General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

TECHNICAL
RESEARCH
REPORT

Optimal Estimation for Discrete Time Jump Processes

Marco V. Vaca and Steven A. Tretter

March 25, 1977

APR 1977
RECEIVED
NASA STI FACILITY
INPUT BRANCH

DEPARTMENT OF
ELECTRICAL ENGINEERING
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742

(NASA-CR-152620) OPTIMAL ESTIMATION FOR DISCRETE TIME JUMP PROCESSES (Maryland Univ.) 30 p HC A03/MF A01 CSCL 12A
N77-20825
Unclas
G3/64 22805

Optimal Estimation for Discrete Time Jump Processes

Marco V. Vaca* and Steven A. Tretter**

ABSTRACT

In this paper we obtain optimum estimates of nonobservable random variables or random processes which influence the rate functions of a discrete time jump process (DTJP).

The approach we follow is based on the a posteriori probability of a nonobservable event expressed in terms of the a priori probability of that event and of the sample function probability of the DTJP. Thus, we obtain a general representation for optimum estimates and recursive equations for MMSE estimates.

In general, MMSE estimates are nonlinear functions of the observations. We examine the problem of estimating the rate of a DTJP when the rate is a random variable with a probability density function of the form $cx^k(1-x)^m$ and show that the MMSE estimates are linear in this case. This class of density functions is rather rich and explains why there are insignificant differences between optimum unconstrained and linear MMSE estimates in a variety of problems.

---

† This work was partially supported by NASA Grant NSG5048.

* Department of Electrical Engineering, University of Maryland, College Park, Maryland, 20742.

** Department of Electrical Engineering, University of Maryland, College Park, Maryland, 20742.

I. INTRODUCTION

Estimation and decision problems arising in communications and control have been studied in detail for continuous time observations. However, not much has been published for the case in which the observation process is a discrete time jump process (DTJP). We define a DTJP as a process having arbitrary jumps at times $t_1, t_2, \ldots$. A more precise definition is given below. Segall [1] obtained some optimum estimates for the special case where the jumps are restricted to be unity by using discrete time martingale techniques. In this paper we derive optimal estimates for more general cases.

In Section 2 we define discrete time jump processes precisely, present some representations, and derive the likelihood function for an observed realization. In Section 3 we derive the a posteriori probability measure for a nonobservable random process we wish to estimate given an observed realization of the DTJP. Recursive optimum estimation equations are derived in Section 4. The problem of optimum linear estimation is briefly discussed in Section 5. An interesting example in which the optimum estimates turn out to be linear is presented in Section 6.

2. DEFINITION, REPRESENTATIONS, AND LIKELIHOOD FUNCTION FOR DISCRETE TIME JUMP PROCESSES

We wish to describe an arbitrary discrete time jump process, taking values on a ℓ-dimensional Euclidean space R^{ℓ}, by means of discrete time counting processes. This approach has been used in the context of processes with independent increments and in general continuous time jump process.

Let T be the countable set

$$T = \{t_0, t_1, t_2, \ldots\}$$

where t_i is a real number, i.e. $t_i \epsilon R$, for $i = 0, 1, 2, \ldots$. Let Ω be the set of all possible piecewise constant right continuous functions defined on R, taking values on R^{ℓ}, and having jumps in T only. An element $\omega \epsilon \Omega$ will be called a sample function. Define the variables $Y_i \triangleq Y(t_i)$ and $y_i \triangleq y(t_i)$ as

$$Y_i(\omega) \triangleq \text{value of } \omega \text{ at time} \quad t = t_i \epsilon T \text{ for } \omega \epsilon \Omega,$$

$$Y_0(\omega) \triangleq 0;$$

$$y_i(\omega) \triangleq Y_i(\omega) - Y_{i-1}(\omega) = \text{jump size of } \omega \text{ at time } t = t_i \epsilon T$$

$$y_0(\omega) \triangleq 0$$

Let $\mathfrak{F}$ be the minimal sigma-algebra of subsets of Ω such that all functions $(Y_i(\omega),\ t_i \in T)$ are measurable. Denote by P any probability measure on $\mathfrak{F}$. The triple $(\Omega, \mathfrak{F}, P)$ will be called the discrete time jump process and will be denoted by Y. Since y_i is a $\mathfrak{F}$-measurable function for all $i \geq 0$, we define $\mathfrak{F}_k$ to be the sub-sigma algebra of $\mathfrak{F}$ generated by $(y_i(\omega),\ i \in \{0, 1, \cdots, k\})$. For any Borel set A of R^{ℓ}, with $0 \notin A$, define the random variables $N_k(A)$ and $n_k(A)$ as

$$N_k(\omega, A) \stackrel{\Delta}{=} \sum_{0<i\leq k} I(Y_i(\omega) - Y_{i-1}(\omega) \in A) \tag{1}$$

$$n_k(\omega, A) \stackrel{\Delta}{=} N_k(\omega, A) - N_{k-1}(\omega, A) = I(y_k(\omega) \in A) \tag{2}$$

where $I(\cdot \in A)$ is the indicator set function of the set A. In accordance with accepted usage, we shall drop the symbol ω and write Y_k, N_k, n_k, etc., for $Y_k(\omega)$, $N_k(\omega)$, $n_k(\omega)$, etc., respectively. Note that $N_k(A)$ represents the number of jumps of the process Y that fall in A during the time interval $\{t_0 < t_i \leq t_k\}$. Thus, $N_k(A)$ is a finite, nondecreasing, $\mathfrak{F}_k$-measurable function of k. Therefore, $(N_k(A),\ k=0,1,\ldots)$ is a submartingale for any Borel set $A \subset R^{\ell}$. The Doob decomposition for submartingales, [2, Chapter VII], implies that there exists a unique decomposition of $N_k(A)$ in terms of a $(\mathfrak{F}_k, P)$-martingale $Q_k(A)$ and a $\mathfrak{F}_{k-1}$ - measurable, increasing process $\Pi_k(A)$ with $\Pi_0(A) = 0$ such that

$$N_k(A) = Q_k(A) + \Pi_k(A)\ ;\ k=0,1,2,\ldots \tag{3}$$

From (1) - (3) we obtain, for $k = 0, 1, 2, \ldots$, and any Borel set A

$$n_k(A) = q_k(A) + \pi_k(A) \tag{4}$$

where

$$q_k(A) \stackrel{\Delta}{=} Q_k(A) - Q_{k-1}(A)$$

$$\pi_k(A) \stackrel{\Delta}{=} \Pi_k(A) - \Pi_{k-1}(A)$$

Note that $(q_k(A),\ k=0,1,\ldots)$ is a martingale difference sequence (MD).

Remark 1. The random variable $\pi_k(A)$ has a simple interpretation in terms of the conditional probability of a jump at time t_k. By taking the conditional expectation with respect to $\mathfrak{F}_{k-1}$ of both sides of (4) we obtain

$$\pi_k(A) = P(y_k \epsilon A | \mathfrak{F}_{k-1}) \tag{5}$$

The Doob decomposition (4) has been defined here i) to model the process Y, ii) to guaranty the existance of $P(y_k \epsilon A | \mathfrak{F}_{k-1})$, and iii) for obtaining estimates of nonobservable events (Section 5).

It is possible to represent the process $(y_k, t_k \epsilon T)$ by means of the process n_k defined in (2). The following lemma is a special case of a result given in Gikhman and Skorokhod [3, Chapter VI] and the proof will be omitted.

Lemma 1. Let $y_k(A) \stackrel{\Delta}{=} y_k I(y_k \epsilon A) = y_k n_k(A)$ for any Borel set $A \subset R^{\ell}$ with $0 \notin A$. Then

$$y_k(A) = \int_A x n_k(dx) = \int_A x q_k(dx) + \int_A x \pi_k(dx) \tag{6}$$

for $k = 1, 2, \ldots$.

Note that $y_k(A)$ is the jump size of Y_k provided that $Y_k - Y_{k-1} \epsilon A$. If $A = R^{\ell} - 0$, $y_k(A)$ becomes $y_k(R^{\ell} - 0) = y_k$, with

$$y_k = \int x n_k(dx) = \int x q_k(dx) + \int x \pi_k(dx) \tag{7}$$

where the integration is on the space R^{ℓ} with the vector 0 excluded. The integrals in (6) and (7) are defined in the sense of Gikhman and Skorokhod [3, Section 3, Chapter VII].

Remark 2 If the space of all possible jumps of Y is countable, say $\mathcal{U} \subset R^{\ell}$, with $\mathcal{U} = \{U_1, U_2, \ldots\}$, the above representation reduces to

$$n_k(U) \triangleq I(y_k = U) \quad \text{for } U \in \mathcal{U} \tag{8}$$

$$n_k(U) = q_k(U) + \lambda_k(U) \tag{9}$$

and
$$y_k = \sum_{i=1}^{\infty} U_i \; n_k(U_i) = \sum_{i=1}^{\infty} U_i \; q_k(U_i) + \sum_{i=1}^{\infty} U_i \; \lambda_k(U_i) \tag{10}$$

where
$$\lambda_k(U) = P\,(n_k(U) = 1 \,|\, \mathfrak{F}_{k-1}) \tag{11}$$

In the estimation problem we will study later on, we will assume, for simplicity, a countable jump space $\mathcal{U}$.

The likelihood function. The likelihood function is a quantity proportional to the probability of observing a particular realization of the jump process $(Y_i,\ t_0 \le t_i \le t_k)$ for $t_i, t_k \in T$ and plays a fundamental role in estimation and decision problems [4].

We wish to find the likelihood function for a discrete time discrete amplitude jump process. Denote by $p_k \triangleq p(t_k)$ the probability of having a particular realization of Y, i.e.

$$p_k = P(y_i = \xi_i,\ i = 0, 1, \cdots, k) \tag{12}$$

where $\xi_i \in \mathcal{U}$. Then

$$p_k = P\,(y_k = \xi_k \,|\, y_i = \xi_i,\ i = 0, 1, \cdots, k-1)\, p_{k-1}$$

$$= \prod_{i=1}^{k} P(y_i = \xi_i \,|\, y_i = \xi_i,\ i = 0, ., \ldots, i-1) P(y_0 = \xi_0) \tag{13}$$

where

$$P(y_k = \xi_k \,|\, y_i = \xi_i,\ i = 0, 1, \cdots, k-1) = P(y_k = \xi_k \,|\, \mathfrak{F}_{k-1})$$

$$= P(n_k(\xi_k) = 1 \,|\, \mathfrak{F}_{k-1}) = \lambda_k(\xi_k) \tag{14}$$

Let $t_{j_1}, t_{j_2}, \cdots,$ be the jump times of the random process Y with jump amplitudes $\xi_{j_1}, \xi_{j_2}, \cdots$. Then $n_{j_i}(\xi_{j_i}) = 1$, $i = 1, 2, \cdots$. The probability of no jump, $n_k = 0$ is given by

$$P(n_k = 0 | y_i = \xi_i,\ i = k, 2, \cdots, k-1) = 1 - \lambda_k$$

where $$\lambda_k = \sum_{i=1}^{k} \lambda_i(\xi_i)$$

Therefore, the likelihood function (13) becomes, with $p_0 \triangleq P(y_0 = \xi_0)$,

$$p_k = p_0 \prod_{i=1}^{k} (\lambda_i(\xi_i))^{n_i(\xi_i)} (1-\lambda_i)^{1-n_i} \tag{15}$$

which can also be written as

$$\begin{aligned} p_k &= \exp\left[\sum_{i=1}^{k} (\ln(\lambda_i(\xi_i)) n_i(\xi_i) + (\ln(1-\lambda_1))(1-n_i)\right] \\ &= \exp\left[\sum_{i=1}^{k} \left(\sum_{m=1}^{\infty} \ln \frac{\lambda_i(U_m)}{1-\lambda_i}\ n_i(U_m) + \ln(1-\lambda_i)\right)\right] \end{aligned} \tag{16}$$

where we have used the fact that $\xi_0 \triangleq 0$, $n_i = \sum_{m=1}^{\infty} n_i(U_m)$, and

$$\sum_{i=1}^{k} \sum_{m=1}^{\infty} \ln \lambda_i(U_m)\, n_i(U_m) = \sum_{i=1}^{k} \ln \lambda_i(\xi_i)\, n_i(\xi_i)$$

Remark 3 If the set of all jump amplitudes of Y is uncountable, then, by assuming that the limit

$$\lambda_i(x) \triangleq \lim_{\max|\Delta x_i| \to 0} \left(\prod_{i=1}^{\ell} (\Delta x_i)\right) \pi_i([x, x+\Delta x)]$$

exists, where $x = (x_1, x_2, \cdots, x\ ) \in R^{\ell}$, then the likelihood function is:

$$P_k = \exp\left[\sum_{i=1}^{k}\left(\int_{R^\ell} \ln\left(\frac{\lambda_i(x)}{1-\lambda_i}\right) n_i(dx) + \ln(1-\lambda_i)\right)\right] \tag{17}$$

where $\lambda_i \triangleq \int_{R^\ell} \lambda_i(x)m(dx)$ and m is a measure on R^ℓ.

3. A POSTERIORI PROBABILITIES FOR ESTIMATION

We shall formulate the estimation problem in a manner motivated by a problem in communication theory [5], [6]. Let $X(t)$ be a nonobservable "signal" which is applied to the input of a general channel. Let $y(t)$ be the output of the channel at time $t \in T$. We will assume that the observation record of the channel output at times $t_0, t_1, t_2, \cdots$ is a sample function of the discrete time jump process described in Section 2. Based on the observation record $(y(\sigma),\ 0 \le \sigma \le t;\ \sigma, t \in T)$ we wish to find an estimate of X, in particular, the minimum mean square error estimate. We formulate the problem in some convenient probability spaces in such a manner that the observations y can influence the "signals" X.

Let $(\Omega_s, \mathcal{B}_s, P_s)$ be a probability space called the "signal space" where the events $B_s \in \mathcal{B}_s$ are nonobservable. Let $(\Omega_m, \mathcal{B}_m, P_m(\omega_s, \cdot))$, $\omega_s \in \Omega_s$, be a probability space called the "transfer space" where the probability measure is parameterized by the elements ω_s. The transfer space models the channel behavior for each $\omega_s \in \Omega_s$. We want to obtain statistical inferences about the nonobservable events $B_s \in \mathcal{B}_s$ by observing events $B_m \in \mathcal{B}_m$.

We will assume that the elements $\omega_m \in \Omega_m$ are the sample functions of the discrete time jump process described in Section 2.

It is convenient to construct the product space $(\Omega, \mathcal{B}, P)$ where $\Omega = \Omega_s \times \Omega_m$, $\mathcal{B} = \mathcal{B}_s \otimes \mathcal{B}_m$ and

$$P(B_s \times B_m) \triangleq \int_{B_s} P_m(\omega_s, B_m) P_s(d\omega_s) \tag{18}$$

For example, let E_m be the event $\{y_1 = U_1, y_2 = U_2, \cdots y_k = U_k\} \in \mathcal{B}$. This event represents a particular realization of the discrete time jump process from $t = t_0$ $(y_o \triangleq 1)$ up to $t = t_k$. Then, from (16)

$$P_m(\omega_s, E_m) = P(\omega_s, Y_i, i \in [0, \cdots, k]) \triangleq p_k(\omega_s, \omega_m^k)$$

$$= \exp\left[\sum_{i=1}^{k}\left(\sum_{n=1}^{\infty} \ln\left(\frac{\lambda_i(U_n;\omega_s,\omega_m^{i-1})}{1-\lambda_i(\omega_s,\omega_m^{i-1})}\right) n_i(U_n;\omega_m^{i-1}) - \ln\frac{1}{1-\lambda_i(\omega_s,\omega_m^{i-1})}\right)\right] \tag{19}$$

where we have indicated, explicitly, that the rates λ_i and $\lambda_i(U_n)$ depend on the "signal" element ω_s and the sample path ω_m from t_0 up to t_{i-1}, i.e. ω_m^{i-1}.

Let us define a new probability measure P_m^0 on the transfer space, functionally independent of ω_s and mutually absolutely continuous with respect to $P_m(\omega_s, \cdot)$. For the event E_m, we define

$$P_m^0(E_m) = \exp\left[\sum_{i=1}^{k}\left(\sum_{n=1}^{\infty} \ln\left(\frac{\gamma_i(U_n;\omega_m^{i-1})}{1-\gamma_i(\omega_m^{i-1})}\right) n_i(U_n) - \ln\frac{1}{1-\gamma_i(\omega_m^{i-1})}\right)\right] \tag{20}$$

where the rates $\gamma_i(U_n)$ and γ_i are not functions of ω_s. We define the likelihood ratio $L_k(\omega_s, \omega_m)$ as

$$L_k(\omega_s, \omega_m^{i-1}) \triangleq \frac{P_m(\omega_s, E_m)}{P_m^0(E_m)}$$

$$= \exp\left[\sum_{i=1}^{k}\left(\sum_{n=1}^{\infty} \ln\left(\frac{(\lambda_i(U_n;\omega_s,\omega_m^{i-1})(1-\gamma_i(\omega_m^{i-1}))}{\gamma_i(U_n;\omega_m^{i-1})(1-\lambda_i(\omega_s,\omega_m^{i-1}))}\right) n_i(U_n)\right.\right.$$

$$\left.\left. - \ln\frac{1-\gamma_i(\omega_m^{i-1})}{1-\lambda_i(\omega_s,\omega_m^{i-1})}\right)\right] \tag{21}$$

The probability of any event $B_s \epsilon \beta_s$ given a sample path realization ω_m of the observation process can be calculated in terms of the likelihood ratio L_k given in (21). In fact, we have

Theorem 1. (Prior-to-posterior probability)

Let (Ω_s, β_s, P_s) and $(\Omega_m, \beta_m, P_m(\omega_s, \cdot))$ be the signal and transfer spaces defined priviously. Let L_k be a likelihood ratio between $P_m(\omega_s, \cdot)$ and P_m^0. Then

$$P(B_s \times \Omega_m | S_0 \otimes \mathcal{F}_k) = \frac{\int_{B_s} L_k(\omega_s, \omega_m) P_s(d\omega_s)}{\int_{\Omega_s} L_k(\omega_s, \omega_m) P_s(d\omega_s)} \triangleq P_s(B_s | \mathcal{F}_k) \tag{22}$$

for every $B_s \epsilon \beta_s$, where $S_0 \triangleq \{\phi_s, \Omega_s\}$.

Proof. The proof of this theorem, in a more general context, is given in [6, section IV]

Note that the right-hand side of (22) does not depend on P_m^0 because it can be written, alternatively, using (21), as

$$P(B_s \times \Omega_m | S_0 \otimes \mathcal{F}_m(t_k)) = \frac{\int_{B_s} P_m(\omega_s, B_m) P_s(d\omega_s)}{\int_{\Omega_s} P_m(\omega_s, B_m) P_s(d\omega_s)} \triangleq P_s(B_s | \mathcal{F}_k) \tag{23}$$

where $B_m = E_m$ and $\mathcal{F}_k \triangleq S_0 \otimes \mathcal{F}_m(t_k)$, therefore P_m^0 is a "ficticious" probability measure used to prove the above theorem.

Remark 4 (Conditional probability density) Let the nonobservable random variable X be defined as $X(\omega_s) = \omega_s \epsilon \Omega_s$ where $\Omega_s = R^n$. We want to obtain the conditional probability density function of X given $\mathcal{F}_k$. For that purpose, let $B_s = [X, X+\Delta X)$. Then (23) becomes

$$P_s\left([X, X+\Delta X) \mid \mathcal{F}_k\right) = \frac{\int_X^{X+\Delta X} P_m(x, B_m) P_s(dx)}{\int_{R^n} P_m(x, B_m) P_s(dx)} \tag{24}$$

Dividing both sides of (24) by $\prod_{i=1}^{n} (\Delta X_i)$, taking the limit when $\max_i |\Delta X_i| \to 0$, and assuming that $P_s(dx) = p_s(x)dx$, we obtain

$$\lim_{\max|\Delta X_i| \downarrow 0} \prod_{i=1}^{n} (\Delta X_i)^{-1} P_s\left([X, X+\Delta X) \mid \mathcal{F}_k\right) = \frac{P_m(X, B_m)}{\int_{R^n} P_m(x, B_m)\, p_s(x)dx}\, p_s(X)$$

$$\triangleq p_s(X \mid \mathcal{F}_k) \tag{25}$$

where $E_m = B_m$ and $P_m(X, B_m)$ is given by (19) with $\omega_s \triangleq X$.

4. RECURSIVE OPTIMUM ESTIMATES

Let $(X_k, k=0,1,2,\ldots)$ be an integrable random process defined on the product space $(\Omega, \mathcal{B})$. We want to obtain the best estimate $\hat{X}_k$ by observing a sample path ω_m^n from t_0 up to t_n. The criteria is the minimum mean square error.

It is well known that the conditional mean minimizes the mean square error. Therefore, the best estimate is

$$\hat{X}_k(\omega_m^n) \triangleq \int_{\Omega_s} X_k(\omega_s, \omega_m^n) P_s(d\omega_s | \mathcal{F}_n) \tag{26}$$

Using (23), we can write (26) as

$$\hat{X}_k(\omega_m^n) = \frac{E_s(p_n(\omega_s, \omega_m^n)\, X_k(\omega_s, \omega_m^n))}{E_s(p_n(\omega_s, \omega_m^n))} \tag{27}$$

where, for simplicity we have defined

$$p_n(\omega_s, \omega_m^n) \triangleq P_m(\omega_s, B_m)\ ;\ \ B_m = \{E_n, N_n = n\} \tag{28}$$

Equation (27) is the best estimate of X_k based on the observation path ω_m^n from t_0 up to t_n, and we have the following cases:

i) smoothing estimate, if $n > k$ (i.e. $t_n > t_k$)

ii) filtering estimate, if $n = k$ (i.e. $t_n = t_k$)

iii) prediction estimate, if $n < k$ (i.e. $t_n < t_k$)

For simplicity, we shall write (27) as

$$\hat{X}_{k|n} = \frac{E_s(p_n X_k)}{E_s(p_n)} \tag{27'}$$

Note that X_k may not be $\mathcal{B}_s \otimes \mathcal{F}_m(t_n)$ - measurable for $k > n$, which implies that X_k may not be $\mathcal{B}_s$ - measurable and (27) does not apply. However, if X_k is constant on Ω_m, then, it is $\mathcal{B}_s$ measurable and (27) applies.

Remark 5. Note that the random process X depends on both the signal ω_s and the observations ω_m which implies that feedback is allowed, and (27) is the best estimate of X.

Recursive filtering estimate. We wish to find a recursive formula for $\hat{X}$ given in (27) for n=k.

From (19) we see that

$$p_k = p_{k-1} \exp\left[\sum_{n=1}^{\infty} \ln\left(\frac{\lambda_k(U_n)}{1-\lambda_k}\right) n_k(U_n) - \ln\frac{1}{1-\lambda_k}\right] \tag{29}$$

where we have dropped ω_s and ω_m for simplicity. We prove now that the denominator of (27) satisfies

$$\bar{p}_k \triangleq E_s(p_k) = p_{k-1} \exp\left(\sum_{n=1}^{\infty}\left(\ln\frac{\lambda_{i|i-1}(U_n)}{1-\hat{\lambda}_{i|i-1}(U_n)} n_i(U_n) - \ln\frac{1}{1-\hat{\lambda}_{i|i-1}}\right)\right) \tag{30}$$

where $\hat{\lambda}_{i|i-1}(U_n) \triangleq \dfrac{E_s(\lambda_i(U_n)\cdot p_{i-1})}{E_s(p_{i-1})}$, $k=1,2,\cdots$.

For k=0 we have $p_o=1$ and

$$p_1 = (1-\lambda_1)\prod_{n=1}^{\infty}\left(\frac{\lambda_1(U_n)}{1-\lambda_1)}\right)^{n_1(U_n)} = \begin{cases} 1-\lambda_1 & \text{if } n_1 = 0 \\ \lambda_1(\xi) & \text{if } n_1(\xi) = 1 \end{cases} \tag{30'}$$

The last equality follows because $n_1=1$ if and only if there is a single jump of size $\xi \epsilon \mathcal{U}$ at $t=t_1$. Notice that (30') can be written as

$$p_1 = p_0\,(\lambda_1(\xi)\,n_1(\xi) + (1-\lambda_1)(1-n_1)) \tag{31}$$

Taking the expectation with respect to E_s, dividing both sides of (31) by $E_s(p_0)$ (which is equal to 1), and using (27), we have

$$\frac{E_s(p_1)}{E_s(p_0)} = \frac{E_s(p_0\lambda_1(\xi))}{E_s(p_0)} n_1(\xi) + \frac{E_s(p_0(1-\lambda_1))}{E_s(p_0)} (1-n_1)$$

Then $$E_s(p_1) = E_s(p_0)\left[\hat{\lambda}_{1|0}(\xi)\, n_1(\xi) + (1-\hat{\lambda}_{1|0}(\xi))\,(1-n_1)\right] \tag{32}$$

or $$\bar{p}_1 = p_0\,(\hat{\lambda}_{1|0}(\xi) n_1(\xi) + (1-\hat{\lambda}_{1|0}(\xi)\,(1-n_1))$$

Since (31) and (32) satisfy the same type of difference equation we conclude that $\bar{p}_1$ satisfies (30) with k=1. Using mathematical induction, it is easy to verify (30). The filtering estimate $X_{k|k}$ becomes

$$\hat{X}_{k|k} = \frac{E_s(p_k X_k)}{E_s(p_k)} = E_s(\Lambda_k X_k) \tag{33}$$

where

$$\Lambda_k = (\bar{p}_k)^{-1} p_k = \exp\left[\sum_{i=1}^{k}\left(\sum_{n=1}^{\infty} \ln \frac{\lambda_i(U_k)(1-\hat{\lambda}_{i|i-1})}{\hat{\lambda}_{i|i-1}(U_n)\,(1-\lambda_i)} n_i(U_n) - \ln \frac{\hat{\lambda}_{i|i-1}}{1-\lambda_i}\right)\right]$$

$$= \Lambda_{k-1} \exp\left[\sum_{n=1}^{\infty} \ln \frac{\lambda_k(U_n)\,(1-\hat{\lambda}_{k|k-1})}{\hat{\lambda}_{k|k-1}(U_n)\,(1-\lambda_k)} n_k(U_n) - \ln \frac{1-\hat{\lambda}_{k|k-1}}{1-\lambda_k}\right]$$

$$\triangleq \Lambda_{k-1}\, \ell_k \;, \quad k=1, 2, \cdots \tag{34}$$

where ℓ_k is the exponential formula, and $\Lambda_0 \equiv 1$.

Theorem 2 (Optimal filtering estimate). The optimum filtering estimate $\hat{X}_{k|k}$ given in (33) satisfies the stochastic difference equation

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + \sum_{i=1}^{\infty} \frac{\hat{E}_{k-1}(X_k \lambda_k(U_i)) - \hat{X}_{k|k-1}\,\hat{\lambda}_{k|k-1}(U_i) + \hat{h}_{k|k-1}(U_i)}{\hat{\lambda}_{k|k-1}(U_i)\,(1-\hat{\lambda}_{k|k-1})}\, \hat{q}_k(U_i) \tag{35}$$

for $k=1, 2, \cdots$, where

$$\hat{E}_{k-1}(X_k\lambda_k(U_i)) \triangleq \frac{E_s(X_k\lambda_k(U_i)p_{k-1})}{E_s(p_{k-1})} \tag{36}$$

$$\hat{h}_{k|k-1}(U_i) \triangleq \sum_{j=1}^{\infty} \hat{E}_{k-1}\left[X_k(\hat{\lambda}_{k|k-1}(U_i)\lambda_k(U_j) - \hat{\lambda}_{k|k-1}(U_j)\lambda_k(U_i)\right]$$

$$= \hat{E}_{k-1}\left[X_k(\hat{\lambda}_{k|k-1}(U_i)\lambda_k - \hat{\lambda}_{k|k-1}\lambda_k(U_i))\right] \tag{37}$$

and

$$\hat{q}_k(U_i) \triangleq n_k(U_i) - \hat{\lambda}_{k|k-1}(U_i) \tag{38}$$

Proof. From (33) and (34) we have

$$\hat{X}_{k|k} = E_s(X_k\Lambda_k) = E_s(X_k\Lambda_{k-1}\ell_k) \tag{39}$$

But
$$\ell_k = \begin{cases} \dfrac{1-\lambda_k}{1-\hat{\lambda}_{k|k-1}} & , n_k = 0 \\[2ex] \dfrac{\lambda_k(U_i)}{\hat{\lambda}_{k|k-1}(U_i)} & , n_k(U_i) = 1 , \quad U_i \in \mathcal{U},\ i=1,2,\cdots \end{cases}$$

$$= \frac{1-\lambda_k}{1-\hat{\lambda}_{k|k-1}}\left(1 - \sum_{i=1}^{\infty} n_k(U_i)\right) + \sum_{i=1}^{\infty} \frac{\lambda_k(U_i)}{\hat{\lambda}_{k|k-1}(U_i)} n_k(U_i)$$

$$= \frac{1-\lambda_k}{1-\hat{\lambda}_{k|k-1}} + \sum_{i=1}^{\infty} \frac{\lambda_k(U_i) - \hat{\lambda}_{k|k-1}(U_i)}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} n_k(U_i)$$

$$+ \sum_{i=1}^{\infty} \frac{\hat{\lambda}_{k|k-1}(U_i)\lambda_k - \lambda_k(U_i)\hat{\lambda}_{k|k-1}}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} n_k(U_i) \tag{40}$$

Thus, using (39) we have:

$$\hat{X}_{k|k} = \frac{\hat{X}_{k|k-1} - \hat{E}_{k-1}(X_k\lambda_k)}{1-\hat{\lambda}_{k|k-1}} + \sum_{i=1}^{\infty} \frac{\hat{E}_{k-1}(X_k\lambda_k(U_i)) - \hat{X}_{k|k-1}\hat{\lambda}_{k|k-1}(U_i)}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} n_k(U_i)$$

$$+ \sum_{i=1}^{\infty} \frac{\hat{E}_{k-1}\left[X_k(\hat{\lambda}_{k|k-1}(U_i)\lambda_k - \hat{\lambda}_{k|k-1}\lambda_k(U_i))\right]}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} n_k(U_i) \tag{41}$$

By noting that $\lambda_k = \Sigma_i \lambda_k(U_i)$ and $\hat{\lambda}_{k|k-1} = \Sigma_i \hat{\lambda}_{k|k-1}(U_i)$, the first two terms on the right hand side of (41) are equal to

$$\hat{X}_{k|k-1} + \sum_{i=1}^{\infty} \frac{\hat{E}_{k-1}(X_k \lambda_k(U_i)) - \hat{X}_{k|k-1}\hat{\lambda}_{k|k-1}(U_i)}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} (n_k(U_i) - \hat{\lambda}_{k|k-1}(U_i)) \qquad (42)$$

and, after some manipulation, the third term on the right hand side of (41) is equal to

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\hat{E}_{k-1}\left[X_k (\hat{\lambda}_{k|k-1}(U_i)\lambda_k(U_j) - \hat{\lambda}_{k|k-1}(U_j)\lambda_k(U_i))\right]}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} \hat{q}_k(U_i)$$

$$= \sum_{i=1}^{\infty} \frac{\hat{h}_{k|k-1}(U_i)}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} \hat{q}_k(U_i) \qquad (43)$$

where $\hat{h}_{k|k-1}(U_i)$ is defined in (37). Combining (42) and (43) and using (38) we obtain (35).

Example 1 - If the observation process is a discrete counting process, i.e., $\mathcal{U} = \{U_1\} = \{1\}$, then $\lambda_k(U_1) = \lambda_k$ and (32) reduces to

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + \frac{\hat{E}_{k-1}(X_k \lambda_k) - \hat{X}_{k|k-1}\hat{\lambda}_{k|k-1}}{\hat{\lambda}_{k|k-1}(1-\hat{\lambda}_{k|k-1})} \hat{q}_k$$

This equation has been obtained by Segall [1].

Example 2 - Let us assume that the nonobservable random process X_k can be represented as

$$X_k = f(k, X_{k-1}, u_{k-1}) + w_k$$

where f is a known function of X_{k-1} and of a $\mathcal{F}_{k-1}$ measurable

control u_k; w_k is a MD on the signal space, and is not a function of ω_m then w_k can be interpreted as noise in the dynamics of X_k. Then, the one step prediction $\hat{X}_{k|k-1}$ is

$$\hat{X}_{k|k-1} = E_s(X_k \Lambda_{k-1}) = E_s(f(k, X_{k-1}, U_{k-1})\Lambda_{k-1}) + E_s(w_k \Lambda_{k-1})$$

$$= \hat{f}_{k|k-1} + E_s(w_k \Lambda_{k-1})$$

We will assume that w_k is a MD with respect to some sigma-algebra $\mathcal{B}_s(t_k) \subset \mathcal{B}_s$, and that Λ_{k-1} is $\mathcal{B}_s(t_{k-1})$ measurable, then

$$E_s(w_k \Lambda_{k-1}) = E_s(E_s(w_k \Lambda_{k-1} | \mathcal{B}_s(t_{k-1})) = 0$$

Thus $\quad \hat{X}_{k|k-1} = \hat{f}_{k|k-1}$

Example 3 - Let us assume that the rate parameter λ_k is a fixed random variable X defined on the signal space, i.e. $\lambda_k = X = X(\omega_s)$ and that X is uniformly distributed on $[0,1]$. The best estimate $\hat{X}_k$ at time $t=t_k$ is given by:

$$\hat{X}_k = \hat{X}_{k-1} + \frac{\hat{E}_{k-1}(X^2) - (\hat{X}_{k-1})^2}{\hat{X}_{k-1} - (\hat{X}_{k-1})^2} (n_k - \hat{X}_{k-1}) \quad : k=1, 2, \cdots \tag{44}$$

where $\hat{X}_0 = \frac{1}{2}$.

Note that in order to solve (44), we need to know $\hat{E}_{k-1}(X^2)$, which can be obtained from another difference equation involving a term $\hat{E}_{k-1}(X^3)$ and so on up to infinity. Therefore (44) is not a closed form solution for the best estimate. However, this is a general characteristic of nonlinear estimation.

Motivated by the problems of solving (44), we develop below a recursive formula for the conditional probability density $\hat{p}_k(X)$.

Theorem 3 - Let $\hat{p}_k(x)$ be the conditional probability density of the random variable $X = X(\omega_s)$ given $\mathcal{F}_k$. Let $\lambda_k(U_i, X)$, the rate of the jump process, for $U_i \in \mathcal{U}$, be a know function of X. Then

$$\hat{p}_k(x) = \hat{p}_{k-1}(x)\left[1 + \sum_{i=1}^{\infty} \frac{\lambda_k(U_i, x) - \hat{\lambda}_{k-1}(U_i) + \hat{g}_{k-1}(U_i, x)}{\hat{\lambda}_{k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} \hat{q}_k(U_i)\right] \tag{45}$$

where $\hat{g}_{k-1}(U_i, x) \triangleq \hat{\lambda}_{k|k-1}(U_i, x)\lambda_k(x) - \hat{\lambda}_{k|k-1}\lambda_k(U_i, x)$

Proof Let us consider the random variable $y = e^{jvx}$, then

$$\hat{E}_k(y) = \hat{E}_{k-1}(y) + \sum_{i=1}^{\infty} \frac{\hat{E}_{k-1}(y\lambda_k(U_i)) - \hat{y}_{k-1}\hat{\lambda}_{k|k-1}(U_i) + \hat{h}_{k|k-1}(U_i)}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} \hat{q}_k(U_i) \tag{46}$$

Since $p_k(\cdot)$ is the conditional probability density, then

$$E_k(y) = \int \exp(jvx)\hat{p}_k(x)dx \quad , \quad \text{for } v \in R \tag{47}$$

Therefore, from (46) and (47) we get

$$\int \exp(jvx)\hat{p}_k(x)dx = \int \exp(jvx)\hat{p}_{k-1}(x)dx$$

$$+ \sum_{i=1}^{\infty} \frac{1}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})}\left[\int \exp(jvx)\lambda_k(U_i)\hat{p}_{k-1}(x)dx - \hat{\lambda}_{k|k-1}(U_i)\int \exp(jvx)\hat{p}_{k-1}(x)dx\right.$$

$$\left. + \int \exp(jvx)(\hat{\lambda}_{k|k-1}(U_i)\lambda_k - \lambda_{k|k-1}\hat{\lambda}_k(U_i))\hat{p}_{k-1}(x)dx\right]$$

for any $v \in R$. Thus

$$\hat{p}_k(x) = \hat{p}_{k-1}(x) + \hat{p}_{k-1}(x) \sum_{i=1}^{\infty} \frac{\lambda_k(U_i) - \hat{\lambda}_{k|k-1}(U_i) + \hat{\lambda}_{k|k-1}(U_i)\lambda_k - \hat{\lambda}_{k|k-1}\lambda_k(U_i)}{\hat{\lambda}_{k|k-1}(U_i)(1-\hat{\lambda}_{k|k-1})} \hat{q}_k(U_i)$$

where $\hat{p}_0(x)$ is the initial probability density function for the random variable X.

Example 3 (cont.) Let $\lambda_k(U_1, X) = \lambda_k(X) = X$ be a uniformly distributed random variable, then (45) reduces to

$$\hat{p}_k(x) = \hat{p}_{k-1}(x)\left[1 + \frac{x - \hat{X}_{k-1}}{\hat{X}_{k-1} - (\hat{X}_{k-1})^2}(n_k - \hat{X}_{k-1})\right], \quad k = 1, 2, \cdots \tag{48}$$

where $\hat{X}_{k-1} = \int x\,\hat{p}_{k-1}(x)dx$, and

$$\hat{p}_0(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & , \text{ otherwise} \end{cases}$$

Notice that knowing $\hat{p}_{k-1}(x)$ we can find $\hat{X}_{k-1}$ which in turn allows us to find $\hat{p}_k(x)$, and so on. Thus we obtain a close form solution for all the conditional moment of X. It is straightforward to verify from (48) that $\hat{X}_k$ satisfies (44)

Recursive Smoothing Estimate - We wish to find a recursive formula for the optimum smoothing estimate $X_{k|n}$ of the random variable X_k given $\mathcal{F}_n$ for $k < n$.

Theorem 4 (Optimal smoothing estimate). The optimum estimate $\hat{X}_{k|n}$, for $k < n$ and k fixed, satisfies the stochastic equation

$$\hat{X}_{k|n} = \hat{X}_{k|k} + \sum_{i=k}^{n-1} \sum_{m=1}^{\infty} \frac{\hat{E}_i(X_k \hat{\lambda}_{i+1}(U_m)) - \hat{X}_{k|i}\hat{\lambda}_{i+1|i}(U_m) + \hat{h}_{k|i}(U_m)}{\hat{\lambda}_{i+1|i}(U_m)\,(1 - \hat{\lambda}_{i+1|i})}\,\hat{q}_{i+1}(U_m) \tag{49}$$

where

$$\begin{aligned} \hat{h}_{k|i}(U_m) &\triangleq \sum_{j=1}^{\infty} \hat{E}_i\left[X_k(\hat{\lambda}_{i+1|i}(U_m)\lambda_{i+1}(U_j) - \hat{\lambda}_{i+1|i}(U_j)\lambda_{i+1}(U_m)\right] \\ &= E_i\left[X_k(\hat{\lambda}_{i+1|i}(U_m)\lambda_{i+1} - \hat{\lambda}_{i+1|i}\lambda_{i+1}(U_m))\right] \end{aligned} \tag{50}$$

and $\hat{q}_{i+1}(U_m) \triangleq n_{i+1}(U_m) - \hat{\lambda}_{i+1|i}(U_m)$

Proof The proof is very similar to that of Theorem 2. In fact, the smoothing estimate is given by

$$\hat{X}_{k|n} = E_s(X_k \Lambda_n) = E_s(X_k \Lambda_{n-1} \ell_n) \tag{51}$$

where

$$\ell_n = \frac{1-\lambda_n}{1-\hat{\lambda}_{n|n-1}} + \sum_{m=1}^{\infty} \frac{\lambda_n(U_m) - \hat{\lambda}_{n|n-1}(U_m)}{\hat{\lambda}_{n|n-1}(U_m)(1-\hat{\lambda}_{n|n-1})} n_n(U_m)$$

$$+ \sum_{m=1}^{\infty} \frac{\hat{\lambda}_{n|n-1}(U_m)\lambda_m - \lambda_n(U_m)\hat{\lambda}_{n|n-1}}{\hat{\lambda}_{n|n-1}(U_m)(1-\hat{\lambda}_{n|n-1})} n_n(U_m) \tag{52}$$

Upon supstitution of (52) in (51), we obtain

$$\hat{X}_{k|n} = \hat{X}_{k|n-1} + \sum_{m=1}^{\infty} \frac{\hat{E}_{n-1}(X_k \lambda_n(U_m)) - \hat{X}_{k|n-1}\hat{\lambda}_{n|n-1}(U_m) + \hat{h}_{k|n-1}(U_m)}{\hat{\lambda}_{n|n-1}(U_m)(1-\hat{\lambda}_{n|n-1})} \hat{q}_n(U_m) \tag{53}$$

for $n = k+1, k+2, \cdots$. Writing the stochastic equations for $\hat{X}_{k|n-1}, \hat{X}_{k|n-2}, \cdots$, etc., we deduce (49).

Recursive Prediction Estimate - We wish to find a recursive equation for optimum prediction estimate $X_{k|n}$ of the random variable X_k given $\mathfrak{F}_n$ for $n < k$. We assume here that X_k is $\mathcal{B}_s$ measurable. A sufficient condition for the $\mathcal{B}_s$ measurability of X_k is that X_k be constant on Ω_m.

Theorem 5 (Optimum prediction estimate). Assume that the random variable X_k is $\mathcal{B}_s$ measurable. The optimum prediction estimate $\hat{X}_{k|n}$, for $k > n$ and k fixed, satisfies the stochastic equation

$$\hat{X}_{k|n} = \hat{X}_{k|0} + \sum_{i=0}^{n-1} \sum_{m=1}^{\infty} \frac{\hat{E}_i(X_k \lambda_{i+1}(U_m)) - \hat{X}_{k|i} \hat{\lambda}_{i+1|i}(U_m) + \hat{h}_{k|i}(U_m)}{\hat{\lambda}_{i+1|i}(U_m)(1-\lambda_{i+1|i})} \hat{q}_{i+1}(U_m) \tag{54}$$

where $\hat{h}_{k|i}(U_m)$ is defined in (50)

Proof - The proof of this theorem is identical to that of Theorem 4 and will be omitted.

A special case of Theorem 4 is for $U = \{U_1\} = \{1\}$. In this case, the recursive formula for $\hat{X}_{k|n}$ becomes

$$\hat{X}_{k|n} = \hat{X}_{k|k} + \sum_{i=k}^{n-1} \frac{\hat{E}_i(X_k \lambda_{i+1}) - \hat{X}_{k|i} \hat{\lambda}_{i+1|i}}{\hat{\lambda}_{i+1|i}(1-\hat{\lambda}_{i+1|i})} \tag{55}$$

which has been derived by Segall [1]. The prediction estimate $\hat{X}_{k|n}$, for $U = \{1\}$, becomes

$$\hat{X}_{k|n} = \hat{X}_{k|0} + \sum_{i=0}^{n-1} \frac{\hat{E}_i(X_k \lambda_{i+1}) - \hat{X}_{k|i} \hat{\lambda}_{i+1|i}}{\hat{\lambda}_{i+1|i}(1-\hat{\lambda}_{i+1|i})} \tag{56}$$

5. COMMENTS ON OPTIMUM LINEAR ESTIMATION

In this section we indicate how to obtain the best linear estimate $\hat{\lambda}^{\ell}_{k|n}(U_m)$ of the intensity function $\lambda_k(U_m)$, $m=1,2,\ldots$, in the sense of minimizing the error covariance function $E(\lambda_k(U_m) - \hat{\lambda}^{\ell}_{k|n}(U_m))^2$ by observing a sample path realization $(n_i(U_m),\ i=1,2,\ldots,n)$, $n \geq k$, of a discrete time point process $(n_i(U_m))$ which is obtained (see Section 2) from an arbitrary discrete time, discrete amplitude jump process $(y_i,\ i=1,2,\ldots,n)$. As we discuss in Section 2, the Doob submartingale decomposition of $n_i(U_m)$ gives

$$n_i(U_m) = \lambda_i(U_m) + q_i(U_m) , \tag{57}$$

for $i=1,2,\ldots,n$; $m=1,2,\ldots$, where $q_i(U_m)$ in a MD sequence, therefore $E(q_i(U_m)q_j(U_m)) = 0$ for all $i,j = 1,2,\ldots,n$.

The best linear estimate $\hat{\lambda}^{\ell}_{k|n}(U_m)$ is of the form

$$\hat{\lambda}^{\ell}_{k|n}(U_m) = E(\lambda_k(U_m)) + \sum_{i=1}^{n} H_{ki}(U_m)(n_i(U_m) - E(\lambda_i(U_m))) \tag{58}$$

where the unit response $H_{ki}(U_m)$ is obtained from the orthogonality principle

$$E\left[(\lambda_k(U_m) - \hat{\lambda}^{\ell}_{k|n}(U_m))(n_j(U_m) - E(\lambda_j(U_m))\right] = 0 \tag{59}$$

for $k,j \leq n$, and $m = 1,2,\ldots$.

When $\lambda_k(U_m)$ is the state of a linear dynamical system, the Kalman filter can be used to recursively compute the optimum linear estimates.

6. A CLASS OF PROBLEMS IN WHICH THE OPTIMUM ESTIMATES ARE LINEAR

It is well known that the unconstrained minimum mean-square error estimates of one set of random variables from another set are linear when the two sets are jointly normal. Few other examples are known where the optimum estimates are linear. In this section we present a problem for discrete time point processes in which the optimum estimates are linear.

We will examine the problem of estimating the rate parameter X for a binary discrete time point process when X is a random variable with the probability density function

$$p_X(x) = \begin{cases} \dfrac{(k+m+1)!}{m!\,k!}\, x^k (1-x)^m & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{60}$$

where k and m are non-negative integers. Let us assume that the observed discrete time point process y_n, n=1, 2, . . . is a sequence of binary numbers with

$$P(y_n = 1 \,|X) = 1 - P(y_n = 0 \,|X) = X \tag{61}$$

and that it is an independent sequence conditioned on X, that is,

$$P(y_i = \xi_i, i = 1, \ldots, n \,|X) = \prod_{i=1}^{n} P(y_i = \xi_i \,|X)$$

$$= X^S (1-X)^{n-S} \tag{62}$$

where

$$S = \sum_{i=1}^{n} y_i$$

From (25) it follows that

$$p_X(x \mid y_i = \xi_i, i = 1, \ldots, n) = p_X(x) \quad \frac{P(y_i = \xi_i, i = 1, \ldots, n \mid X = x)}{\int_0^1 P(y_i = \xi_i, i = 1, \ldots, n \mid X = x) p_X(x) dx}$$

$$= \frac{x^{k+S} (1-x)^{m+n-S}}{\int_0^1 x^{k+S} (1-x)^{m+n-S} dx} \qquad \text{for} \quad 0 \le x \le 1 \tag{63}$$

Using the fact that

$$\int_0^1 x^m (1-x)^n dx = m!\ n!/(m+n+1)!$$

yields

$$p_X(x \mid y_i, i=1, \ldots, n) = \frac{(k+m+n+1)!}{(k+S)!\ (m+n-S)!} \quad x^{k+S}(1-x)^{m+n-S} \text{ for } 0 \le x \le 1 \tag{64}$$

Therefore, the minimum mean-square error estimate of X given $y_i, \ldots, y_n$ is

$$\hat{X}_n = E\{X \mid y_i, \ldots, y_n\} = \int_0^1 x\, p_X(x \mid y_i, i=1, \ldots, n)\, dx$$

$$= (k+1+\sum_{i=1}^{n} y_i)/(n+k+m+2) \tag{65}$$

which is a linear estimate. This result is not at all obvious from the recursive estimation formula of Example 1 in Section 4. Notice that as n becomes large, the optimum estimate converges to the proportion of one's in the observed sequence.

The optimum estimate is unbiased. This follows since

$$E\{\hat{X}_n\} = (k+1+n\,E\{y_i\})/(n+k+m+2) \tag{66}$$

and

$$E\{y_i\} = E\{E\{y_i|X\}\} = E\{X\} = (k+1)/(k+m+2) \tag{67}$$

so that

$$E\{\hat{X}_n\} = E\{X\} \tag{68}$$

The linear minimum mean-square error estimate can also be derived by appealing to the Doob decomposition and expressing the observations as $y_i = X + q_i$. The sequence $q_i = y_i - X$ is a martingale difference sequence with

$$E\{q_i q_j\} = E\{E\{(y_i - X)^2 \,|X\}\}\,\delta_{ij} = E\{X(1-X)\}\,\delta_{ij}$$

$$= \frac{(m+1)(k+1)}{(m+k+2)(m+k+3)}\,\delta_{ij} \tag{69}$$

The observations can be arranged in the matrix form

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} X + \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \tag{70}$$

or

$$\underline{Y} = \underline{A}\;X + \underline{Q} \tag{70'}$$

Then the optimum linear estimate is [7, Ch. 13]

$$\hat{X}_n^{\ell} = E\{X\} + (\underline{A}^t \underline{R}^{-1} \underline{A} + V^{-1})^{-1} \underline{A}^t \underline{R}^{-1} (\underline{Y} - E\{\underline{Y}\}) \tag{71}$$

where $\underline{R} = \text{cov}\, \underline{Q}$ and $V = \text{var}\, X$. This reduces to the conditional mean $\hat{X}_n$ derived above. The corresponding mean-square error is

$$E\{(X - \hat{X}_n)^2\} = (\underline{A}^t \underline{R}^{-1} \underline{A} + V^{-1})^{-1} = \frac{(m+1)(k+1)}{(k+m+2)(k+m+3)(n+k+m+2)} \tag{72}$$

The Kalman filter [8] can be used to obtain the optimum linear estimates recursively. If we consider X to be the state of the dynamical system $X_{n+1} = X_n$ with $X_0 = X$, then the observations are $y_n = X_n + q_n$ and the Kalman filter equations become

$$\hat{X}_n^{\ell} = \hat{X}_{n-1}^{\ell} + (\Gamma_n / \text{var}\, q_n)(y_n - \hat{X}_{n-1}^{\ell}) \tag{73}$$

and

$$\Gamma_n = \Gamma_{n-1} (\text{var}\, q_n) / (\Gamma_{n-1} + \text{var}\, q_n) \tag{74}$$

with the initial conditions

$$\hat{X}_0^{\ell} = E\{X\} \quad \text{and} \quad \Gamma_0 = \text{var}\, X \tag{75}$$

The mean-square error at time n is Γ_n.

In computer simulations of the optimum nonlinear recursive estimator given by (44) and (48) and the optimum linear recursive estimator of previous paragraph with probability density functions for X not belonging to the class in this section, we found only small differences in the estimates. This can be explained by the

fact that after a few steps the a posteriori probability density function becomes peaked about the true value of X and can be closely approximated by a density of the class in this section. Then the two estimates become nearly identical.

REFERENCES

1. A. Segall, "Recursive Estimation for Discrete-Time Point Processes", IEEE Trans. on Information Theory, Vol. IT-22, No. 44, pp. 422-431, July 1976.

2. P. A. Meyer, Probability and Potentials, Blaisdell Publishing Company, Waltham, 1966.

3. I. I. Gikhman and A. Y. Skorokhod, Introduction to the Theory of Random Processes, W. B. Saunders, Philadelphia, 1965.

4. D. L. Snyder, Random Point Processes, New York: Wiley, 1975.

5. T. Kadota, "A New Look at the Mathematical Foundations of Communication Theory", Proc. of 9th Annual Allerton Conf. on Circuits and Systems Theory, pp. 485-496, University of Illinois, Urbana, October 1971.

6. M. V. Vaca and D. L. Snyder, "Estimation and Decision for Observations Derived from Martingales: Part 1, Representations", IEEE Trans. on Information Theory, Vol. IT-22, No. 6, pp. 691-707, November 1976.

7. S. A. Tretter, Introduction to Discrete-Time Signal Processing, John Wiley & Sons, 1976.

8. R. E. Kalman, "A New Approach to Linear Filtering and Prediction," Transactions of the ASME, Journal of Basic Engineering, Vol. 82D, pp. 34-45, March 1960.

ACKNOWLEDGEMENTS

This research was partially supported by NASA Grant NSG5048.