t \in (X - \delta X, X], \Delta x_t \in (Y - \delta Y, Y], x_{t_0, t}, x_t = \zeta_i) \]

\[ = 1 - \sum_{\zeta_k \neq \zeta_i} \lambda_{\zeta_i, \zeta_k}(t, x_{t_0, t}, X) \Delta t + o(\Delta t) \]

and where

\[ \Pr(\Delta z_t = 0 \mid x_{t_0, t}, x_t = \zeta_i) \]

\[ = E[\Pr(\Delta z_t = 0 \mid x_{t_0, t}, x_t = \zeta_i, x_t) \mid x_{t_0, t}, x_t = \zeta_i] \]

\[ = 1 - \sum_{\zeta_k \neq \zeta_i} \hat{\lambda}_{\zeta_i, \zeta_k} \Delta t + o(\Delta t). \]
Here, \( \lambda_{t_i, t_k} = E[\lambda_{t_i, t_k}(t, z_{t_0, t}, x_t) \mid z_{t_0, t}, z_t = z_{t_i}] \). Putting these results together, we obtain

\[
c_{t+\Delta t}(v \mid z_{t_0, t}, z_t = z_{t_i}, \Delta z_t = 0) = E\left\{ \exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) \times \left[ 1 - \sum_{t_k \neq t_i} (\lambda_{t_i, t_k} - \lambda_{t_i, t_k}) \Delta t \right] \mid z_{t_0, t}, z_t = z_{t_i} \right\} + o(\Delta t).
\]

Similarly, by using the same procedure, we obtain, for \( i \neq j \),

\[
c_{t+\Delta t}(v \mid z_{t_0, t}, z_t = z_{t_i}, \Delta z_t = z_{t_j} - z_{t_i}) = E\{\exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) \lambda_{t_i, t_j} \mid z_{t_0, t}, z_t = z_{t_i}\} \lambda_{t_i, t_j}^{-1} \lambda_{t_i, t_j} + o(\Delta t).
\]

Thus, we have for \( z_t \) and \( \Delta z_t \) variable

\[
c_{t+\Delta t}(v \mid z_{t_0, t}, \Delta z_t) = E\left\{ \exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) \times \left[ 1 - \sum_{t_k \neq t_i} (\lambda_{z_{t_i}, \Delta z_{t_k}} - \lambda_{z_{t_i}, \Delta z_{t_k}}) \Delta t \right] \mid z_{t_0, t} \right\} + o(\Delta t).
\]

Because \( \delta_{0, \Delta N_t} \xrightarrow{a.s.} 1 - \Delta N_t \) and \( \delta_{1, \Delta N_t} \xrightarrow{a.s.} \Delta N_t \) as \( \Delta t \downarrow 0 \), we can write

\[
c_{t+\Delta t}(v \mid z_{t_0, t}, \Delta z_t) = E\{\exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) \mid z_{t_0, t}\} - E\left\{ \exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) \sum_{t_k \neq z_{t_i}} (\lambda_{z_{t_i}, \Delta z_{t_k}} - \lambda_{z_{t_i}, \Delta z_{t_k}}) \mid z_{t_0, t} \right\} \Delta t
\]

\[
+ E\{\exp(j\langle v, x_t \rangle) \exp(j\langle v, \Delta x_t \rangle) (\lambda_{z_{t_i}, \Delta z_{t+k}} - \lambda_{z_{t_i}, \Delta z_{t+k}}) \mid z_{t_0, t}\} \lambda_{t_i, \Delta z_{t+k}}^{-1} \lambda_{t_i, \Delta z_{t+i}} \Delta N_t + o(\Delta t).
\]

The desired result (8) is now obtained by subtracting from this equation the characteristic function for \( x_t \) given \( z_{t_0, t} \) and taking the limit as \( \Delta t \downarrow 0 \). The required interchange of limit and expectation to establish the first term on
the right in (8) is justified by assumption (a) and the bounded convergence theorem.

ACKNOWLEDGMENT

I would like to thank Dr. P. Frost and Professor T. Kailath for bringing the paper by Yashin (1970) to my attention.

Received: September 22, 1971


REFERENCES


Snyder, D. L. (1970a), Estimation of stochastic intensity functions of conditional Poisson processes, Monograph No. 128, Biomedical Computer Laboratory, Washington University School of Medicine, St. Louis, Mo. 63110.

