The Annals of Mathematical Statistics
1971, Vol. 42, No. 3, 1054-1067

THE STRUCTURE OF RADON-NIKODYM DERIVATIVES WITH RESPECT TO WIENER AND RELATED MEASURES[1]

BY THOMAS KAILATH

Stanford University and Indian Institute of Science

The Radon-Nikodym derivative (RND) with respect to Wiener measure of a measure determined by the sum of a differentiable (random) signal process and a Wiener process is shown, under rather general conditions, to have the same form as the RND for the case of a known (nonrandom) signal plus a Wiener process. The role of the known signal is played by the causal least-squares estimate of the signal process given the sum process. This formula can be shown to be equivalent to all previously known explicit formulas for RND's relative to Wiener measure. Moreover, and more important, the formula suggests a general structure for engineering approximation and implementation of signal detection schemes.

Secondly, an explicit necessary and sufficient characterization, in signal plus noise form, is given of all processes absolutely continuous with respect to a Wiener process. Finally, the results are extended to some reference measures related to Wiener measure, in particular to measures induced by martingales of a Wiener process. We also note that the case where both measures are Gaussian permits some stronger results.

The proofs are based on several recent results in martingale theory.

1. Introduction. The problem of determining the Radon-Nikodym derivative of the absolutely continuous part of one measure with respect to another has, especially in recent years, been of considerable interest—we mention here only the survey papers of Yaglom [20] and Skorokhod and Gikhman [6]. The Radon-Nikodym derivative can be used as a tool to calculate probabilities and to evaluate function space integrals (see, e.g., Shepp [26]), but it also has major applications in the statistical theories of estimation and signal detection (hypothesis testing) and in information theory.

The basic problem in these latter applications is to decide, given a sample function, whether it comes from one stochastic process or another. Regarding the stochastic processes as measures, say P_1 and P_0, in a function space, it can be shown that an 'optimal' solution, under a wide variety of error criteria, is obtained by comparing the Radon-Nikodym derivative, dP_1/dP_0, with a preset threshold that depends upon the costs and prior probabilities.

Our aim in this paper will be to present some results that elucidate the structure of the Radon-Nikodym derivative (RND). We shall also discuss the considerable engineering significance of the results. To introduce such practical considerations, we first note that "structural" information on the RND is often more valuable in applications than the explicit analytical or numerical evaluation of the Radon-

[1] This work was partially supported by the Applied Mathematics Division of the AFOSR under contract AF 44-620-69-C-0101 and by a Fellowship of the John Simon Guggenheim Memorial Foundation.

Nikodym derivative (RND), or likelihood ratio (LR), as it is often known in statistical applications. This aspect has been illustrated in some detail in the engineering papers [13]–[15], where we may remark that stronger conditions were imposed than in the present paper. However, to provide some perspective for the present paper, we shall summarize some of those earlier discussions. Briefly, we first observe that the LR can always, roughly speaking, be written down as a ratio of probability densities, which can with sufficient effort be numerically evaluated in any problem; this would seem to be the end of the matter! However, in many instances, such brute-force evaluation misses the point. For one thing, all mathematical models are only idealizations, to varying degrees, of the real physical problems. Furthermore, even if the models are good, our knowledge of the parameters in them (e.g., the covariance functions of the processes) may not be good enough to justify a direct numerical evaluation of formulas derived from the mathematical models. Finally, even if the evaluation can be performed, it may involve an undue amount of work and simpler, "suboptimum," schemes must be sought. Therefore, the chief aim of most mathematical analyses of engineering problems is to gain insight into the structure of the solution, insight that can then be used to intelligently modify and adapt the mathematical solution to a particular physical problem. We shall try to present below such an examination and interpretation of our major mathematical results.

Mathematical assumptions. For the presentation of the mathematical results, we start with a probability space $(\Omega, \mathscr{B}, P)$ and three families of random variables $w(t, \omega)$, $z(t, \omega)$, $x(t, \omega)$, $\omega \in \Omega$, $t \in [0, T)$, $T < \infty$. We assume that

(i) the $\{w(t, \omega)\}$ constitute a Wiener process with

$$E[w(t)] = 0, E[w(t)w(s)] = \min(t, s) = t \wedge s, \tag{1}$$

(ii) the $\{z(t, \omega)\}$ form a, not necessarily Gaussian, process obeying

$$E \int_0^T |z(t, \omega)| \, dt < \infty \tag{2}$$

and

$$\int_0^T z^2(t, \omega) \, dt < \infty \text{ a.s.}, \tag{3}$$

(iii) "future" increments $\{w(t, \omega) - w(s, \omega)\}$, are independent of the Borel fields

$$\mathscr{B}_s = \sigma\{w(\tau, \omega), z(\tau, \omega), \tau \leqq s\}$$

generated by "past" $z(\cdot)$ and $w(\cdot)$, and augmented by all the null sets. Following K. Itô, such independence conditions will often be compactly written

$$w(t, \omega) - w(s, \omega) \perp\!\!\!\perp \mathscr{B}_s, \qquad 0 \leqq s < t \leqq T. \tag{4}$$

Next, let

$$\mathscr{F}_t = \sigma\{x(\tau, \omega), \tau \leqq t\}. \tag{5}$$

We shall assume that the $\{\mathscr{F}_t\}$ are each augmented by all null sets. Then on $(\Omega, \mathscr{F}_T, P)$ we shall define two measures P_1 and P_0 by the relations

(6) $$P_0\{\omega: x(t_1, \omega) \leqq x_1, \cdots, x(t_n, \omega) \leqq x_n\} = P\{\omega: w(t_1, \omega) \leqq x_1, \cdots, w(t_n, \omega) \leqq x_n\}$$

and

(7) $$P_1\{\omega: x(t_1, \omega) \leqq x_1, \cdots x,(t_n, \omega) \leqq x_n\} = P\{\omega: \textstyle\int_0^{t_1} z(s, \omega)\, ds + w(t_1) \leqq x_1, \cdots, \int_0^{t_n} z(s, \omega)\, ds + w(t_n) \leqq x_n\}$$

for all finite collections $\{t_i\}$ in $[0, T]$ and for arbitrary real numbers $\{x_i\}$.

The problem of deciding whether an observed sample function of the process $x(\cdot)$ is governed by measure P_1 or P_0 is the statistical hypothesis testing problem or, in the language of communication engineering, the signal detection problem, where the random process $z(\cdot)$ is the signal. Such signal detection problems arise, for example, in radio-astronomy, molecular spectroscopy, and in multipath communication channels subject to fading and scatter propagation.

As noted earlier, the basic procedure in our discrimination problem is to compute the RND (or likelihood ratio (LR)). Our most important result is the following.

THEOREM 2. *Under the assumptions* (i) *to* (iii), *the measure* P_1 *is absolutely continuous with respect to the measure* $P_0 (P_1 \ll P_0)$, *and the Radon-Nikodym derivative can be expressed as*

(8a) $$\frac{dP_1}{dP_0} = \exp\left\{\int_0^T \hat{z}_1(t)\, dx\,(t) - \frac{1}{2}\int_0^T \hat{z}_1{}^2(t)\, dt\right\} \quad on\ A$$

(8b) $$\frac{dP_1}{dP_0} = 0 \quad on\ \bar{A}$$

where

(9) $$\hat{z}_1(t) = E_{P_1}[z(t) \mid \mathscr{F}_t]$$

and

(10) $$A = \left\{\omega: \frac{dP_1}{dP_0} > 0\right\}.$$

We note that $P_1(A) = 1$. The stochastic integral is taken in the Itô sense (cf., Doob [3], Chapter 9).

REMARK 1. If $P_1 \equiv P_0$, i.e., $P_1 \ll P_0$ and $P_0 \ll P_1$, then the set A will have measure one under both P_1 and P_0. In that case (8b) and (10) can be ignored. Sufficient conditions for $P_1 \equiv P_0$ are that (cf., [12], [16]) $z(\cdot)$ be a.s. uniformly bounded in amplitude or energy or that $z(\cdot)$ be completely independent of $w(\cdot)$ and of a.s. finite energy, i.e., that $z(\cdot)$ be a.s. square integrable.

REMARK 2. The formula (8) is very general and all known explicit RND formulas relative to Wiener measure can, with the added help of Theorem 1 below, be shown to be equivalent to it. Several examples are worked out in references [13]–[15], so we shall not add any here.

The special case of (8) where $z(\cdot)$ is deterministic, so that

$$z(\cdot) = \text{its mean value, say } \quad m(\cdot) \tag{11}$$

and

$$\frac{dP_1}{dP_0} = \exp\left\{\int_0^T m(t)\,dx(t) - \frac{1}{2}\int_0^T m^2(t)\,dt\right\} \tag{12}$$

is well known and was perhaps first studied by Cameron and Martin and in greater generality by Maruyama (see the references in Cameron and Graves [1]). The engineering meaning and practical implementation of devices, known as matched filters and correlation detectors, computing (12) have been studied in considerable detail in the engineering literature (see, e.g., [9], [28] and [29]). It turns out that there is an intuitively pleasing relationship between (8) and (12) that derives from the well-known fact that the conditional mean $\hat{z}_1(\cdot)$ of (8) can be described as

(13) $\hat{z}_1(t)$ = the minimum mean-square error estimate of the signal process $z(t)$, given past observations $\{x(s), 0 \leqq s \leqq t\}$ and assuming that measure P_1 is operative.

This fact leads to the following nice interpretation of (8). When the signal $z(\cdot)$ is known (deterministic), we have the formula (12); when $z(\cdot)$ is unknown (random), we obviously cannot use (12) directly, but we should first estimate the signal and then behave as if the estimate were perfect, i.e., as if the signal is now completely known!

But what is the engineering significance of the formula (8) and of the interpretation we have just described? In the first place, when the estimate $\hat{z}_1(\cdot)$ can be directly computed, the interpretation enables us to extend to the general random-signal problem a lot of the engineering insight and experience that has been gained for the deterministic-signal problem. Other ways of describing the LR will, to varying degrees, deprive us of this benefit. Secondly, a lot of effort has gone into the problem of determining $\hat{z}_1(\cdot)$, especially in control theory where such calculations have become important for trajectory determination. But despite this, in many problems, $\hat{z}_1(\cdot)$ is difficult to obtain, either because the formula for it is very complicated and/or because we may have too gross a knowledge of the parameters of our mathematical model to justify too great an effort on the exact calculation of $\hat{z}_1(\cdot)$. But we feel that the very general result (8) suggests a "reasonable" procedure in many such instances, viz., that we try to use the "best available" estimate. Thus we might use a suboptimum, but simpler estimate; or we may fix the complexity of the estimator to allow say three variable parameters, whose values can then be optimized to give the best (constrained) estimate or whose values could even be set by Monte-Carlo trials, or by field experiments. We should point out that such a philosophy has already been successfully demonstrated in some engineering solutions (cf., [2], [24], [25]), though not with as much theoretical backing as for the structure we have proposed here.

After this long discussion of Theorem 2, we turn to Theorem 1, which plays an important part in the proof of Theorem 2, and also in certain extensions of it. Moreover, Theorem 1 also shows that the type of "signal" plus "noise" structure that we used to define the measure P_1 in Theorem 2 is, in a sense, quite general.

THEOREM 1. *A stochastic process* $\{x(t, \omega), \mathscr{F}_t, Q\}$ *will be absolutely continuous with respect to Wiener process* $\{x(t, \omega), \mathscr{F}_t, P_0\}$ *if and only if we can write* $x(\cdot)$ *in the form*

$$x(t, \omega) = \int_0^t \phi(s, \omega)\, ds + \mu(t, \omega), \qquad 0 \leqq t \leqq T, \text{ a.s. } Q \tag{14}$$

where $\{\mu(t, \omega), \mathscr{F}_t, Q\}$ *is a Wiener process and* $\phi(\cdot, \cdot)$ *is a unique* [a.e. (t, ω)] *function that is* (i) (t, ω) *measurable,* (ii) $\mathscr{F}_t$*-measurable for each* $t \in [0, T]$, *and* (iii) *such that* $\int_0^T \phi^2(t, \omega)\, dt < \infty$ a.s. Q. *Moreover, when* $Q \ll P_0$ *(i.e.* Q *is absolutely continuous with respect to* P_0), *we can write* RND *as*

$$\frac{dQ}{dP_0} = \exp\left\{\int_0^T \phi(t, \omega)\, dx(t, \omega) - \frac{1}{2}\int_0^T \phi^2(t, \omega)\, dt\right\}, \qquad \text{on } A \tag{15a}$$

$$\frac{dQ}{dP_0} = 0 \qquad \text{on } \bar{A} \tag{15b}$$

where

$$A = \left\{\omega : \frac{dQ}{dP_0} > 0\right\}. \tag{15c}$$

This theorem is more general than Theorem 2 because it holds for any process absolutely continuous with respect to (w.r.t.) a Wiener process, not just for signal plus noise processes of the form $\{x(t, \omega), \mathscr{F}_t, P_1\}$ used in Theorem 2. However, from an engineering point of view, the value of Theorem 1 is limited because, unlike the function $\hat{z}_1(\cdot)$ of Theorem 2, we have no physical idea what the function $\phi(\cdot)$ represents and therefore we have no guidance in approximating $\phi(\cdot)$ in a physical problem. This is why we have given more importance to Theorem 2.

However, the greater generality of Theorem 1 enables us to show that our assumptions in the more special Theorem 2, though quite weak, are not the best possible. Thus, we assumed that the signal process $z(\cdot)$ obeyed the condition

$$E\int_0^T |z(t, \omega|\, dt < \infty \tag{16}$$

which implies in particular that

$$E|z(t, \omega)| < \infty \text{ a.e. } t. \tag{17}$$

However even the constraint (17) on $z(\cdot)$ is not necessary, despite the fact that such a constraint is usually imposed when conditional expectations, like our $\hat{z}_1(\cdot)$, are to be considered. For an example, suggested in a conversation by L. A. Shepp of the Bell Telephone Laboratories, let us take

(18) $z(\cdot) = \eta$, a Cauchy distributed random variable, independent of $w(t)$, $0 \leqq t \leqq T$.

In this case, though

$$E|z(t)| = E|\eta| = \infty, \qquad 0 \leqq t \leqq T \tag{19}$$

the conditional expectation $\hat{z}_1(\cdot)$ can still be defined; in fact, it is easily verified that

$$\hat{z}_1(t) = \frac{\int_0^t \eta A(t,\eta)p(\eta)\,d\eta}{\int_0^t A(t,\eta)p(\eta)\,d\eta}, \qquad A(t,\eta) = \exp\left[\eta w(t) - \eta^2 \frac{t}{2}\right]. \tag{20}$$

It is not hard to see that this $\hat{z}_1(\cdot)$ satisfies the requirements on the function $\phi(\cdot)$ of Theorem 1 and therefore that the RND will still be given by (8) of Theorem 2, even though one of the hypotheses of that theorem is not met.

Similarly with the help of Theorem 1, we can show that the hypothesis in Theorem 2 that, for all $0 \leqq s \leqq t \leqq T$, $\{w(t) - w(s)\}$ be independent of $\mathscr{F}_t = \sigma\{x(\tau), \tau \leqq s\}$ is not necessary. For let us take

(21) $z(\cdot) = w(T)$, the value of the Wiener process at $t = T$.

In this case, if we take

$$\hat{z}_1(t) = \frac{3}{T+3t}\,x(t), \qquad 0 \leqq t \leqq T \tag{22}$$

and it is not hard to verify that $\hat{z}_1(\cdot)$ satisfies the conditions on $\phi(\cdot)$ of Theorem 1. Therefore again, though a hypothesis of the theorem is violated, the RND continues to be given by (8) of Theorem 2.

We conjecture that the minimum requirements for Theorem 2 to be valid are that $z(\cdot)$ be such that

(23) $P_1 \ll P_0$ and $\hat{z}_1(\cdot)$ exists and is square-integrable a.s. P_1.

However, as yet no simpler conditions on $z(\cdot)$ than (2)–(4) have been found that will ensure that the conditions (23) are met. Actually, it seems likely that it suffices for P_1 to be absolutely continuous with respect to P_0. In our proof the only role of (3)–(4) is to ensure such absolute continuity. The other difficulty, at least with our present line of proof, arises in a certain key lemma, Lemma 4. At present, we need the assumption (2) for a proof of this lemma, though again the example already presented in (18)–(20) shows that this assumption is not necessary. The lemma in question is:

LEMMA 4. *Let* $dx(t,\omega) = z(t,\omega)\,dt + dw(t,\omega)$, *where* $w(\cdot)$ *and* $z(\cdot)$ *satisfy conditions* (1), (2) *and* (4). *Then the process* $\{v(t,\omega), \mathscr{F}_t, P_1\}$, *where*

$$dv(t) = dx(t) - \hat{z}_1(t)\,dt, \qquad i.e., \quad v(t) = x(t) - \int_0^t \hat{z}_1(s)\,ds \tag{24}$$

is a Wiener process with variance t.

We have called [13] the process $v(\cdot)$ the *innovations* process of $x(\cdot)$ because $dv(t)$ may be regarded, keeping in mind the interpretation (13) of $\hat{z}_1(\cdot)$ as an estimate of $z(\cdot)$, as the part of $dx(t)$ that cannot be estimated from the past $\{x(s), 0 \leqq s \leqq t\}$; in other words, $dv(t)$ may be regarded as the "new information" or the "innovation" in $x(\cdot)$ at time t. The term *innovations process* is due to Wiener and Masani [20]. The result of Lemma 4 under the stronger assumptions (1), (4) and the integrability of $E|z(\cdot)|^2$ was obtained independently by Frost [5], Shiryaev [27], Kallianpur [personal communication] and the author [13]. The role of Lemma 4 is in helping to identify $\phi(\cdot)$ of Theorem 1 with $\hat{z}_1(\cdot)$, though it takes some more work to establish this. However, heuristically we can briefly argue as follows: by Lemma 4 we can write

$$x(t,\omega) = \int_0^t \hat{z}_1(s,\omega)\,ds + v(t,\omega) \quad \text{a.s.} \quad P_1.$$

In this representation, $v(\cdot)$ is Wiener and $\hat{z}_1(\cdot)$, though random, is conditionally known given $x(\cdot)$. If $\hat{z}_1(\cdot)$ were completely known, the RND would have the Cameron–Martin–Maruyama form (12); it seems plausible that this form should continue to hold so long as $\hat{z}_1(\cdot)$ is at least conditionally known. Theorem 2 establishes this fact rigorously.

We should note here that Lemma 4 can also be shown to be valid [18] when the process $w(\cdot)$ is replaced by a square-integrable martingale of Brownian motion. Theorems 1 and 2 have certain corresponding generalizations which we shall examine in Section 3. Finally, in Section 4 we shall note that slightly stronger results can be obtained when all the processes are Gaussian. In this connection we note that Duncan [4], following Itô and Watanabe [11], has stated a more general version of formula (15) in which P_0 is not necessarily Wiener measure. Under the assumptions that the fields $\{\mathscr{F}_t\}$ are continuous and that the only martingales on $\{\Omega, \mathscr{F}_t, P_0\}$ are continuous, he states that (15) is true with the stochastic integral replaced by a continuous martingale and the second integral replaced by twice the natural increasing process associated with this martingale.

2. Proofs of Theorems 1 and 2.

The proofs will be developed in a series of lemmas. The reader will note that a surprising variety of martingale results is used in the proof. The basic notations have been presented in Section 1, especially (1)–(6), and it may be helpful to review them now.

LEMMA 1. *A sufficient condition for absolute continuity. The process* $\{x(t,\omega), \mathscr{F}_t, P_1\}$ *will be absolutely continuous with respect to the Wiener process* $\{x(t,\omega), \mathscr{F}_t, P_0\}$ *if*

$$\int_0^T z^2(t,\omega)\,dt < \infty \text{ a.s.} \tag{25}$$

PROOF. This result has been proved by Kadota and Shepp [12] and by Kailath and Zakai [16]. The proof in [16] is based on some results of Girsanov [7] combined with certain random time-change (McKean [21] page 29) and random stopping-time arguments.

LEMMA 2. *If a stochastic process* $\{x(t, \omega), \mathscr{F}_t, Q\}$ *is absolutely continuous with respect to a Wiener process* $\{x(t, \omega), \mathscr{F}_t, P_0\}$, *which we may denote as* $Q \ll P_0$, *then there exists a unique* (a.e. (t, ω)) *function* $\phi(t, \omega)$ *that is* (t, ω)*-measurable and* $\mathscr{F}_t$*-measurable for each* t, *obeying*

$$\int_0^T \phi^2(t, \omega)\, dt < \infty \text{ a.s. } Q \tag{26}$$

and such that the RND *can be written*

$$\frac{dQ}{dP_0} = L_T = \exp\left\{\int_0^T \phi(t, \omega)\, dx\,(t, \omega) - \frac{1}{2}\int_0^T \phi^2(t, \omega)\, dt\right\}, \qquad \text{on } A \tag{27a}$$

$$\frac{dQ}{dP_0} = L_T = 0 \qquad \text{on } \bar{A} \tag{27b}$$

where

$$A = \{\omega : L_T(\omega) > 0\}. \tag{27c}$$

PROOF. Our proof will follow one given by Hitsuda [10] and Kunita and Watanabe ([19] Section 6) for the case where Q and P_0 are equivalent, i.e., $Q \ll P_0$ and $P_0 \ll Q$. The case where only $Q \ll P_0$ will require some preliminary refinements.

To begin, let

$$L_t(\omega) = E_{P_0}[L_T(\omega) \mid \mathscr{F}_t]. \tag{28}$$

If Q and P_0 are equivalent, L_T and L_t will be strictly positive a.e. Q and P_0, a fact that is essential in the proofs of [10] and [19]. When only $Q \ll P_0$, there may be sets whose P_0-measure is nonzero, but whose Q-measure is zero and on such sets the RND, L_T, must be zero a.s. P_0. With this in mind, let us define a stopping time.

$$\tau(\omega) = \inf\{t \geqq 0 : L_t(\omega) = 0\}. \tag{29}$$

We note that since L_t is a nonnegative supermartingale, by a theorem of Meyer ([22] page 99) we will have

$$L_t(\omega) = 0, \qquad t \geqq \tau(\omega).$$

We shall prove that L_t has the form

$$L_t(\omega) = \exp\{\textstyle\int_0^t \phi(s, \omega)\, dx\,(s, \omega) - \frac{1}{2}\int_0^t \phi^2(s, \omega)\, ds\}, \qquad t < \tau(\omega). \tag{30}$$

By using the just-quoted theorem of Meyer, (27a) will follow immediately.

To prove (30), for $t < \tau(\omega)$, we can now proceed as in Hitsuda [10]. Briefly, by first proving the a.s. continuity of $L(t, \omega)$, it is shown that $L(t, \omega)$ is a local martingale of Brownian motion, in the sense of Kunita–Watanabe [19]. Then by using a result in ([19] page 312), it noted that L_t can be written as a stochastic integral,

$$L(t, \omega) = 1 + \int_0^t g(s, \omega)\, dx\,(s, \omega) \tag{31}$$

where $g(\cdot, \cdot)$ satisfies the conditions described for $\phi(\cdot, \cdot)$ in the statement of this lemma. We may note that the Kunita–Watanabe result only shows that $g(\cdot, \cdot)$ is square-integrable a.s. P_0. However, since $Q \ll P_0$, we can make the same assertion for $g(\cdot, \cdot)$ a.s. Q. The uniqueness of $g(\cdot, \cdot)$ also needs a separate proof, but this is easy. One proof can be given as in ([15] Lemma 4), by using Meyer's theorem on the unique Doob decomposition of certain continuous martingales. For variety, we shall give here an alternative proof.

If $L(t, \omega) = 1 + \int_0^t g_i(s, \omega)\, dx\,(s, \omega)$, $i = 1, 2$ let $h(t, \omega) = g_1(t, \omega) - g_2(t, \omega)$ and $J(t, \omega) = \int_0^t h(s, \omega)\, dx\,(s, \omega)$.

By hypothesis, $J(t, \omega) = 0$ a.e. (t, ω), and we have to show the same for $h(\cdot, \cdot)$. By the Itô differential rule (see, e.g., [19], Theorem 2.2),

$$J^2(t, \omega) = \int_0^t 2J(s, \omega)h(s, \omega)\, dx\,(s, \omega) + \int_0^t h^2(s, \omega)\, ds$$

which leads to

$$\int_0^t h^2(s, \omega)\, ds = 0 \text{ a.e. } (t, \omega)$$

and thence to

$$h(t, \omega) = 0 \text{ a.e. } (t, \omega).$$

Returning to the main proof, we note that by applying the Itô differential rule to (30), $L(t, \omega)$ can be rewritten as

$$L(t, \omega) = \exp\left\{\int_0^t \frac{g(s, \omega)}{L(s, \omega)}\, dx\,(s, \omega) - \frac{1}{2}\int_0^t \frac{g^2(s, \omega)}{L^2(s, \omega)}\, dx\right\}, \quad t < \tau(\omega).$$

But we can now obtain (30) by letting $\phi(t, \omega) = g(t, \omega)/L(t, \omega)$, $t < \tau(\omega)$.

LEMMA 3. *Let $\phi(t, \omega)$ be the function in Lemma 2 and let*

$$\mu(t, \omega) = x(t, \omega) - \int_0^t \phi(s, \omega)\, ds. \tag{32}$$

Then, $\{\mu(t, \omega), \mathscr{F}_t, Q\}$ is a Wiener process.

PROOF. This follows by a direct application of the main theorem of Girsanov [7].

PROOF OF THEOREM 1. The statement is given in Section 1. The sufficiency part of the representation (14) follows from Lemma 1; the necessity from Lemma 3. The RND formula follows from Lemma 2.

LEMMA 4. *The innovations process. Let $x(\cdot)$, $z(\cdot)$, P_1, and $\hat{z}_1(\cdot)$ be as in* (1), (2), (4), (7), *and* (9), *and let*

$$\nu(t, \omega) = x(t, \omega) - \int_0^t \hat{z}_1(s, \omega)\, ds.$$

Then $\{v(t, \omega), \mathscr{F}_t, P_1\}$ *is a Wiener process, which we have called the innovations process of* $\{x(t, \omega), \mathscr{F}_t, P_1\}$.

PROOF. We have to verify that $v(\cdot)$ is a continuous locally square-integrable martingale with quadratic variation equal to t, for then $v(\cdot)$ is Wiener by a famous theorem of Lévy and Doob ([3] page 384), as extended by Kunita and Watanabe ([19] page 217). The details of the calculation, and some extensions, are given in [18]; a proof under stronger conditions can also be found in ([4] Appendix I), [16], [17].

PROOF OF THEOREM 2. The statement is given in Section 1. The result follows by setting $Q = P_1$ in Theorem 1 and then using Lemmas 3 and 4, and the uniqueness of $\phi(\cdot, \cdot)$, to identify $\phi(\cdot, \cdot)$ as $\hat{z}_1(\cdot, \cdot)$ (see also the proof in [15], Lemma 4).

3. Other reference measures. In Theorem 2, the process $x(\cdot)$ was required to be a Wiener process under the reference measure P_0. In this section, we shall show how to relax this restriction. In the first place, instead of a Wiener process we can, of course, have any process equivalent (mutually absolutely continuous with respect) to a Wiener process—just use the chain rule for Radon-Nikodym derivatives. More interesting generalizations can be obtained by replacing the Wiener process by an a.s. continuous square-integrable martingale, say $M(\cdot)$, of a Wiener process or, more generally, by what is sometimes called (see, e.g., [7]) an Itô process, viz., the sum of an a.s. differentiable process and a martingale of the above type.

To begin, we recall that by a result of Kunita and Watanabe ([19] page 227), we can write $M(\cdot)$ as a stochastic integral

$$M(t, \omega) = \int_0^t g(s, \omega)\, dw(s, \omega), \qquad \int_0^t g^2(s, \omega)\, ds < \infty \text{ a.s.} \tag{33a}$$

The increasing function associated with the martingale $M(\cdot)$ is

$$\langle M(t, \omega)\rangle = \int_0^t g^2(s, \omega)\, ds. \tag{33b}$$

We shall replace the measures P_0 and P_1 by two other measures P_{00} and P_{11} defined on the fields $\{\mathscr{F}_t\} = \{\sigma(x(s, \omega), s \leqq t)\}$ by the formulas

$$P_{00}\{\omega : x(t_1, \omega) \leqq a_1, \cdots, x(t_n, \omega) \leqq a_n\} = P\{\omega : M(t_1, \omega) \leqq a_1, \cdots, M(t_n, \omega) \leqq a_n\} \tag{34a}$$

$$P_{11}\{\omega : x(t_1, \omega) \leqq a_1, \cdots, x(t_n, \omega) \leqq a_n\} \quad P\{\omega : M(t_1, \omega) + \int_0^{t_1} z(s, \omega)\, ds \leqq a_1, \cdots\} \tag{34b}$$

where

$$E \int_0^T |z(s)|\, ds < \infty, \int_0^T z^2(s)\, ds < \infty, \quad \text{a.s.} \tag{35a}$$

and $M(\cdot)$ obeys (33) and

$$M(t) - M(s) \perp\!\!\!\perp \sigma\{z(\tau), M(\tau), \tau \leqq s\}. \tag{35b}$$

LEMMA 5. *Consider a stochastic process* $x(\cdot, \cdot)$ *with the representation*

$$dx(t, \omega) = z(t, \omega)\, dt + dM(t, \omega) = z(t, \omega)\, dt + g(t, \omega)\, dw(t, \omega) \tag{36}$$

where $z(\cdot)$ *and* $M(\cdot)$ *satisfy* (35). *Then* $|g(t, \omega)|$ *is measurable with respect to* $\mathscr{F}_t = \sigma\{x(\tau, \omega), \tau \leqq t\}$ *for almost all* $t \in [0, T]$.

PROOF. By direct calculation, and repeated use of Schwarz's inequality, we can prove that for a sequence of partitions $\{0 \leqq t_1 \leqq t_2 \cdots \leqq t\}$,

$$p-\lim \textstyle\sum_{\{t_i\}} [x(t_{i+1}, \omega) - x(t_i, \omega)]^2 = \int_0^t g^2(s, \omega)\, ds \tag{37}$$

as $\max |t_{i+1} - t_i| \to 0$. Now the conclusion is obvious.

THEOREM 3. *Consider the two processes* $\{x(t, \omega), \mathscr{F}_t, P_{ii}\}$, $i = 1, 0$, *when the* $\{P_{ii}\}$ *are defined by* (33)–(35). *Assume that*

$$E \textstyle\int_0^T |z(t, \omega)/g(t, \omega)|\, dt < \infty, \qquad \int_0^T |z(t, \omega)/g(t, \omega)|^2\, dt < \infty \text{ a.s.} \tag{38}$$

Then $P_1 \ll P_0$ *and*

$$\frac{dP_{11}}{dP_{00}}(\omega) = \exp\left\{\int \frac{\hat{z}_{11}(t, \omega)}{|g(t, \omega)|^2}\, dx\,(t, \omega) - \frac{1}{2}\int \frac{\hat{z}_{11}^2(t, \omega)}{|g(t, \omega)|^2}\, dt\right\}, \quad \text{on } A \tag{39a}$$

$$\frac{dP_{11}}{dP_{00}}(\omega) = 0 \quad \text{on } \bar{A} \tag{39b}$$

where

$$\hat{z}_{11}(t, \omega) = E_{P_{11}}[z(t, \omega) \mid \mathscr{F}_t], \qquad A = \left\{\omega: \frac{dP_{11}}{dP_{00}} > 0\right\}. \tag{39c}$$

REMARK 1. If $g(t, \omega) \equiv 0$ on any interval, then the measures P_{11} and P_{00} will be trivially singular unless $z(t, \omega) \equiv 0$ on the same interval; in order not to burden the notation, we shall henceforth assume that $|g(t, \omega)| > 0$ a.e. $t \in [0, T]$.

REMARK 2. Instead of $|g(t, \omega)|^2$ in (39a) we can write $d\langle M(t, \omega)\rangle/dt$.

PROOF OF THEOREM 3. From the definition (34) of P_{11} and P_{00} we can write

$$dx\,(t, \omega) = z(t, \omega)\, dt + g(t, \omega)\, dw\,(t, \omega) \quad \text{under } P_{11}$$

$$dx\,(t, \omega) = g(t, \omega)\, dw\,(t, \omega) \quad \text{under } P_{00}.$$

On dividing by the $\mathscr{F}_t$-measurable function $|g(t, \omega)|$ we can write (whenever $|g(t, \omega)| > 0$, cf., Remark 1),

(40a)[2] $$\text{under } P_{11}: \frac{1}{|g(t, \omega)|}\, dx\,(t, \omega) = \frac{z(t, \omega)}{|g(t, \omega)|}\, dt + d\tilde{w}(t, \omega)$$

$$\text{under } P_{00}: \frac{1}{|g(t, \omega)|}\, dx\,(t, \omega) = d\tilde{w}(t, \omega) \tag{40b}$$

[2] We take $\{1/g(t, \omega)\} = 1$ on the t-set where $g(t, \omega) = 0$.

where

(41) $$\tilde{w}(t, \omega) = \int_0^t [g(s, \omega)/|g(s, \omega)|]\, dw\,(s, \omega).$$

Now it is easy to prove by the Lévy–Doob theorem ([7] page 384) that $\{^W(t, \omega), \mathscr{B}_t P\}$ is again a Wiener process (see also Nisio [23] Lemma 1). But now we have a problem to which our previous Theorem 2 can directly be applied. Doing so immediately yields the statements of Theorem 3.

Having obtained a generalization to martingale reference measures, we can go further and by the chain rule treat more general problems, e.g., those in which we have

$$P_{111}: dx(t, \omega) = z(t, \omega)\, dt + a(t, \omega)\, dt + g(t, \omega)\, dw\,(t, \omega)$$

$$P_{000}: dx\,(t, \omega) = a(t, \omega)\, dt + g(t, \omega)\, dw\,(t, \omega)$$

where $g(\cdot)$, $a(\cdot)$, and $[a(\cdot)+z(\cdot)]$ obey the hypotheses of Theorem 3. In other words, the reference measure P_{000} now describes what has been called an Itô process [7]. Theorem 3 is an extension of Theorem 2; Theorem 1 can be similarly extended but we shall not give the details here. We turn instead to the much-studied important special case in which the process $\{x(t, \omega), \mathscr{F}_t\}$ is Gaussian under both measures P_1 and P_0.

4. Gaussian processes. When the measures P_0 and P_1 are Gaussian, and P_0 is Wiener, then the results of Theorems 1 and 2 can be somewhat strengthened. In the first place, for Gaussian processes it is by now well known that absolute continuity implies mutual absolute continuity. Therefore, the set A in those theorems will have probability one under both P_1 and P_0 and therefore the statements (8b) and (15b) can be omitted. Secondly, in Theorem 2 we imposed conditions that are in general stronger than necessary for absolute continuity. In the Gaussian case, the necessary and sufficient conditions have been found by Shepp [26] (see also Kailath[18]).

THEOREM 4. (*Shepp*). *A Gaussian process $\{x(t, \omega), \mathscr{F}_t, P_1\}$ is equivalent to a Wiener process, say $\{x(t, \omega), \mathscr{F}_t, P_0\}$ if and only if*

(42) (1) $E_{P_1}[x(t, \omega)]$ *is square-integrable on* $[0, T]$

(2) $$\mathrm{Cov}_{P_1}[x(t, \omega), x(s, \omega)] = t \wedge s + \int_0^t \int_0^s K(u, v)\, du\, dv$$

where $\int_0^T \int_0^T K^2(u, v)\, du\, dv < \infty$ *and* -1 *is not an eigenvalue of* $K(\cdot, \cdot)$ *on* $[0, T] \times [0, T]$.

For such a kernel K, we can define another Volterra kernel as the unique square-integrable solution of the Wiener–Hopf type of integral equation

(43) $$h(t, s) + \int_0^t h(t, \tau) K(\tau, s)\, d\tau = K(t, s), \qquad 0 \leqq s \leqq t \leqq T,$$

$$h(t, s) = 0, \qquad s > t.$$

Also let us define

$$\hat{z}_1(t, \omega) = \int_0^t h(t, s)\, dx\,(s, \omega), \qquad 0 \leqq t \leqq T. \tag{44}$$

Then we have the following result.

THEOREM 5. *The Radon-Nikodym derivative for the processes defined in Theorem* 4 *can be written*

$$\frac{dP_1}{dP_0} = \exp\left\{\int_0^T \hat{z}_1(t, \omega)\, dx\,(t, \omega) - \frac{1}{2}\int_0^T \hat{z}_1{}^2(t, \omega)\, dt\right\}. \tag{45}$$

PROOF. One proof has been given by Hitsuda [10], who shows that the function $\phi(\cdot, \cdot)$ of Theorem 1 can be written as a Wiener integral

$$\phi(t, \omega) = \int_0^t h(t, s)\, dx\,(s, \omega) \tag{46}$$

where $h(\cdot, \cdot)$ can then be identified as the unique square-integrable solution of (43). We may note that though the $\phi(\cdot)$ of Theorem 1 is readily seen to be Gaussian, the fact that it can be written as in (46) is not immediate. For by a result of Nisio [23], it is possible that the Gaussian process $\phi(\cdot)$ may have to be written with a stochastic integrand, viz., as

$$\phi(t, \omega) = \int_0^t h(t, s, \omega)\, dx\,(s, \omega). \tag{47}$$

Another proof of the theorem has been given by Kailath [14] who used certain Hilbert-space results of Gohberg and Krein [8] on the factorization of operators and certain identities for Fredholm–Carleman determinants. We may note that some further discussion of the formula (45) is given in [14], including in particular the reasons for adopting the notation $\hat{z}_1(\cdot)$ for the integral in (44). We may also note that, as will be described elsewhere, the Hilbert-space proof in [14] can be used, along with the concept of a reproducing-kernel Hilbert space, to extend the formula (45) to the case of any two mutually absolutely continuous Gaussian measures. The Hilbert-space derivations of these extensions should have martingale analogs, which suggests the possibility of some interesting generalizations of martingale theory.

REFERENCES

[1] CAMERON, R. H. and GRAVES, R. E. (1951). Additive functionals on a space of continuous functions. *Trans. Amer. Math. Soc.* **70** 160–176.

[2] DOELZ, M., HEALD, E. and MARTIN, D. (1967). Binary data transmission techniques for linear systems. *Proc. IRE* **45** 656–661.

[3] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[4] DUNCAN, T. (1970). On the absolute continuity of measures. *Ann. Math. Statist.* **41** 30–38.

[5] FROST, P. A. (1968). Estimation of continuous-time nonlinear systems. Ph.D. dissertation, Department of Electrical Engineering, Stanford Univ.

[6] GIKHMAN, I. I. and SKOROKHOD, A. V. (1966). On the densities of probability measures in function space. *Russian Math. Surveys* **21** 83–156.

[7] GIRSANOV, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theor. Probability Appl.* **5** 285–301.

[8] GOHBERG, I. Ts. and KREIN, M. G. (1966). The factorization of operators in Hilbert space. *Amer. Math. Soc. Transl. Ser.* 2 **51** 155–188.

[9] GREEN, P. E., ed. (1960). Special issue on matched filters. *IEEE Trans. Information Theory* **IT-6**.

[10] HITSUDA, M. (1968). Representation of Gaussian processes equivalent to a Wiener process. *Osaka J. Math.* **5** 299–312.

[11] ITÔ, K. and WATANABE, S. (1965). Transformation of Markov processes by multiplicative functionals. *Ann. Inst. Fourier* **15** 13–30.

[12] KADOTA, T. T. and SHEPP, L. A. Conditions for the absolute continuity between a certain pair of probability measures (1970). To appear in *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.*

[13] KAILATH, T. (1969). A general likelihood ratio formula for random signals in Gaussian noise. *IEEE Trans. Information Theory* **IT-15** 350–361.

[14] KAILATH, T. (1970). Likelihood ratios for Gaussian processes. *IEEE Trans. Information Theory* **IT-16** 276–288.

[15] KAILATH, T. (1970). A further note on a general likelihood ratio formula for random signals in noise. *IEEE Trans. Information Theory* **IT-16** 393–396.

[16] KAILATH, T. and ZAKAI, M. (1971). Absolute continuity and Radon–Nikodym derivatives for certain measures relative to Wiener measure. *Ann. Math. Statist.* **42** 000–000.

[17] KAILATH, T. (1967). On measures equivalent to Wiener measure. *Ann. Math. Statist.* **38** 261–263.

[18] KAILATH, T. Some extensions of the innovations theorem. To appear in Bell System Technical Journal, April, 1971.

[19] KUNITA, H. and WATANABE, S. (1967). On square integrable martingales. *Nagoya Math. J.* **30** 209–245.

[20] MASANI, P. and WIENER, N. (1958). The prediction theory of multivariate stochastic processes, Part I: The regularity condition. *Acta Math.* **98** 111–150.

[21] MCKEAN, H. P. (1969). *Stochastic Integrals.* Academic Press, New York.

[22] MEYER, P. A. (1966). *Probability and Potentials.* Blaisdell, Waltham, Mass.

[23] NISIO, M. (1961). Remarks on the canonical representation of strictly stationary processes. *J. Math. Kyoto Univ.* **11** 129–146.

[24] PRICE, R. and GREEN, P. E. (1958). A communication technique for multipath channels. *Proc. IRE* **46** 555–570.

[25] PROAKIS, J. G., DROUILHET, P. R., and PRICE, R. (1964). Performance of coherent detection systems using decision-directed channel measurement. *IEEE Trans. Comm. Systems.* **CS-12** 54–63.

[26] SHEPP, L. A. (1966). Radon-Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.* **37** 321–354.

[27] SHIRYAEV, A. N. (1966). Stochastic equations of nonlinear filtration for purely discontinuous Markov processes. *Problemy Peredaci Informacii* **2** 3–22.

[28] SKOLNIK, M. I. (1962). *Introduction to Radar Systems.* McGraw-Hill, New York.

[29] WOODWARD, P. M. (1953). *Probability and Information Theory, With Applications to Radar.* Pergamon Press, New York.

[30] YAGLOM, A. M. (1963). On the equivalence and perpendicularity of two Gaussian probability measures in function space. *Proc. Symp. Time Series Analysis* (Murray Rosenblatt, ed.). Wiley, New York.