SIAM J. APPL. MATH.
Vol. 19, No. 1, July 1970

ON THE CALCULATION OF MUTUAL INFORMATION*

TYRONE E. DUNCAN†

1. Introduction. Calculating the amount of information about one random function contained in another random function has many applications in communication theory. For continuous time stochastic processes an expression for the mutual information has been obtained by Gel'fand and Yaglom [1], Chiang [2] and Pèrez [3] by generalizing Shannon's result [4] in a natural way. With a certain absolute continuity condition the expression for the mutual information of continuous parameter real-valued processes has the same form as Shannon's result. For some Gaussian processes Gel'fand and Yaglom [1] express the mutual information in terms of a mean square estimation error. We generalize their result to calculating the mutual information between one process and the sum of the first process and white noise. The expression for the mutual information is in a form different from that obtained by Gel'fand and Yaglom but more naturally related to a corresponding filtering problem. With the expression for the mutual information some information rates are calculated.

2. Problem statement. We shall consider two stochastic processes Y and Z as follows:

$$dY_t = Z_t\,dt + dB_t, \tag{1}$$

where the n-dimensional process Z is independent of the n-dimensional standard Brownian motion B, $t \in [0, 1]$, $Y_0 \equiv 0$ and

$$\iint Z_t^T Z_t\,dP\,dt < \infty \tag{2}$$

where the superscript T denotes transpose.

We wish to calculate the amount of information in the process Y about the process Z.

3. Preliminaries. Generalizations of Shannon's definition of mutual information have been obtained by Gel'fand and Yaglom [1], Chiang [2] and Pèrez [3]. They obtain the following result as the natural extension of Shannon's mutual information.

THEOREM 1. *Let ξ and η be two random vectors with joint probability measure $P_{\xi\eta}$ and marginal probability measures P_ξ and P_η respectively. Assume that $P_{\xi\eta} \ll P_\xi P_\eta$. Then the mutual information $J(\xi, \eta)$ between ξ and η is*

$$J(\xi, \eta) = \int \alpha(x, y) \log \alpha(x, y)\,dP_\xi(x)\,dP_\eta(y), \tag{3}$$

* Received by the editors May 21, 1968, and in revised form April 21, 1969.

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

where

$$\alpha(x, y) = \frac{dP_{\xi\eta}(x, y)}{dP_{\xi}(x)\, dP_{\eta}(y)}. \tag{4}$$

From Theorem 1 we see that an appropriate Radon–Nikodym derivative must be calculated to evaluate the mutual information. So before establishing our main result we shall prove an absolute continuity result that will be useful there. While the proof will appear elsewhere in a detection theory context [5] we shall include it here for completeness.

THEOREM 2. *Consider the processes B, Y and Z described in* (1) *and* (2). *Then* $\mu_{YZ} \ll \mu_B\mu_Z$ *and* $\mu_Y \ll \mu_B$, *where* μ_{YZ} *is the joint probability measure for the processes Y and Z and* μ_B, μ_Y *and* μ_Z *are the measures for the processes B, Y and Z respectively. The Radon–Nikodym derivatives are*

$$\frac{d\mu_{YZ}}{d\mu_B\, d\mu_Z} = \exp\left[\int Z_s^T\, dB_s - \frac{1}{2}\int Z_s^T Z_s\, ds\right], \tag{5}$$

$$\frac{d\mu_Y}{d\mu_B} = \exp\left[\int \hat{Z}_s^T\, dB_s - \frac{1}{2}\int \hat{Z}_s^T \hat{Z}_s\, ds\right], \tag{6}$$

where $\hat{Z}_s = E[Z_s | Y_u, 0 \leqq u \leqq s]$ *with Y given by* (1) *while in* (6) *it has the* μ_B *probability law.*

Proof. We shall initially assume that Z is a bounded uniformly stepwise process, i.e., there exists a finite subdivision of $[0, 1]$, $0 = t_0 < t_1 < \cdots < t_n = 1$ and a finite constant M such that

$$Z_t(\omega) = Z_{t_i}(\omega), \qquad t_i \leqq t < t_{i+1}, \quad i = 0, 1, \cdots, n-1, \tag{7}$$

and $|Z_t(\omega)| < M$. Considering each partition interval we can easily establish that $\mu_{YZ} \ll \mu_{BZ}$. The Radon–Nikodym derivative, ϕ, is

$$\phi_t = \exp\left[\int_0^t Z_s^T\, dY_s - \frac{1}{2}\int_0^t Z_s^T Z_s\, ds\right]. \tag{8}$$

We shall now show that $\mu_Y \ll \mu_B$. By the independence of the processes B and Z the measure μ_{BZ} is the product measure $\mu_B\mu_Z$. Thus, we merely integrate on the measure μ_Z. Define

$$\psi_t = E_{\mu_Z}\phi_t, \tag{9}$$

where E_{μ_Z} denotes integration with respect to the measure μ_Z. Therefore,

$$d\mu_Y/d\mu_B = \psi. \tag{10}$$

Applying the formula for stochastic differentials [6] to ϕ_t we have

$$\phi_t = 1 + \int_0^t \phi_s Z_s^T\, dB_s. \tag{11}$$

A simple verification shows that

$$\int \phi_s^2 Z_s^T Z_s\, ds < \infty \quad \text{a.s. } \mu_{BZ} \tag{12}$$

so that the stochastic integral in (11) can be defined as an $L^1(dP)$ limit of finite sum approximations to the integral. For the finite sums we have

$$\int E_{\mu_Z} \sum_{i=1}^{n} \phi_{t_i} Z_{t_i}^T (B_{t_{i+1}} - B_{t_i})\, d\mu_B = \int \sum [E_{\mu_Z} \phi_{t_i} Z_{t_i}^T](B_{t_{i+1}} - B_{t_i})\, d\mu_B. \tag{13}$$

Since the limit of the integrand on the R.H.S. of (13) is well-defined, we have

$$E_{\mu_Z} \int_0^t \phi_s Z_s^T\, dB_s = \int_0^t E_{\mu_Z} \phi_s Z_s^T\, dB_s \qquad \text{a.s. } \mu_B. \tag{14}$$

Therefore,

$$\psi_t = 1 + \int_0^t E_{\mu_Z} \phi_s Z_s^T\, dB_s. \tag{15}$$

Let $\Gamma_t = \ln \psi_t$ and apply the formula for stochastic differentials [6] (which can be easily verified to be valid here) to obtain

$$d\Gamma_t = \frac{E_{\mu_Z} \phi_t Z_t^T\, dB_t}{E_{\mu_Z} \phi_t} - \frac{1}{2} \frac{E_{\mu_Z} \phi_t Z_t^T E_{\mu_Z} \phi_t Z_t\, dt}{[E_{\mu_Z} \phi_t]^2}. \tag{16}$$

Consider the expression

$$E_{\mu_Z} \phi_t Z_t / (E_{\mu_Z} \phi_t). \tag{17}$$

Since $\phi = d\mu_{YZ}/d\mu_{BZ}$ the expression (17) is the conditional expectation $E[Z_t | Y_u, 0 \leqq u \leqq t]$, i.e., (17) has the proper measurability properties for $E[Z_t | Y_u, 0 \leqq u \leqq t]$ and it calculates the correct probabilities. Thus

$$\hat{Z}_t \triangleq E[Z_t | Y_u, 0 \leqq u \leqq t] = E_{\mu_Z} \phi_t Z_t / (E_{\mu_Z} \phi_t) \tag{18}$$

and

$$\psi_t = \exp\left[\int_0^t \hat{Z}_s^T\, dB_s - \frac{1}{2} \int_0^t \hat{Z}_s^T \hat{Z}_s\, ds\right].$$

For the case of a process Z satisfying (2) and independent of the process B, a sequence of bounded uniformly stepwise processes which converge to the process Z in $L^2(dt\, dP)$ can be obtained. By the Kolmogorov–Doob inequality [7] for the stochastic integral and the usual $L^1(dt)$ bound for the ordinary integral we have that

$$\phi^{(n)} \to \phi \quad \text{uniformly in } t \quad \text{a.s. } \mu_{BZ}. \tag{19}$$

All that remains to verify is that the absolute continuity has been preserved, in other words, that $\phi^{(n)} \to \phi$ in $L^1(d\mu_{BZ})$. A necessary and sufficient condition for $\phi^{(n)} \to \phi$ in $L^1(d\mu_{BZ})$ is that the sequence $\{\phi^{(n)}\}$ be uniformly integrable [8]. Since the process Z satisfies (2) we have that

$$\sup_n \int \phi^{(n)} \ln \phi^{(n)}\, d\mu_{BZ} < \infty \tag{20}$$

which implies uniform integrability of the sequence $\{\phi^{(n)}\}$ (see [8]). Arguments similar to those for a bounded uniformly stepwise process Z show that ψ in (10) is given by (17).

4. Main result. Having sufficient preliminaries established we shall now characterize the mutual information between the processes Y and Z.

THEOREM 3. *Consider the processes Y and Z given in* (1) *and* (2). *The mutual information, $J(Y,Z)$, contained in $\{Y_u, 0 \leqq u \leqq 1\}$ about $\{Z_u, 0 \leqq u \leqq 1\}$ is given by the following expression*:

$$J(Y,Z) = \frac{1}{2} E \int_0^1 [Z_u - \hat{Z}_u]^T [Z_u - \hat{Z}_u] \, du, \tag{21}$$

where $\hat{Z}_u = E[Z_u | Y_s, 0 \leqq s \leqq u]$.

Proof. To calculate the mutual information between Y and Z using Theorem 1 we must show that $\mu_{YZ} \ll \mu_Y \mu_Z$ and compute the Radon–Nikodym derivative. Let

$$\Phi = \frac{d\mu_{YZ}}{d\mu_Y \, d\mu_Z}. \tag{22}$$

By Theorem 2, $\mu_{YZ} \ll \mu_B \mu_Z$, and by using the entropy property (20) with μ_B and μ_Y reversed we have that $\mu_B \ll \mu_Y$. Using the chain rule for Radon–Nikodym derivatives we have that $\mu_{YZ} \ll \mu_Y \mu_Z$ and $\Phi = \phi \psi^{-1}$. The mutual information is

$$J(Y,Z) = \int \Phi \log \Phi \, d\mu_Y \, d\mu_Z \tag{23}$$

and

$$\log \Phi = \int (Z_s - \hat{Z}_s)^T \, dY_s - \frac{1}{2} \int (Z_s - \hat{Z}_s)^T (Z_s - \hat{Z}_s) \, ds. \tag{24}$$

Substituting $dY_t = Z_t \, dt + dB_t$ and using the fact that the stochastic integral is a martingale from (2) we have

$$J(Y,Z) = \frac{1}{2} E \int (Z_s - \hat{Z}_s)^T (Z_s - \hat{Z}_s) \, ds. \tag{25}$$

Remark 1. Gel'fand and Yaglom [1] obtain an expression for the mutual information of Gaussian processes and for some Gaussian processes express the mutual information in terms of a filtering error. While their filtering error expression is in a different form from (21) the equivalence of the results can be obtained from some results on Fredholm integral equations. The mean square error expression (21) for mutual information for Gaussian processes in white noise is known [9, p. 585]. Some recent work on mutual information for Gaussian processes has been done by Baker [10].

Remark 2. With additional assumptions on the structure of the process Z the assumption of independence of the processes B and Z can be removed and the mutual information can be expressed in a form similar to (21).

5. An application to information rate. When the process Z is a Gauss–Markov process some information rates can be obtained. These results extend and simplify some results of Gel'fand 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.

To clarify terms we give the definition that we shall use for information rate (Gel'fand and Yaglom [1], Pinsker [11]).

DEFINITION. The *rate of generation of information* about a process η by a process ξ is

$$\bar{I}(\xi, \eta) = \lim_{T \to \infty} \frac{1}{T} J(\xi_u, \eta_u; 0 \leqq u \leqq T), \tag{26}$$

where $J(\xi_u, \eta_u; 0 \leqq u \leqq T)$ is the mutual information between the processes ξ and η on the interval $[0, T]$ and $\bar{I}$ is defined only when the limit exists.

We shall obtain a result for the existence of an information rate in terms of some system theory results. The methods used to obtain the existence of the mutual information will give some useful bounds on finite-time approximations to information rate.

We shall calculate the rate of generation of information about a Gaussian process X by another Gaussian process Y described by the following stochastic differential equations:

$$dX_t = a(t)X_t\,dt + b(t)\,d\tilde{B}_t, \tag{27}$$

$$dY_t = cX_t\,dt + dB_t, \tag{28}$$

where the processes B and $\tilde{B}$ are independent n- and m-dimensional standard Brownian motions respectively, the matrices a and b with suitable dimensions have nonrandom elements which are continuous functions of t, the matrix c with suitable dimensions has elements which are constants, the interval of solution is the positive half-line $[0, \infty)$ and the initial conditions are $X_0 = \alpha$, a zero mean Gaussian random vector independent of the processes B and $\tilde{B}$ and $Y_0 \equiv 0$.

We shall also consider the case where the coefficients of the stochastic differential equation (27) are not functions of time. In this case, we shall use the same symbols for the coefficients deleting the variable t, i.e.,

$$dX_t = aX_t\,dt + b\,d\tilde{B}_t, \tag{29}$$

$$dY_t = cX_t\,dt + dB_t, \tag{30}$$

where the appropriate assumptions for (27) and (28) are still in effect.

If the process Y is a process of observations from which the minimum mean square error estimate of process X is sought then we have a well-known filtering problem. By Theorem 3 the mutual information for (27) and (28) (or (29) and (30)) is obtained from the integral of the trace of the optimal error covariance matrix for estimating the process X from the process Y. Information rates for (27) and (28) and (29) and (30) will be obtained by showing that the error covariance matrix for the associated filtering problem converges to a steady state solution.

Assuming that the system (27) and (28) is uniformly completely controllable and uniformly completely observable, Kalman and Bucy [12] and Kalman [13] have shown that for an arbitrary covariance for $\alpha = X_0$ the error covariance for the filtering problem (27) and (28) is bounded and converges uniformly and exponentially to a unique matrix.

By assuming complete controllability and complete observability for the system (29) and (30) the error covariance converges uniformly to a constant matrix [12], [13] which is the unique positive definite equilibrium state of the Riccati equation

$$dP/dt = aP + Pa^T + Pc^TcP + bb^T. \tag{31}$$

This constant matrix is the error covariance for the Wiener–Kolmogorov solution to the filtering problem (29) and (30) given the infinite past $\{Y_u, -\infty < u \leqq t\}$.

Applying the above results to the calculation of mutual information for (27) and (28) and (29) and (30), we can easily obtain the following result.

PROPOSITION. *Suppose that the system* (27) *and* (28) *is uniformly completely controllable and uniformly completely observable. Then the rate of generation of information about the process X by the process Y exists and is one-half the trace of $c\bar{P}c^T$, where $\bar{P}$ is the steady-state error covariance for the filtering problem for* (27) *and* (28).

COROLLARY. *Suppose that the system* (29) *and* (30) *is completely controllable and completely observable. Then the rate of generation of information about the process X by the process Y exists and is one-half the trace of $c\bar{P}c^T$, where $\bar{P}$ is the error covariance matrix for the Wiener–Kolmogorov solution to the filtering problem* (29) *and* (30).

Remark 3. Mutual information and information rate can be calculated when the processes B and $\tilde{B}$ are correlated, and with appropriate absolute continuity conditions calculations can be made when the process B is replaced by a "smooth" process.

REFERENCES

[1] I. M. GEL'FAND AND A. M. YAGLOM, *Calculation of the amount of information about a random function contained in another such function*, Uspekhi Mat. Nauk, 12 (1957), pp. 3–52; English transl., Amer. Math. Soc. Transl. (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 transl., Amer. Math. Soc. Transl. (2), 12 (1959), pp. 247–250.

[3] A. PÈREZ, *Notions generalisées d'incertitude, d'entropie et d'information du point de vue de la theorie de martingales*, Trans. 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*, University of Illinois Press, Urbana, 1949.

[5] T. E. DUNCAN, *Likelihood functions for stochastic signals in white noise*, Information and Control, to appear.

[6] K. ITÔ, *On a formula concerning stochastic differentials*, Nagoya Math. J., 3 (1951), pp. 55–65.

[7] J. L. DOOB, *Stochastic Processes*, John Wiley, New York, 1953.

[8] P. A. MEYER, *Probability and Potentials*, Blaisdell, Waltham, Mass., 1966.

[9] H. L. VAN TREES, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley, New York, 1968.

[10] C. R. BAKER, *Mutual information for Gaussian processes*, to appear.

[11] M. S. PINSKER, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.

[12] R. E. KALMAN AND R. S. BUCY, *New results in linear filtering theory*, Trans. ASME Ser. D. J. Basic Engrg., 83 (1961), pp. 95–108.

[13] R. E. KALMAN, *New methods in Wiener filtering theory*, Proc. First Symposium on Engineering Applications of Random Function Theory and Probability, J. L. Bodganoff and F. Kozin, eds., John Wiley, New York, 1963, pp. 270–388.