AFOSR 69-1112TR

AD686819

ON THE CALCULATION OF MUTUAL INFORMATION

Tyrone E. Duncan

Information and Control Engineering
The University of Michigan
Ann Arbor, Michigan 48104

1968

13

ON THE CALCULATION OF MUTUAL INFORMATION

Tyrone E. Duncan*

1. Introduction

Calculating the amount of information about a random function contained in another random function has important uses in communication theory. An expression for the mutual information for continuous time random processes has been given by Gelfand and Yaglom [1], Chiang [2] and Perez [3] by generalizing Shannon's result [4] in a natural way. Under a condition of absolute continuity of measures the continuous time expression has the same form as Shannon's result. For two Gaussian processes Gelfand and Yaglom express the mutual information in terms of a mean square estimation error. We generalize this result to diffusion processes and express the solution in a different form which is more naturally related to a corresponding filtering problem. We also use these results to calculate some information rates.

2. Problem Statement

We shall consider two random processes which are the solutions of the following stochastic differential equations

$$dX_t = a(t, X_t)dt + b(t, X_t)dB_t \qquad (1)$$

$$dY_t = c(t, X_t)dt + h(t)d\tilde{B}_t \qquad (2)$$

where the solution is obtained for the interval $[0, 1]$ and (for notational simplicity) $X_0 = 0$ and $Y_0 = 0$. The processes $\{X_t\}$ and $\{Y_t\}$ are n and m dimensional respectively and the processes $\{B_t\}$ and $\{\tilde{B}_t\}$ are n and m dimensional standard Brownian motions. The elements of the vectors a and c and the matrices b and h are continuous in t and globally Lipschitz continuous in X_t. The inverse of the matrix $h(t)$, $h^{-1}(t)$, exists and is continuous for all $t \in [0, 1]$. We wish to calculate the amount of information in the

*Information and Control Engineering, The University of Michigan, Ann Arbor, Michigan 48104. This research was supported by United States Air Force Grant AF-AFOSR 814-66.

process $\{Y_t\}$ about the process $\{X_t\}$.

3. Preliminaries

In order to calculate the mutual information we must determine some appropriate Radon-Nikodym derivatives. We shall use the following result due to Girsanov [5].

Theorem 1: Suppose that

$$dX_t = a(t, X_t)dt + b(t, X_t)dB_t$$

$$dY_t = (a(t, Y_t) + b(t, Y_t)h(t, Y_t))dt + b(t, Y_t)dB_t$$

where

i) $t \in [s, 1]$, $X(s) = Y(s)$

ii) a and h are n vectors and b is an $n \times n$ matrix

iii) $a(\cdot, \cdot)$, $b(\cdot, \cdot)$ and $h(\cdot, \cdot)$ are measurable in both variables. In particular a and b are continuous in their first variable and globally Lipschitz continuous in their second variable.

iv)

$$\int_s^1 |h(t, X_t)|^2 dt < \infty \qquad \text{a.s.}$$

v) $|h(t, X_t)| < h_0(\sup_t |X_t|)$

where h_0 is a nondecreasing function of a real variable.

Then the measures μ_X and μ_Y induced on $C_n[s, 1]$ (the space of all continuous functions with values in $\mathcal{R}^n$) by $\{X_t\}$ and $\{Y_t\}$ respectively, are mutually absolutely continuous.

The Radon-Nikodym derivative $d\mu_Y/d\mu_X$ is given by

$$\frac{d\mu_Y}{d\mu_X} = \exp\left[\int_s^1 h^T(u, X_u)dB_u - \tfrac{1}{2}\int_s^1 |h(u, X_u)|^2 du\right] \tag{3}$$

We shall also use a result of Duncan [6] giving the expression for the likelihood function of a related detection problem.

Theorem 2: Consider the following detection problem

$$
\begin{aligned}
dY_t &= c(t, X_t)dt + h(t)d\tilde{B}_t \qquad &&\text{for signal present} \\
&= \qquad\qquad\qquad\;\; h(t)d\tilde{B}_t \qquad &&\text{for signal not present}
\end{aligned}
\tag{4}
$$

where X_t is the solution of (1) with the assumptions indicated there

Then the Radon-Nikodym derivative, Λ_t, for this detection problem is given by

$$
\begin{aligned}
\Lambda_t &= E_{\mu_X}\left\{\exp\left[\int_0^t c^T(u, X_u)h_u^{-1^T}\, d\tilde{B}_u - \tfrac{1}{2}\int_0^t c^T(u, X_u)g_u^{-1}c(u, X_u)du\right]\right\} \\
&= \tilde{\psi}_t = \exp\left[\int_0^t \hat{c}^T(u, X_u)h_u^{-1^T}\, d\tilde{B}_u - \tfrac{1}{2}\int_0^t \hat{c}^T(u, X_u)g_u^{-1}\hat{c}(u, X_u)du\right]
\end{aligned}
\tag{5}
$$

where E_{μ_X} corresponds to integration with respect to the measure μ_X generated by the solution of (1), $\hat{c}(t, X_t) = E[c(t, X_t)|Y_u,\ 0 \le u \le t]$ (the conditional expectation of $c(t, X_t)$ given the augmented Borel field generated by $\{Y_u,\ 0 \le u \le t\}$), and $g = h^T h$.

Generalizations of Shannon's mutual information have been discussed by Gelfand and Yaglom [1], Chiang [2] and Perez [3] They obtain the following result as the natural extension of Shannon's mutual information.

Theorem 3: Let ξ and η be two random vectors. The mutual information $J(\xi, \eta)$ is given by

$$
J(\xi, \eta) = \int \alpha(x, y)\log \alpha(x, y)dP_\xi(x)dP_\eta(y) \tag{6}
$$

where

$$
\alpha(x, y) = \frac{dP_{\xi\eta}(x, y)}{dP_\xi(x)dP_\eta(y)}
$$

4. Main Result

We have now established sufficient preliminaries to obtain the main result of this paper

Theorem 4: Consider the processes $\{X_t\}$ and $\{Y_t\}$ obtained as solutions of (1) and (2). The mutual information contained in $\{Y_u, 0 \le u \le 1\}$ about $\{X_u, 0 \le u \le 1\}$ is given by the following expression

$$J(X, Y) = \tfrac{1}{2} E \int_0^1 [c(u, X_u) - \hat{c}(u, X_u)]^T g^{-1}(u)[c(u, X_u) - \hat{c}(u, X_u)]du \tag{7}$$

where $\hat{c}$ is defined in Theorem 2.

Proof: To calculate the mutual information we must compute an appropriate Radon-Nikodym derivative. Let

$$\Phi = \frac{d\mu_{XY}}{d\mu_X d\mu_Y}$$

where μ_{XY} is the product measure generated by (1) and (2) and μ_X and μ_Y are the marginals. By a simple calculation, essentially using only the absolute continuity results of Theorems 1 and 2 we have

$$\frac{d\mu_{XY}}{d\mu_X d\mu_Y} = \frac{\psi_t}{\tilde{\psi}_t}$$

where $\tilde{\psi}_t = E_{\mu_X}[\psi_t]$ is given in Theorem (2) for the detection problem. Thus

$$J(X, Y) = \int \Phi \log \Phi d\mu_X d\mu_Y$$

$$\log \Phi = \int [c(t, X_t) - \hat{c}(t, X_t)]^T g_t^{-1} dY_t$$

$$- \tfrac{1}{2}\int [c^T(t, X_t) g_t^{-1} c(t, X_t) - \hat{c}^T(t, X_t) g_t^{-1} \hat{c}(t, X_t)]dt$$

Substituting $dY_t = c(t, X_t)dt + h(t)d\tilde{B}_t$ and using the fact that the integral with respect to $\tilde{B}_t$ is a martingale we have

$$J(X,Y) = E\left\{ \int [c^T(t,X_t)g_t^{-1} c(t,X_t) - \hat{c}^T(t,X_t)g_t^{-1}c(t,X_t)]dt \right.$$

$$\left. - \tfrac{1}{2}\int [c^T(t,X_t)g_t^{-1}c(t,X_t) - \hat{c}^T(t,X_t)g_t^{-1}\hat{c}(t,X_t)]dt \right\}$$

$$= \tfrac{1}{2}E\int (c - \hat{c})^T g^{-1}(c - \hat{c})dt \qquad \blacksquare$$

Remark 1. This result is in a different form from that obtained by Gelfand and Yaglom for Gaussian processes but by using some resolvent identities obtained by Siegert [7] we can show the equivalence of the two results and the relation of Siegert's work to the recursive linear filtering of Kalman and Bucy [8].

Remark 2. Some obvious simple extensions of the above result are, for example, letting $c(t, X_t)$ be a function of Y_t - actually c can be a functional of the past of both the processes $\{X_t\}$ and $\{Y_t\}$ with a suitable function space Lipschitz assumption (cf K. Itô and Nisio [9]).

Remark 3. If we let $c(t, X_t) = X_t$ and consider the process $\{Y_t\}$ as observations of the process $\{X_t\}$ in noise, then twice the mutual information is merely the integral of the trace of the optimal mean square filtering error for estimating $\{X_t\}$ from $\{Y_t\}$. In the one dimensional case the "new data" ($d\tilde{B}_t$) is weighted according to the additional amount of conditional mutual information it possesses (cf Kushner [10] or Duncan [11]).

5. An Application to Information Rate

We shall consider a Gaussian problem in more detail and obtain some results for information rate. These results extend and simplify some results of Gelfand and Yaglom [1] and indicate rates of convergence for some of their approximations. The methods used here require only time-domain techniques which indicate more clearly the necessary properties for the existence of the information rates.

First, though, we give the definition that we shall use for information rate (cf Gelfand and Yaglom [1] or Pinsker [12]).

Definition: The rate of generation of information about a process η by a process ξ is

$$\overline{I}(\xi, \eta) = \lim_{T\to\infty} \frac{1}{T} J(\xi_0^T, \eta_0^T) \tag{8}$$

where ξ_0^T and η_0^T denote the processes on the interval $[0, T]$ and $\overline{I}$ is only defined when the limit exists.

We shall obtain a result for the existence of information rate in terms of some system theory results. This result will indicate some useful bounds on approximations that one obtains by using finite time calculations for information rate.

We shall calculate the rate of generation of information about a Gaussian process $\{X_t\}$ by another Gaussian process $\{Y_t\}$. Specifically we have the following equations

$$dX_t = a(t)X_t dt + b(t)dB_t \tag{9}$$

$$dY_t = f(t)X_t dt + d\widetilde{B}_t \tag{10}$$

where $\{B_t\}$ and $\{\widetilde{B}_t\}$ are independent n and m dimensional Brownian motions respectively. We shall assume also that the matrices a, b, and f have elements which are continuous functions of t. The interval of solution is the half line $[0, \infty)$. The initial conditions are $X_0 = \alpha$, a zero mean Gaussian random vector, and $Y_0 = 0$.

We shall also consider the case where the coefficients of the stochastic differential equations (9) and (10) are not functions of time. In this case we shall use the same symbols for the coefficients deleting the variable t, i.e.,

$$dX_t = aX_t dt + bdB_t \tag{11}$$

$$dY_t = fX_t dt + dB_t \tag{12}$$

where the appropriate assumptions for (9) and (10) are still in effect.

Whenever the process $\{Y_t\}$ is a process of observations from which we wish to obtain a best mean square estimate of the process $\{X_t\}$, then we have a well known filtering problem. In fact, by Theorem 4 the mutual information for (9) and (10) (or (11) and (12)) is obtained from the integral of the trace of the error covariance matrix for this filtering problem. What we intend to show is that this mean square error converges to a steady-state solution which will then give us the appropriate information rate.

In the subsequent discussion we shall use the following definitions [8,13].

Definition: The system (9) and (10) is <u>uniformly completely observable</u> if there exist fixed positive constants σ, α and β such that

$$0 < \alpha I \leq M(t - \sigma, t) \leq \beta I$$

for all t where

$$M(t_1, t_2) = \int_{t_1}^{t_2} \Phi^T(t, t_2) f^T(t) f(t) \Phi(t, t_2)\, dt \tag{13}$$

For symmetric matrices $A < B (A \leq B)$ implies that $A - B$ is positive definite (non-negative definite). The matrix Φ is the transition matrix for the ordinary differential equation

$$\frac{dX}{dt} = a(t)X$$

Definition: The system (9) and (10) is <u>uniformly completely controllable</u> if there exist fixed positive constants σ, α, β such that

$$0 < \alpha I \leq W(t - \sigma, t) \leq \beta I$$

where

$$W(t_1, t_2) = \int_{t_1}^{t_2} \Phi(t_2, t) b(t)\, b^T(t) \Phi^T(t_2, t)\, dt \tag{14}$$

Definition: The system (11) and (12) is <u>completely observable</u> if the matrix $M(t_1, t_2)$ (cf eq. (13)) is positive definite.

Definition: The system (11) and (12) is completely controllable if the matrix $W(t_1, t_2)$ (cf eq. (14)) is positive definite.

Assuming that the system (9) and (10) is uniformly completely controllable and uniformly completely observable, Kalman and Bucy [8] and Kalman [13] have shown that with an arbitrary initial covariance for X_0 the conditional error covariance for the filtering problem (9) and (10) is bounded and converges uniformly and exponentially to a unique solution.

For the system (11) and (12), assuming complete controllability and complete observability, the conditional error covariance converges uniformly to a constant matrix which is the unique positive definite equilibrium state of

$$\frac{dP}{dt} = aP + Pa^T - Pf^T fP + bb^T \tag{15}$$

which is the optimal matrix mean square error for the Wiener-Kolmogorov solution to the filtering problem (11) and (12) given the infinite past $\{Y_u, -\infty < u \le t\}$. With these results it is easy to obtain the following:

Proposition: <u>Given that the system (9) and (10) is uniformly completely controllable and uniformly completely observable then the rate of generation of information about the process $\{X_t\}$ by the process $\{Y_t\}$ exists and is one half the trace of the steady-state covariance error for the filtering problem for (9) and (10).</u>

Corollary: <u>Given that the system (11) and (12) is completely controllable and completely observable then the rate of generation of information about the process $\{X_t\}$ by the process $\{Y_t\}$ exists and is one half the trace of the</u>

optimal matrix mean square error for the Wiener-Kolmogorov solution to the filtering problem (11) and (12).

Remark 1. We can also calculate mutual information and information rate when the noise processes $\{B_t\}$ and $\{\tilde{B}_t\}$ are correlated and with appropriate absolute continuity conditions we can make calculations for "smooth" noise processes.

Remark 2. From the convergence properties for the conditional error covariance, rates of convergence can be given for some problems that Gelfand and Yaglom [1] consider of information and information rate about a stationary process over a finite interval, contained in a sum of this process and white noise when the interval is allowed to become unbounded.

REFERENCES

1. I. M. Gelfand and A. M. Yaglom, "Calculation of the Amount of Information About a Random Function Contained in Another Such Function," Usp. Mat. Nauk 12 (1957) pp 3-52. English Translation in Amer. Math. Soc. Transl. Series 2, 12 (1959) pp.199-246

2. Chiang Tse-Pei, "Remark on the Definition of the Quantity of Information," Teor. Veroyatnost. i. Primenen 3 (1958) pp 99-103. English Translation in Amer. Math. Soc. Transl., Series 2, 12 (1959) pp.247-250.

3. A. Perez, "Notions generalisees d'incertitude, d'entropie et d'information du point de vue de la theorie de martingales," Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. Publishing House, Czech Acad. Sci., Prague (1957) pp.183-208.

4. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, Univ. of Illinois Press, Urbana 1949.

5. I. V. Girsanov, "On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measures," Theor. Probability Appl., V, (1960), pp.285-301.

6. T. E Duncan, "Evaluation of Likelihood Functions," to appear.

7. A.J.F. Siegert, "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena - Part II, Examples," IRE Trans. Information Theory IT-3 (1957) pp. 38-43. "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena - Part III, Examples," IRE Trans. Information Theory IT-4 (1958) pp. 4-14.

8. R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering Theory," J. Basic Engr. (ASME Trans), 83 D, (1961) pp. 95-108.

9. K. Itô and M. Nisio, "On Stationary Solutions of a Stochastic Differential Equation," J. Math. Kyoto Univ., 4, (1964-1965) pp. 1-75.

10. H. J. Kushner, "Dynamical Equations for Optimum Nonlinear Filtering," J. Differential Equations, 3, (1967) pp.171-190.

11. T. E. Duncan, "On the Nonlinear Filtering Problem," to appear.

12. M.S. Pinsker, Information and Information Stability of Random Variables and Processes, Holden-Day, Inc., 1964.

13. R.E. Kalman, "New Methods in Wiener Filtering Theory," Proc. of the First Symposium on Engineering Applications of Random Function Theory and Probability, eds. J.L. Bodganoff and F. Kozin, John Wiley and Sons, Inc., New York, 1963, pp. 270-388.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) University of Michigan Information and Control Engineering Ann Arbor, Michigan 48105	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

ON THE CALCULATION OF MUTUAL INFORMATION

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Tyrone E. Duncan

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1969	11	13

8a. CONTRACT OR GRANT NO. AF-AFOSR 814-66 b. PROJECT NO. 9749-01 c. 6144501F d. 681304	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR 69-1112TR

10. DISTRIBUTION STATEMENT

1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES TECH, OTHER	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research (SRMA) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

Calculating the amount of information about a random function contained in another random function has important uses in communication theory. An expression for the mutual information for continuous time random processes has been given by Gelfand and Yaglom, Chiang, and Perez by generalizing Shannon's result in a natural way. Under a condition of absolute continuity of measures the continuous time expression has the same form as Shannon's result. For two Gaussian processes Gelfand and Yaglom express the mutual information in terms of a mean square estimation error. We generalize this result to diffusion processes and express the solution in a different form which is more naturally related to a corresponding filtering problem. We also use these results to calculate some information rates.

DD FORM 1473