VIII. DETECTION AND ESTIMATION THEORY*

Academic and Research Staff

Prof. H. L. Van Trees
Prof. A. B. Baggeroer
Prof. D. L. Snyder

Graduate Students

I. L. Bruyland
A. E. Eckberg
J. E. Evans
T. P. McGarty
J. R. Morrissey

A. LIKELIHOOD RATIOS AND MUTUAL INFORMATION FOR POISSON PROCESSES WITH STOCHASTIC INTENSITY FUNCTIONS

1. Introduction

In this report we formally obtain some results regarding likelihood ratios and mutual information for (conditional) Poisson processes with stochastic intensity functions. Our results for likelihood ratios are an extension of results obtained recently by Snyder[1,2] by a somewhat different approach (Snyder considers the case where the stochastic intensity function is Markovian while our results include the case of non-Markovian stochastic intensities). To the best of our knowledge, the results for mutual information are new.

2. Likelihood Ratios

We consider the following problem. Given the observations $\{N(u), u \leq u \leq t\}$ of a conditional Poisson counting process, determine so as to minimize an expected risk function, which of the following hypotheses is true.[3]

$$H_1: \lambda_r(t) = \lambda_o + \lambda(t) \tag{1}$$

$$H_o: \lambda_r(t) = \lambda_o, \tag{2}$$

where $\lambda_r(t)$ is the (stochastic) intensity function of N(t). λ_o might represent the rate of arrival of photons from background radiation. Following Snyder,[2] by a conditional Poisson counting process, we mean that if $\{\lambda_r(t), t \geq 0\}$ is given, then $\{N(t), t \geq 0\}$ is an integer valued process with independent increments, and for $t \geq 0$,

*This work was supported in part by the Joint Services Electronics Programs (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DA 28-043-AMC-02536(E).

$$\Pr\left[N(t+\Delta t)-N(t)=1 \,|\, \lambda_r(t)\right] = \lambda_r(t)\,\Delta t + o(\Delta t)$$
$$\Pr\left[N(t+\Delta t)-N(t)=0 \,|\, \lambda_r(t)\right] = 1 - \lambda_r\Delta t + o(\Delta t). \tag{3}$$

The method by which we obtain the likelihood ratio for the detection problem is quite similar to that used by Duncan[4-6] to obtain results for additive white Gaussian noise channels. We first consider the likelihood ratio conditional on knowledge of $\lambda_r(t)$. This is equivalent to the detection problem with known intensity function and the form of the likelihood ratio is well known.[7-13] The unconditional likelihood ratio is then obtained by averaging over $\lambda(t)$ and utilizing certain properties of stochastic integrals.

The likelihood ratio for the detection problem above conditional on knowledge of $\lambda(t)$ has been shown by Reiffen and Sherman[7] to be

$$\frac{p\left(N_o^t \,|\, H_1, \lambda\right)}{p\left(N_o^t \,|\, H_o\right)} = \chi(t) = \exp\left[-\int_0^t \lambda(u)\,du + \int_0^t \ln\left[\frac{\lambda_o+\lambda(u)}{\lambda_o}\right] dN(u)\right], \tag{4}$$

where

$p\left(N_o^t \,|\, H_1, \lambda\right)$ = conditional probability density of the observed counting process $N_o^t = \{N(u), 0 \leq u \leq t\}$, given $\lambda(t)$, under the assumption that H_1 is true.

$p\left(N_o^t \,|\, H_o\right)$ = conditional probability density of N_o^t, under the assumption that H_o is true.

The second integral on the right-hand side of (4) (and all other integrals involving $dN(u)$) will be interpreted in the sense introduced by Itô (see Doob[14]).

Before averaging (3) over possible paths of $\lambda(t)$, we shall put (3) into a more convenient form by use of the Itô differential rule. Snyder[2] has shown that if z_t is a scalar Markov process satisfying the stochastic differential equation,

$$dz_t = a_t\,dt + b_t\,dN_t, \tag{5}$$

where a_t and b_t are nonanticipative functionals of the Poisson counting process N_t, and $\phi_t(z_t)$ is a differentiable function of z_t and t, then

$$d\phi_t(z_t) = \left[\frac{\partial}{\partial t}\phi_t(z_t) + \frac{\partial}{\partial z_t}\phi_t(z_t)\,a_t\right] dt + \left[\phi_t(z_t+b_t)-\phi_t(z_t)\right]dN_t. \tag{6}$$

Applying (6) to (4), we find that $\chi(t)$ satisfies the following stochastic differential equation:

$$d\chi(t) = -\chi(t)\,\lambda(t)\,dt + \left[\chi(t)\,e^{\ln\frac{\lambda_o+\lambda(t)}{\lambda_o}} - \chi(t)\right] dN(t)$$
$$= -\chi(t)\,\lambda(t)\,dt + \frac{1}{\lambda_o}\chi(t)\,\lambda(t)\,dN(t), \tag{7}$$

or equivalently

$$\chi(t) = \chi(o) - \int_0^t \chi(u)\,\lambda(u)\,du + \frac{1}{\lambda_o}\int_0^t \chi(u)\,\lambda(u)\,dN(u). \tag{8}$$

We now take the expectation of both sides of (7) with respect to all possible paths of $\lambda(t)$. From (4), we see that $\chi(t)$ is related to the unconditional likelihood ratio, $\Lambda(t)$, by

$$\Lambda(t) = E_\lambda[\chi(t)], \tag{9}$$

where $E_\lambda[\cdot]$ denotes averaging over all sample paths of $\lambda(t)$. Thus

$$\Lambda(t) = \Lambda(o) - E_\lambda\left[\int_0^t \chi(u)\,\lambda(u)\,du\right] + \frac{1}{\lambda_o}E_\lambda\left[\int_0^t \chi(u)\,\lambda(u)\,dN(u)\right]$$
$$= \Lambda(o) - \int_0^t E_\lambda[\chi(u)\lambda(u)]\,du + \frac{1}{\lambda_o}\int_0^t E_\lambda[\chi(u)\lambda(u)]\,dN(u). \tag{10}$$

In Eqs. 8-10, it is important to keep in mind that N(u) is the observation and thus fixed, as far as the expectation in (9) and (10) is concerned.

The stochastic differential equation corresponding to (10) is

$$d\Lambda(t) = -E_\lambda[\chi(t)\lambda(t)]\,dt + \frac{1}{\lambda_o}E_\lambda[\chi(t)\lambda(t)]\,dN(t)$$
$$= -\frac{E_\lambda[\chi(t)\lambda(t)]}{E_\lambda[\chi(t)]}\Lambda(t)\,dt + \frac{1}{\lambda_o}\frac{E_\lambda[\chi(t)\lambda(t)]}{E_\lambda[\chi(t)]}\Lambda(t)\,dN(t). \tag{11}$$

But

$$\frac{E_\lambda[\chi(t)\lambda(t)]}{E_\lambda[\chi(t)]} = \frac{E_\lambda\left[\lambda(t)p\left(N_o^t|\lambda, H_1\right)\right]}{E_\lambda\left[p\left(N_o^t|\lambda, H_1\right)\right]} = \hat{\lambda}(t) = E\left[\hat{\lambda}(t)|N_o^t, H_1\right]$$

= conditional expectation of $\lambda(t)$, given N_o^t, under the assumption that H_1 is true.

= least-squares realizable estimate of $\lambda(t)$, under the assumption that H_1 is true.

Thus, we find that the likelihood ratio, $\Lambda(t)$, is the solution to the stochastic differential equation

$$d\Lambda(t) = -\hat{\lambda}(t)\,\Lambda(t)\,dt + \frac{1}{\lambda_o}\hat{\lambda}(t)\,\Lambda(t)\,dN(t). \tag{13}$$

The solution to (13) is

$$\Lambda(t) = \exp\left[-\int_0^t \hat{\lambda}(u)\,du + \frac{1}{\lambda_o}\int_0^t \ln\left[\frac{\lambda_o + \hat{\lambda}(u)}{\lambda_o}\right] dN(u)\right]. \tag{14}$$

Our result (14) can be verified by application of the Itô differential rule for Poisson counting processes to (14) and noting that we obtain (13). Furthermore, (14) clearly has the proper value at $t = 0$.

Comparing (14) with (4), we see that the likelihood ratio for stochastic intensity functions is equivalent to that for known intensity functions, in the sense that the causal minimum mean-square estimate of the intensity is incorporated in the detector in the same way as if it were known. An analogous relationship has been obtained for stochastic signals on additive white Gaussian noise channels by Duncan.[4, 5]

The form above for the likelihood ratio is identical to that obtained by Snyder[2] for Markovian $\lambda(t)$. Snyder[1] also has some results concerning means of obtaining causal minimum mean-square estimates of $\lambda(t)$.

3. Mutual Information Results

We shall now apply our results for likelihood ratios to the problem of computing the mutual information between the message process m and the conditional Poisson counting process $\{N(t), t \geq 0\}$ when the stochastic intensity function of $N(t)$ is $\lambda_o + \lambda(m, t)$. Our approach is quite similar to that of Duncan[6] and Zakai[15] who obtained mutual information results for additive white Gaussian noise channels: We convert the problem of calculating certain probabilities into a detection problem by introducing a dummy hypothesis corresponding to H_o of the previous section. By using appropriate likelihood ratio results and some properties of stochastic integrals, we are able to obtain results in terms of certain causal filtering estimates.

Let us define

N_o^t = channel output in time interval $[u, t]$ = counting process

$\lambda_r(t) = \lambda(t, m) + \lambda_o$ = intensity function of $N(t)$

m = message.

The mutual information between N_o^t and m is given (see Gelfand and Yaglom[15] by

$$I\left[N_o^t, m\right] = E\left[\log \frac{p\left(N_o^t | m\right)}{p\left(N_o^t\right)}\right]. \tag{15}$$

The likelihood ratio appearing in the expression for the mutual information can be written

$$\frac{p\left(N_o^t | m\right)}{p\left(N_o^t\right)} = \frac{p\left(N_o^t | m\right)}{\tilde{p}\left(N_o^t\right)} \cdot \frac{1}{\dfrac{p\left(N_o^t\right)}{\tilde{p}\left(N_o^t\right)}}, \tag{16}$$

where

$\dfrac{p\left(N_o^t | m\right)}{\tilde{p}\left(N_o^t\right)}$ = likelihood ratio for detection between the hypothesis that $\lambda_r(t) = \lambda(t, m) + \lambda_o$ with the particular m and the hypothesis that $\lambda_r(t) = \lambda_o$.

$\dfrac{p\left(N_o^t\right)}{\tilde{p}\left(N_o^t\right)}$ = likelihood ratio for detection between the hypothesis $\lambda_r = \lambda(t, m) + \lambda_o$ with m random and the hypothesis that $\lambda_r = \lambda_o$.

Now, from section 2,

$$\frac{p\left(N_o^t | m\right)}{\tilde{p}\left(N_o^t\right)} = \exp\left\{-\int_0^t \lambda(u, m)\, du + \int_0^t \ln\left[\frac{\lambda_o + \lambda(u, m)}{\lambda_o}\right] dN(u)\right\} \tag{17}$$

$$\frac{p\left(N_o^t\right)}{\tilde{p}\left(N_o^t\right)} = \exp\left\{-\int_0^t \hat{\lambda}(u)\, du + \int_0^t \ln\left[\frac{\lambda_o + \hat{\lambda}(u)}{\lambda_o}\right] dN(u)\right\}, \tag{18}$$

where $\hat{\lambda}$ is the conditional mean of $\lambda(u, m)$, given $N(u)$, $o \leq u \leq t$. That is,

$$\hat{\lambda}(t) = E\left[\lambda(t, m) \,|\, N_o^t, \lambda_r(t) = \lambda(t, m) + \lambda_o\right]. \tag{19}$$

Thus we have

$$I\left(N_o^t, m\right) = E\left\{\int_0^t [-\lambda(u,m)+\hat{\lambda}(u)]\, du + \int_0^t \ln\left[\frac{\lambda_o+\lambda(u,m)}{\lambda_o+\hat{\lambda}(u)}\right] dN(u)\right\}. \tag{20}$$

The first integral can be shown to be zero by utilizing the properties of a conditional expectation:

$$E\left\{\int_0^t [\lambda(u,m)-\hat{\lambda}(u)]\,du\right\} = E\left[\int_0^t \left\{E\left[\lambda(u,m)\,|\,N_o^t\right]-\hat{\lambda}(u)\right\} du\right] = 0. \tag{21}$$

To evaluate the second integral, we write

$$\int_0^t \ln\left[\frac{\lambda_o+\lambda(u,m)}{\lambda_o+\hat{\lambda}(u)}\right] dN(u) =$$

$$\int_0^t \ln\left[\frac{\lambda_o+\lambda(u,m)}{\lambda_o+\hat{\lambda}(u)}\right] [dN(u)-\lambda(u,m)du] + \int_0^t \ln\left[\frac{\lambda_o+\lambda(u,m)}{\lambda_o+\hat{\lambda}(u)}\right] \lambda(u,m)\, du.$$

Now

$$E\int_0^t \ln\left[\frac{\lambda_o+\lambda(u,m)}{\lambda_o+\hat{\lambda}(u)}\right] [dN(u)-\lambda(u,m)du] = 0 \tag{22}$$

from the definition of the stochastic integral, since $N(u) - \int_0^u \lambda(\tilde{u},m)\, d\tilde{u}$ is a martingale (Doob[14]).

Thus we have

$$I\left[N_o^t, m\right] = \int_0^t E\left\{\ln\left[\frac{\lambda_o+\lambda(u,m)}{\lambda_o+\hat{\lambda}(u)}\right] \lambda(u,m)\right\} dt. \tag{23}$$

Next, we note that

$$I\left[N_o^t, m\right] = \int_0^t E\{\ln[\lambda_o+\lambda(u,m)]\lambda(u,m) - \ln[\lambda_o+\hat{\lambda}(u)]\lambda(u,m)\}\, du.$$

Again taking expectations conditional on N_o^t we obtain

$$I\left[N_o^t, m\right] = \int_0^t E\left[\widehat{\ln[\lambda_o+\lambda(u)]\lambda(u)} - \ln[\lambda_o+\hat{\lambda}(u)]\hat{\lambda}(u)\right] du, \tag{24}$$

where

$\widehat{\ln [\lambda_o + \lambda(u)] \lambda(u)}$ = realizable least-squares estimate of $\lambda(u,m) \ln [\lambda_o + \lambda(u,m)]$, given the observations $\{N(s), 0 \leq s \leq u\}$, under the assumption that $\lambda_r(t) = \lambda(t,m) + \lambda_o$.

One can readily verify that $I\left[N_o^t, m\right]$ is a monotone increasing function of t.

In the case of large background intensity, that is,

$$\lambda_o \gg \lambda(u,m)$$

$$\lambda_o \gg \hat{\lambda}(u),$$

we can relate $I\left[N_o^t, m\right]$ to a realizable least-squares filtering error by using the expansion

$$\ln [\lambda_o + \chi] = \ln \lambda_o + \chi/\lambda_o$$

to obtain

$$I\left[N_o^t, m\right] = \frac{1}{\lambda_o} \int_0^t E[\widehat{\lambda^2}(u) - \hat{\lambda}^2(u)]\, du = \frac{1}{\lambda_o} \int_0^t E\{p(u)\}\, du, \tag{25}$$

where p(u) is the conditional variance of the estimate $\hat{\lambda}(u)$.

A somewhat more interesting result can be obtained by using this expansion on Eq. 23 to obtain

$$I\left[N_o^t, m\right] = \frac{1}{\lambda_o} \int_0^t E\{[\lambda(u,m) - \hat{\lambda}(u)] \lambda(u,m)\}\, du. \tag{26}$$

The error, $\lambda(u,m) - \hat{\lambda}(u)$, is seen to be orthogonal to $\hat{\lambda}(u)$ as follows:

$$E\{[\lambda(u,m) - \hat{\lambda}(u)]\hat{\lambda}(u)\} = E\left(E\left\{[(u,m) - \hat{\lambda}(u)]\hat{\lambda}(u) \,|\, N_o^t\right\}\right) = 0,$$

so that (12) is equivalent to

$$I\left[N_o^t, m\right] = \frac{1}{\lambda_o} \int_0^t E\{[\lambda(t,m) - \hat{\lambda}(u)]^2\}\, du. \tag{27}$$

Equation 27 is our desired result for the relationship between the mutual information and the realizable filtering error for the estimation of $\lambda(t)$.

A result quite similar to (27) was obtained for the mutual information on the white Gaussian channel with (or without) feedback by Duncan and Zakai. They consider the problem

$$y(t) = \int_0^t \phi\left(s, y_o^s, m\right) ds + \omega(t), \tag{28}$$

where m denotes the message, y(t) denotes the channel output at time t, y_o^s denotes the path y(u), $o \leq u \leq s$, ω(t) is a Brownian motion, and m is the channel input. The result that they obtain is that the amount of information between the message m and the output path y_o^T, $I\left[y_o^T, m\right]$, is related to a causal mean-square filtering error by

$$I\left[y_o^T, m\right] = \frac{1}{2}\int_0^T E\left\{\left[\phi\left(s, y_o^s, m\right) - \hat{\phi}\left(s, y_o^s\right)\right]^2\right\} ds, \tag{29}$$

where

$$\hat{\phi}\left(s, y_o^s\right) = E\left[\phi\left(s, y_o^s, m\right) \mid y_o^s\right].$$

4. Summary

We have formally derived some results regarding likelihood ratios and mutual information for (conditional) Poisson processes with stochastic intensity functions. The results emphasize the role of realizable least-squares estimates of the stochastic intensity. Several comments on extensions and utilization of the results presented here are appropriate.

1. The derivation here is formal; for example, we have assumed that several interchanges of integration and expectation are valid without giving sufficient conditions. It would be of interest to rigorously establish these results by using appropriate Radon-Nikodym derivative results,[17, 18] Borel fields, and so forth.

2. We can readily extend our result to cover the case of feedback channels, by making a minor extension of the likelihood ratio results. Vector channels can also be readily considered, by minor modifications to the likelihood-ratio results.

3. These results may be of interest in optical communication,[9] biomedical data processing,[1] and studies of auditory psychophysics.[19]

J. E. Evans

[Mr. James E. Evans is an M. I. T. Lincoln Laboratory Staff Associate.]

References

1. D. Snyder, "Estimation of Stochastic Intensity Functions of Conditional Poisson Processes," Monograph No. 128, Washington University Biomedical Computer Laboratory, St. Louis, Missouri, April 1970.

2. D. Snyder "Detection of Nonhomogeneous Poisson Processes Having Stochastic Intensity Functions," Monograph No. 129, Washington University Biomedical Computer Laboratory, St. Louis, Missouri, April 1970.

3. The likelihood ratio when $\lambda_r = \lambda_o + \hat{\lambda}(t)$ on hypothesis H_o can be obtained by using the chain rule for likelihood ratios. See Snyder[2] for details.

4. T. E. Duncan, "Evaluation of Likelihood Functions," Inform. Contr. 13, 62-74 (1968).

5. T. E. Duncan, "Likelihood Functions for Stochastic Signals in White Noise" (unpublished paper, dated 1969).

6. T. E. Duncan, "On the Calculation of Mutual Information" (submitted for publication).

7. B. Reiffen and H. Sherman, "An Optimum Demodulator for Poisson Processes: Photon Source Detectors," Proc. IEEE 51, 1316-1320 (1963).

8. J. W. Goodman, "Optimum Processing of Photoelectron Counts," IEEE J. Quantum Electronics, Vol. QE-1, pp. 180-181, July 1965.

9. R. M. Gagliardi and S. Karp, "N-ary Poisson Detection and Optical Communications," IEEE Trans., Vol. COM-17, No. 2, pp. 208-216, April 1969.

10. G. L. Fillmore and G. Lachs, "Information Rates for Photocount Detection Systems," IEEE Trans., Vol. IT-15, No. 4, pp. 465-468, July 1969.

11. J. Bar-David, "Communication under Poisson Regime," IEEE Trans., Vol. IT-15, No. 1, pp. 31-37, January 1969.

12. R. W. Chang, "Photon Detection for an Optical Pulse Code Modulation System," IEEE Trans., Vol. IT-15, pp. 725-728, November 1969.

13. K. Abend, "Optimum Photon Detection," IEEE Trans., Vol. IT-12, pp. 64-65, January 1966.

14. J. L. Doob, Stochastic Processes (John Wiley and Sons, Inc., New York, 1953).

15. M. Zakai, "On the Mutual Information of the White Gaussian Channel with or without Feedback" (submitted for publication).

16. I. M. Gelfand and A. M. Yaglom, "Calculation of the Amount of Information about a Random Function Contained in Another Such Function," Am. Math. Soc. Translations (Series 2) 12, 199-246 (1959).

17. A. V. Skorokhod, "On the Differentiability of Measures Which Correspond to Stochastic Processes I. Processes with Independent Increments," in Theory of Probability and Its Applications, Vol. 2, No. 4, pp. 407-432, 1957.

18. A. V. Skorokhod, "Absolute Continuity of Infinitely Divisible Distributions under Translations," in Theory of Probability and Its Applications, Vol. 10, No. 4, pp. 465-472, 1965.

19. W. M. Siebert, "Stimulus Transformations in the Peripheral Auditory System," in P. A. Kolers and M. Eden (eds.), Recognizing Patterns in Living and Automatic Systems (The M.I.T. Press, Cambridge, Mass., 1968), pp. 104-133.