INFORMATION AND CONTROL **13**, 534–540 (1968)

A Note on Least-Squares Estimates from Likelihood Ratios*

THOMAS KAILATH

Department of Electrical Engineering, Stanford University, Stanford, California 94305

In another paper we have shown that the likelihood ratio for detecting the presence of a finite-variance signal in additive white *Gaussian* noise can be expressed in terms of the causal least-squares estimate of the signal. In this paper a converse relation is derived. Comparisons are made with some related but nontrivially different results in the discrete-time case that were recently discussed in this journal by Esposito.

1. INTRODUCTION

This brief note complements a recent article [1] by Esposito that obtains a relation between likelihood ratios and least-squares estimates for discrete-time processes. The formula in [1], which was motivated by a result [2] of the present author, is here shown to have a nontrivially different analog for continuous-time processes. A detailed comparison is made in Section 4. The result in [2] is that the likelihood ratio for the detection of a random signal in additive white Gaussian noise can be expressed in terms of the *causal* least-squares estimate of the signal. Here we show that conversely the causal least-squares estimate can be obtained from the likelihood ratio. The case of a narrowband random-phase signal is worked out as an example.

We first describe the major result of [2]. Consider the two hypotheses

$$\left.\begin{aligned} H_1 &: \dot{x}(t) = z(t) + \dot{w}(t) \\ H_0 &: \dot{x}(t) = \dot{w}(t) \end{aligned}\right\}, \qquad 0 \leqq t \leqq T \tag{1}$$

where $\dot{w}(\cdot)$ is zero-mean white Gaussian noise with covariance function

$$\overline{\dot{w}(t)\dot{w}(s)} = \delta(t - s) \tag{2}$$

* This work was supported by the Applied Mathematics Division of the Air Force Office of Scientific Research under contract AF 49(638)1517 and by the the JSEP at Stanford University, contract Nonr 225(83).

[In other words, $w(\cdot)$ is the so-called *Wiener process*.] The signal process $z(\cdot)$ is a random process, not necessarily Gaussian, that has integrable variance over [0, T] and that is independent of $\dot{w}(\cdot)$. Then the likelihood ratio can be written [2]

$$L(T) = \exp\left[\oint_0^T \hat{z}(t \mid t)\, dx(t) - \frac{1}{2} \int_0^T \hat{z}^2(t \mid t)\, dt \right] \tag{3}$$

where

$$\hat{z}(t \mid t) = \text{the least-squares estimate of } z(t) \text{ given } x(\tau), 0 \leqq \tau \leqq t, \text{ and assuming } H_1 \text{ is true.} \tag{4}$$

and $\oint$ denotes a special kind of stochastic integral known as the Itô integral [3]. The Itô integral has some special properties that derive basically from the so-called Lévy property that the increments $dw(t)$ of a Wiener process are of the order of $\sqrt{dt}$, and not $O(dt)$ as they would be for a smoother random process. This means that second-order terms $(dw)^2$ cannot be neglected in the Itô stochastic calculus. This will be brought out strikingly by the Itô differential rule, which we shall briefly describe here since it will be needed below. In fact, the major result of this note is that, if the likelihood ratio $L(T)$ is known, then $\hat{z}(t \mid t)$ can be obtained from it by the formula

$$\hat{z}(t \mid t) = \frac{dL(t)}{L(t)\, dx(t)} \tag{5}$$

where $d(\cdot)$ denotes the Itô differential.

2. THE ITO DIFFERENTIAL RULE

We say $\{x(t), 0 \leqq t \leqq T\}$ is an *Itô process* if it can be written in the form

$$x(t) = \int_0^t a(s)\, ds + \oint_0^t b(s)\, dw(s), \qquad 0 \leqq t \leqq T \tag{6}$$

where $a(\cdot)$ and $b(\cdot)$ are stochastic processes that depend at most on past values of $w(\cdot)$ and that satisfy

$$\int_0^T \overline{\mid a(s) \mid}\, ds < \infty, \int_0^T \overline{\mid b(s) \mid^2}\, ds < \infty \tag{7}$$

It is usual to write (6) symbolically in differential form as

$$dx(t) = a(t)\, dt + b(t)\, dw(t) \tag{8}$$

Note that in this notation $d(\cdot)$ and $\mathcal{J}(\cdot)$ are inverse operations, as in the usual calculus.

Now suppose $f(u, t)$ is a function of two variables with continuous second-order partial derivatives in u and t, which we shall denote by f_t, f_u, f_{uu}, etc. Then Itô's differential rule [3], p. 24, states that $f(x(t), t)$, where $x(t)$ is the Itô process (6) will also be an Itô process defined by

$$df(x, t) = f_t(x, t)\, dt + f_{uu}(x, t)\, dx + \tfrac{1}{2} f_{uu}(x, t) b^2(t)\, dt \quad (9)$$

This formula can be heuristically obtained by a formal Taylor expansion of $f(x + dx, t + dt)$ and use of the *symbolic* relations

$$dt\, dw = 0 = dx\, dw; \qquad (dw)^2 = dt$$

We note that it is the Lévy property of $w(\cdot)$ that makes (9) different from the ordinary formula for the differential in that (9) contains the second-order term $\frac{1}{2} f_{xx}(x, t) b^2(t)\, dt$. Of course, if $b(t) \equiv 0$, then the Itô formula coincides with the usual one.

There will sometimes be a need to consider vector Itô processes

$$d\mathbf{x}(t) = \mathbf{a}(t)\, dt + \mathbf{B}(t)\, d\mathbf{w}(t) \quad (10)$$

where $\mathbf{a}$, and $\mathbf{x}$ are vectors, $\mathbf{B}$ is a matrix, and $\mathbf{w}$ is a vector of independent Wiener processes. Let $f(\mathbf{u}, t)$ be a scalar function of $\mathbf{u}$ and t and let $\mathbf{f}_u$ denote the vector of first-partial derivatives and $\mathbf{f}_{uu}$ the matrix of second-partials. Then the Itô rule is

$$df(\mathbf{x}, t) = f_t(\mathbf{x}, t)\, dt + \mathbf{f}_x'(\mathbf{x}, t)[\mathbf{a}(t)\, dt + \mathbf{B}(t)\, d\mathbf{w}(t)] + \tfrac{1}{2} \operatorname{tr} [\mathbf{B}'(t)\mathbf{f}_{xx}(x, t)\mathbf{B}(t)]\, dt \quad (11)$$

where the prime denotes transpose.

EXAMPLE

Let

$$f(L, t) = \ln L(t), \quad \text{where} \quad dL(t) = a(t)\, dt + b(t)\, dw(t) \quad (12)$$

Then

$$d \ln L(t) = \frac{dL(t)}{L(t)} - \frac{1}{2} \frac{1}{L^2(t)} b^2(t)\, dt \quad (13)$$

or equivalently

$$\int_0^T \frac{dL(t)}{L(t)}\, dt = \ln\left[\frac{L(T)}{L(0)}\right] + \frac{1}{2} \int_0^T \frac{b^2(t)}{L^2(t)}\, dt \quad (14)$$

which shows well the difference from the usual integration rules.

3. THE ESTIMATION FORMULA

If by some means we have been able to directly determine the likelihood ratio $L(T)$ for the problem (1), then the formula (3) suggests that we should be able to determine $\hat{z}(t \mid t)$ from it. This is true, because direct application of the Itô differential rule to the formula (3) yields

$$dL(t) = L(t) \cdot \hat{z}(t \mid t) \cdot dx(t)$$

Therefore

$$\hat{z}(t \mid t) = \frac{dL(t)}{L(t)} \cdot \frac{1}{dx(t)} \tag{15}$$

where $dL(t)$ and $dx(t)$ are the Itô differentials of $L(t)$ and $x(t)$, respectively. We stress that (cf. Eq. (14)) in the Itô calculus $dL(t)/L(t)$ is not equal to $d \ln L(t)$.

An Example

The formula (15) can be applied in any problem where $L(t)$ can be easily determined. A rather trivial case is where $z(t) = \alpha m(t)$, with $m(t)$ a known signal but α a Gaussian random variable. Here we shall treat the somewhat less obvious case of a "narrowband" signal of random phase: we consider

$$H_1 : \dot{x}(t) = A(t) \cos (\omega_0 t + \theta) + \dot{w}(t), \qquad H_0 : \dot{x}(t) = \dot{w}(t) \tag{15}$$

where $A(\cdot)$ and ω_0 are known but θ is a random variable with a uniform distribution over $[-\pi, \pi]$. We wish to find the least squares estimate of the nonGaussian process $z(\cdot)$ by use of the formula (5). The likelihood ratio for the problem (15) is well known (see e.g., Helstrom [4], Sec. VI.2) to be

$$L(t) = I_0(V(t)) \exp \left[-\frac{1}{4} \int_0^t A^2(\tau) \, d\tau \right] \tag{16}$$

where $I_0(\cdot)$ is the modified Bessel function and

$$V^2(t) = V_c^2(t) + V_s^2(t), V_{\left\{ c \atop s \right\}}(t) = \int_0^t A(\tau) \begin{Bmatrix} \text{Cos } \omega_0 \tau \\ \text{Sin } \omega_0 \tau \end{Bmatrix} dx(\tau)$$

To apply our formula (5) we first calculate (cf., Eq. (14))

$$\frac{dL(t)}{L(t)} = \frac{dI_0(V(t))}{I_0(V(t))} - \frac{1}{4} A^2(t) \, dt \tag{17}$$

Furthermore, after some algebra and use of the narrowband assumption, viz., that "terms in $2\omega_0$" can be dropped, we obtain

$$dV(t) = \frac{V_c(t)\, dV_c + V_s(t)\, dV_s(t)}{V(t)} + \frac{1}{4}\frac{A^2(t)}{V(t)}dt$$

$$= A(t)\, \mathrm{Cos}(\omega_0\, t + \phi(t))\, dx(t) + \frac{1}{4}\frac{A^2(t)}{V(t)}\, dt,$$

where

$$\phi(t) = -\tan^{-1}[V_s(t)/V_c(t)]$$

Then $(dV(t))^2 = (\frac{1}{2})A^2(t)\, dt + o(dt)$ and

$$dI_0(V(t)) = \frac{dI_0}{dV}dV + \frac{1}{2}\cdot\frac{d^2I_0}{dV^2}\cdot\frac{1}{2}A^2(t)\, dt$$

$$= I_1(V)\cdot A(t)\cdot \mathrm{Cos}(\omega_0 t + \phi(t))\, dx + \frac{1}{4}\frac{A^2(t)}{V(t)} \tag{18}$$

$$\cdot\left[I_1(V) + V\frac{dI_1}{dV}\right]dt.$$

Finally, by using the Bessel function identity

$$I_1(V) + V\frac{dI_1}{dV} = VI_0(V)$$

and combining (18) and (17) we obtain

$$\hat{z}(t \mid t) = \frac{dL(t)}{L(t)}\cdot\frac{1}{dx(t)} = \frac{I_1(V)}{I_0(V)}A(t)\, \mathrm{Cos}\,[\omega_0\, t + \phi(t)] \tag{19}$$

This formula can be directly verified, e.g., by use of Bayes' rule.

4. THE DISCRETE-TIME CASE

A discussion of the likelihood ratio formula (1) with R. Esposito led him to develop [2] a relation in the discrete-time case between the likelihood ratio and the noncausal estimate of a random signal in Gaussian noise. Thus consider the hypotheses

$$H_1 : \mathbf{x} = \mathbf{z} + \mathbf{n},\ H_0 : \mathbf{x} = \mathbf{n} \tag{20}$$

where $\mathbf{n}$ is a (zero-mean) Gaussian vector with identity covariance matrix and $\mathbf{z}$ is an independent random vector with density function $p_z(\cdot)$, say.

Then the likelihood ratio is readily seen to be

$$\Lambda(\mathbf{x}) = \int \exp \{\mathbf{u}' \mathbf{x} - \tfrac{1}{2} \mathbf{u}' \mathbf{u}\} \, p_z(\mathbf{u}) \, d\mathbf{u} \tag{21}$$

The least-squares estimate of $\mathbf{z}$ given the whole vector $\mathbf{z} + \mathbf{n} = \mathbf{x}$ is well known to be the conditional mean.

$$\hat{\mathbf{z}} = \frac{\int \mathbf{u} \exp \{\mathbf{u}' \mathbf{x} - 1/2 \, \mathbf{u}' \mathbf{x}\} \, p_z(\mathbf{u}) \, d\mathbf{u}}{\int \exp\{\mathbf{u}' \mathbf{x} - 1/2 \, \mathbf{u}' \mathbf{u}\} \, p_z(\mathbf{u}) \, d\mathbf{u}} \tag{22}$$

The expressions in (21) and (22) seem closely related; in fact, by direct differentiation of $\Lambda(\mathbf{x})$ with respect to $\mathbf{x}$ we have

$$\nabla_x (\mathbf{x}) = \int \mathbf{u} \exp \{\mathbf{u}' \mathbf{x} - \tfrac{1}{2} \mathbf{u}' \mathbf{u}\} \, p_z(\mathbf{u}) \, d\mathbf{u} = \mathbf{z}\Lambda(\mathbf{x}) \tag{23}$$

so that

$$\hat{\mathbf{z}} = \frac{\nabla_x(\mathbf{x})}{\Lambda(\mathbf{x})} = \nabla_{\mathbf{x}} \ln \Lambda(\mathbf{x}) \tag{23}$$

This is Esposito's relation, which he derived somewhat less directly. There is some similarity between (23) and our continuous-time formula (5). However, note that in discrete time there is no Itô rule and $\nabla_x L/L$ can be written $\nabla x \ln L$. But the important difference is that there is no discrete-time formula for the L.R. corresponding to our general continuous-time formula (4). From (23) all we can conclude is that

$$L(\mathbf{x}) = \exp [\hat{\mathbf{z}}' \, d\mathbf{x} + \text{constant}] \tag{24}$$

The formula (24) uses *noncausal* estimates and is not as explicit as (4); moreover, unless an analytical expression is available for $\hat{\mathbf{z}}$, so that the integral can be evaluated analytically, it does not seem possible to implement (24) for any given observation $\mathbf{x}$. On the other hand, in (4), the integrations are with respect to time and moreover (4) yields a receiver structure into which suboptimum estimates for $z(\cdot)$ can be easily introduced. However, it should be of value to study the relationships between the discrete-time and continuous-time analyses in more detail and in particular to see how to carry one over into the other.

In conclusion we note that the above results can be extended to the case of colored additive Gaussian noise by use of a noise-whitening filter; the results can also be easily extended to the vector case.

RECEIVED: August 7, 1968; revised October 10, 1968

REFERENCES

ESPOSITO, R., On a relation between detection and estimation in decision theory. *Inform. Control* **12**, (1968) 116–120.

KAILATH, T., A general likelihood-ratio formula for random signals in gaussian noise. *IEEE Trans. on Inform. Theory*. To appear in 1969.

SKOROKHOD, A. V., "Studies in the Theory of Random Processes." Addison Wesley, Massachusetts, 1965. [translation].

HELSTROM, C. W., "Statistical Theory of Signal Detection." Pergamon Press, 1960.