The Annals of Probability
1975, Vol. 3, No. 3, 412–423

PREDICTION FROM A RANDOM TIME POINT

BY GEORG LINDGREN

University of Umeå

In prediction (Wiener-, Kalman-) of a random normal process $\{X(t), t \in R\}$ it is normally required that the time t_0 from which prediction is made does not depend on the values of the process. If prediction is made only from time points at which the process takes a certain value u, given a priori, ("prediction under panic"), the Wiener-prediction is not necessarily optimal; optimal should then mean best in the long run, for each single realization.

The main theorem in this paper shows that when predicting only from upcrossing zeros t_ν, the Wiener-prediction gives optimal prediction of $X(t_\nu + t)$ as t_ν runs through the set of zero upcrossings, if and only if the derivative $X'(t_\nu)$ at the crossing points is observed. The paper also gives the conditional distribution from which the optimal predictor can be computed.

1. Introduction. Prediction of a covariance stationary random process $\{X(t), t \in R\}$ from a time t_0 onwards is an important and thoroughly penetrated problem. This paper deals with one side of it which does not seem to have received much attention in the literature.

In the classical situation the process is observed in a set of times $T \subset R$ chosen a priori, and prediction is based on the observed values $x(s)$, $s \in T$. Optimal prediction can be achieved by means of the conditional distribution of $X(t)$ given $X(s)$, $s \in T$; optimal with respect to some criterion then means best on the average over the possible outcomes of $X(s)$ for $s \in T$. This point of view also applies if the set T is chosen at random but independently of the X-process.

If the set T from which prediction is made depends on the process to be predicted the object of course must be modified accordingly: the predictor should show a good performance over the possible occasions under which prediction is made, but these occasions are no longer the same as above.

To be specific we will consider prediction from times t_0 at which the process takes the value 0. Similar arguments and results appear when predicting from crossings of arbitrary but fixed levels, local maxima and minima etc.

The clue to the problem is contained in the following simple example. Let us "predict" the value of the derivative $X'(t_0)$ of a normal process with mean zero if it is known that $X(t_0) = 0$. If t_0 is a nonrandom time, the conditional distribution of $X'(t_0)$ given that $X(t_0) = 0$ is equal to the unconditional distribution, which is $N(0, \lambda_2^{\frac{1}{2}})$ where $\lambda_2 = V[X'(t)]$. Thus prediction should make use of this normal distribution.

Received *Annals of Statistics*, February 24, 1974; revised 9 September 1974; received *Annals of Probability*, September 25, 1974.

AMS 1970 *subject classifications*. 60G25; 62M20, 60G40.

Key words and phrases. Prediction, zero-crossings, stopping times.

If, however, the time t_0 was selected just because the process happened to be zero there, the conditional distribution of $X'(t_0)$ is no longer normal but is a double Rayleigh distribution with density

$$\frac{1}{2\lambda_2} |x| \exp(-\tfrac{1}{2}x^2/\lambda_2) ,$$

and prediction should use this distribution.

The example shows that when predicting from a time t_0 at which $X(t_0) = 0$ it can be essential whether the set T of observation times is random or not. The result of this paper shows that the conditional distributions of the process given $X(s)$, $s \in T$ are the same, whether t_0 and T are random or nonrandom, if and only if the derivative $X'(t_0)$ is a function of the observed values.

2. The model process. Let $\{X(t), t \in R\}$ be a stationary, sample path continuously differentiable normal process with mean zero and unit variance, and with the covariance function $r(\tau) = C[X(t + \tau), X(t)]$. By choosing an appropriate time scale we can assume that the derivative has unit variance:

$$\lambda_2 = V[X'(t)] = -r''(0) = 1 .$$

A sufficient condition for the sample differentiability is

$$r''(\tau) = -\lambda_2 + O(|\log|\tau||^{-a}) \qquad \text{as } \tau \to 0 \tag{1}$$

for some $a > 1$.

We are now going to construct a model process which will describe the behaviour of the X-process conditioned by the presence of a zero upcrossing. Let $\{\Omega, \mathscr{F}\}$ be the measurable space of continuous functions with the topology for uniform convergence on compact sets, where $\mathscr{F}$ is the completed Borel σ-algebra. Define two probability measures P^{vw} and P^{hw} on $\{\Omega, \mathscr{F}\}$ by letting

$$P^{vw}(B) = \lim_{h\downarrow 0} P(X(t_0 + \cdot) \in B \mid X(t_0) \in [0, h], X'(t_0) > 0)$$

and

$$P^{hw}(B) = \lim_{h\downarrow 0} P(X(t_0 + \cdot) \in B \mid X(s) \text{ has a zero upcrossing in } [t_0 - h, t_0])$$

for any finite-dimensional event $B = \{\xi \in \Omega; \xi(t_k) < u_k, k = 1, \cdots, n\}$.

The superscripts *vw* and *hw* refer to vertical window and horizontal window conditioning; the terms were introduced by Kac and Slepian (1959). The measure P^{vw} gives the values of the ordinary conditional probabilities $P(X(t_0 + \cdot) \in B \mid X(t_0), I_{\{X'(t_0)>0\}})$ at such outcomes for which $X(t_0) = 0$, $X'(t_0) > 0$, (I_A being the indicator function of A). The measure P^{hw} gives the Palm measure of the X-process given the occurrence of an event in the point process generated by zero upcrossings. As was shown by Kac and Slepian, P^{hw} also gives the ergodic distribution of the process after a zero upcrossing, i.e. $P^{hw}(B)$ is equal to the long run relative frequency of zero upcrossings t_ν such that $X(t_\nu + \cdot) \in B$, in relation to the total number of upcrossings; see Kac and Slepian (1959), Slepian (1962), Cramér and Leadbetter (1967) Chapter 11, and Lindgren (1970).

Both P^{vw} and P^{hw} are actually concentrated to the set Ω_0 of continuously differentiable functions which vanish at the origin, and there they can be described by means of the following very simple "model process," due to Slepian (1962). Let ξ denote an element of Ω_0 and define a random variable Z and a process $\{\kappa(t),\ t \in R\}$ by

$$Z = \xi'(0)\,,$$
$$\kappa(t) = \xi(t) + Zr'(t)\,.$$

This means that the process $\{\xi(t),\ t \in R\}$ can be represented as

$$\xi(t) = \kappa(t) - Zr'(t)\,,$$

where, by definition, $\kappa(0) = \kappa'(0) = 0$, $\xi(0) = 0$, $\xi'(0) = Z$. As was shown by Slepian (1962), the measures P^{vw} and P^{hw} differ only with respect to the distribution of Z: under P^{vw} the distribution of Z is one-sided normal, under P^{hw} it is Rayleigh; it has the densities

$$f_Z^{vw}(z) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp(-z^2/2) \qquad \text{for } z > 0\,, \tag{2}$$
$$f_Z^{hw}(z) = z\exp(-z^2/2) \qquad \text{for } z > 0\,.$$

Further, by Slepian, the process $\{\kappa(t),\ t \in R\}$ has the same distributions under P^{vw} as under P^{hw} and is non-stationary normal with mean zero and covariance function

$$R(s, t) = C[\kappa(s), \kappa(t)] = r(s-t) - r(s)r(t) - r'(s)r'(t)\,, \tag{3}$$

and it is independent of the derivative Z. It is easy to see that if r fulfills condition (1) then R fulfills the sufficient condition for sample function differentiability of non-stationary normal processes given, for example, in Cramér and Leadbetter (1967), Chapters 4 and 9. This gives that both P^{vw} and P^{hw} give probability one to the set Ω_0 of continuously differentiable functions.

Without loss of generality we can assume that R is nonsingular. If it is singular our main result will still hold but the proof is considerably simplified.

We will call the process $\{\kappa(t) - Zr'(t),\ t \in R\}$ the model process, and its distributions coincide with the conditional distributions of the original process $\{X(t)\}$ given an upcrossing zero at $t = 0$, the conditioning being in vw and hw sense, respectively. Especially, for fixed t the P^{hw}-distribution of $\kappa(t) - Zr'(t)$ gives the long run distribution of the X-process when observed at times t after zero upcrossings. This important remark shows that prediction of $X(t_\nu + t)$, where t_ν runs through the set of zero upcrossings, is most efficiently performed in the long run, if one uses the hw-measure P^{hw} of $\kappa(t) - Zr'(t)$.

Remark 1. The distribution of $X(t_k + t)$, where t_k is the kth zero upcrossing on the positive side,

$$t_1 = \inf\{t > 0;\ X(t) = 0,\ X'(t) > 0\}\,,$$
$$t_k = \inf\{t > t_{k-1};\ X(t) = 0,\ X'(t) > 0\}\,,$$

is given neither by P^{vw} nor by P^{hw}. The exact distribution contains an extremely complicated condition and is therefore not readily accessible in general.

REMARK 2. Model processes similar to $\{\kappa(t) - Zr'(t)\}$ can be constructed for the process $X(t_\nu + t)$ conditioned by the presence at time t_ν of an upcrossing of an arbitrary prescribed level u or of a local maximum. To account for an upcrossing of u one just adds a term $ur(t)$ to $\xi(t)$ and gets a model process

$$\xi_u(t) = ur(t) + \kappa(t) - Zr'(t),$$

where $Z = \xi_u'(0)$ has the same meaning as before, while a local maximum requires a model process

$$\xi_*(t) = \Delta(t) - YB(t)$$

where B is a certain function, $\Delta(t)$ is a process just a little more complicated than $\kappa(t)$, and $Y = \xi_*''(0)$ is a random variable with a distribution that depends on the definition used in the conditioning; see Lindgren (1970), where the model process $\xi_*(t)$ is further specialized to have the height of the maximum equal to u. Similar results will hold and the same technique will do as in the zero-crossing problem.

We end this section with two examples of prediction in the model process.

EXAMPLE 1. Suppose we want to predict the value of $X(t_\nu + t)$ knowing only that t_ν is a zero upcrossing. The optimal predictor in least square sense of a random variable is given by the expectation of the variable, and since $E^{vw}[\kappa(t)] = E^{hw}[\kappa(t)] = 0$ and

$$E^{vw}[Z] = \left(\frac{2}{\pi}\right)^{\frac{1}{2}},$$

$$E^{hw}[Z] = \left(\frac{\pi}{2}\right)^{\frac{1}{2}}, \qquad V^{hw}[Z] = 2 - \frac{\pi}{2},$$

the optimal least square prediction of $\xi(t) = \kappa(t) - Zr'(t)$ is given by

$$\hat{\xi}_{vw}(t) = E^{vw}[\xi(t)] = -E^{vw}[Z]r'(t) = -\left(\frac{2}{\pi}\right)^{\frac{1}{2}} r'(t)$$

and

$$\hat{\xi}_{hw}(t) = E^{hw}[\xi(t)] = -E^{hw}[Z]r'(t) = -\left(\frac{\pi}{2}\right)^{\frac{1}{2}} r'(t),$$

respectively.

As mentioned above, the long run distribution of $X(t_\nu + t)$, when t_ν runs over the set of zero upcrossings, is given by P^{hw} and thus the predictor $\hat{\xi}_{hw}(t) = -(\pi/2)^{\frac{1}{2}} r'(t)$ gives optimal least square prediction in the long run. The mean square error for the two predictor functions are, taking $\hat{\xi}_{hw}(t)$ first,

$$\begin{aligned} E^{hw}[|\xi(t) - \hat{\xi}_{hw}(t)|^2] &= E^{hw}\left[\left|\kappa(t) - \left(Z - \left(\frac{\pi}{2}\right)^{\frac{1}{2}}\right) r'(t)\right|^2\right] \\ &= V^{hw}[\kappa(t)] + V^{hw}[Z]r'(t)^2 \\ &= 1 - r(t)^2 + \left(1 - \frac{\pi}{2}\right) r'(t)^2 \end{aligned} \tag{4}$$

and

$$\begin{aligned} E^{hw}[|\xi(t) - \hat{\xi}_{vw}(t)|^2] &= E^{hw}[|\xi(t) - \hat{\xi}_{hw}(t) + \hat{\xi}_{hw}(t) - \hat{\xi}_{vw}(t)|^2] \\ &= E^{hw}[|\xi(t) - \hat{\xi}_{hw}(t)|^2] + |\hat{\xi}_{hw}(t) - \hat{\xi}_{vw}(t)|^2 \\ &= 1 - r(t)^2 + \left(1 - \frac{\pi}{2}\right) r'(t)^2 + \left(\left(\frac{\pi}{2}\right)^{\frac{1}{2}} - \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\right)^2 r'(t)^2 . \end{aligned}$$

Comparison with (4) shows that the use of the inferior predictor $\hat{\xi}_{vw}(t)$ increases the mean square error by a quantity $((\pi/2)^{\frac{1}{2}} - (2/\pi)^{\frac{1}{2}})^2 r'(t)^2$.

Thus $X(t_\nu + t)$ should be predicted by $-(\pi/2)^{\frac{1}{2}} r'(t)$ and the mean square prediction error is given by (4).

EXAMPLE 2. Suppose now that, in addition to the zero crossing, we have observed the derivative $X'(t_\nu) = z$ and that we want to use this extra information to improve the prediction of $X(t_\nu + t)$.

In the model process the derivative $\xi'(0) = Z$ is independent of $\kappa(t)$. Conditioned on $Z = z$ (the conditioning being in ordinary vw or Radon–Nikodym meaning) we therefore predict $\xi(t) = \kappa(t) - Zr'(t)$ with the optimal least square predictor

$$\hat{\xi}_{vw}(t) = \hat{\xi}_{hw}(t) = -zr'(t) .$$

Since the distribution of $\kappa(t)$ is the same under P^{vw} as under P^{hw} the long run mean square error now is

$$E^{hw}[|\xi(t) + Zr'(t)|^2] = E[|\kappa(t)|^2] = V[\kappa(t)] = 1 - r(t)^2 - r'(t)^2 . \tag{5}$$

Thus $X(t_\nu + t)$ should be predicted by $-X'(t_\nu)r'(t)$ and the mean square prediction error is given by (5).

As the two examples show, the distributions P^{vw} and P^{hw} give rise to the same predictor function if we know the derivative at the zero crossing. The theorem in the next section shows that this is the only case in which P^{vw} and P^{hw} give the same result.

3. Main theorem. We now turn to the general problem of prediction from a zero upcrossing t_0 and onwards. The predictor may use observations of $X(t_0 + s)$ for s belonging to some fixed finite or infinite set T. A typical example of a set T is $\{s \leqq 0\}$, in which case one has observations of the entire past preceding the zero crossing. Another typical T-set is $\{-n, -n + 1, \cdots, 0\}$; the predictor may then use the actual value of the process $X(t_0)$ (which is zero) and the delayed values $X(t_0 - 1), \cdots, X(t_0 - n)$; see Section 4. This prediction problem can be solved by use of the conditional distributions of the model process $\xi(t) = \kappa(t) - Zr'(t)$ given $\xi(s)$ for $s \in T$.

Therefore, let P^{vw} and P^{hw} be the previously defined probability measures on the space $\{\Omega, \mathscr{F}\}$ of continuous functions with the completed Borel σ-algebra $\mathscr{F}$, and let

$$\mathscr{A}_T = \sigma\{\xi(s), s \in T\}$$

be the sub-σ-algebra generated by the observed ξ-variables. For any $B \in \mathscr{F}$, let

$$P^{vw}(B \mid \mathscr{A}_T) \quad \text{and} \quad P^{hw}(B \mid \mathscr{A}_T) \tag{6}$$

be conditional probabilities of B given the σ-algebra $\mathscr{A}_T$, i.e. $\mathscr{A}_T$-measurable functions such that

$$P^i(A \cap B) = \int_A P^i(B \mid \mathscr{A}_T)\, dP^i \qquad (i = vw, hw)$$

for all $A \in \mathscr{A}_T$. Since the topology on Ω is metrizable, complete and separable, there exist regular conditional probability measures given $\mathscr{A}_T$, i.e. (6) defines probability measures on $\{\Omega, \mathscr{F}\}$ for every fixed outcome; see Breiman (1968), Theorems 4.34 and A.46.

Also let $\bar{\mathscr{A}}_T$ be the completion of $\mathscr{A}_T$ with respect to P^{vw} and P^{hw},

$$\bar{\mathscr{A}}_T = \{A \in \mathscr{F}; \exists \bar{A} \in \mathscr{A}_T, P^{vw}(A \bigtriangleup \bar{A}) = P^{hw}(A \bigtriangleup \bar{A}) = 0\}.$$

(It follows from the definition of P^{vw} and P^{hw} that they are absolutely continuous with respect to each other, and thus give rise to the same $\bar{\mathscr{A}}_T$.)

THEOREM. *The conditional probabilities $P^{vw}(\cdot \mid \mathscr{A}_T)$ and $P^{hw}(\cdot \mid \mathscr{A}_T)$ for the model process $\xi(t) = \kappa(t) - Zr'(t)$ coincide (almost surely) if and only if $Z = \xi'(0)$ is measurable with respect to $\bar{\mathscr{A}}_T$.*

REMARK 3. The theorem implies that the *vw*- and *hw*-conditional distributions give the same predictor function if and only if the derivative at the crossing point is included in the observed variables. This is therefore the only case in which the classical (*vw*) conditional-distribution predictor gives us long run optimal prediction; in all other cases the *hw*-predictor performs better in the long run.

PROOF. To simplify notations, if f is a nonrandom function or a stochastic process, let $(f)_T$ denote its restriction to the set T. If $T^n = \{t_1, \cdots, t_n\}$ is a finite subset of T, let $(f)_n$ be the restriction of f to T^n.

The if-part of the theorem is then almost trivial: if Z is measurable $\bar{\mathscr{A}}_T$ then it is a function of the observed variables and is therefore itself observable. The only uncertainty in the outcome is therefore contributed by the κ-process and since that process is independent of Z and has the same distribution under P^{vw} as under P^{hw} the conditional probabilities will be the same. To put it formally, let B be an arbitrary finite-dimensional set of functions, define

$$B_z = B + zr' = \{\xi \in \Omega; \xi - zr' \in B\},$$

and let

$$g_z(y) = P^{vw}(\kappa \in B_z \mid (\kappa)_T = y) = P^{hw}(\kappa \in B_z \mid (\kappa)_T = y)$$

be the joint, under P^{vw} and P^{hw}, conditional probability of the event $\kappa \in B_z$ given that $(\kappa)_T = y$, y being a function on T. Then g is jointly measurable in z and y, which implies that

$$g_Z((\xi)_T + Z(r')_T)$$

is measurable with respect to $\bar{\mathscr{A}}_T$ and can serve as the conditional probability of B given $(\xi)_T$ for both P^{vw} and P^{hw}. To see this, let A be an event in $\mathscr{A}_T \subset \bar{\mathscr{A}}_T$. Since $\xi \in A \wedge Z = z$ if and only if $\kappa = \xi + zr' \in A_z \wedge Z = z$ one has, for

$i = vw, hw$,

$$\begin{aligned}\int_A g_Z((\xi)_T + Z(r')_T)\, dP^i &= \int_z \{\int_{\{\kappa \in A_z\}} g_z((\kappa)_T)\, dP_\kappa^i\}\, dP_Z^i \\ &= \int_z \{\int_{y \in A_z} P^i(\kappa \in B_z \mid (\kappa)_T = y)\, dP_\kappa^i\}\, dP_Z^i \\ &= \int_z P^i(\{\kappa - zr' \in A\} \cap \{\kappa - zr' \in B\})\, dP_Z^i \\ &= P^i(\{\kappa - Zr' \in A\} \cap \{\kappa - Zr' \in B\}) \\ &= P^i(\{\xi \in A\} \cap \{\xi \in B\})\,,\end{aligned}$$

which shows that $g_Z((\xi)_T + Z(r')_T)$ fulfills the requirements on a conditional probability. Since it is independent of $i = vw, hw$ the if-part is proved.

We now turn to the more interesting only-if part. Let us for a moment consider $\xi = \kappa - Zr'$ as a process only on T, i.e. we consider $(\kappa - Zr')_T$. For any fixed value z of Z, the process $(\kappa - zr')_T$ is a normal process with mean $-z(r')_T$ and the covariance function R restricted to $T \times T$. The process $(\kappa)_T$ is also normal with the same covariance function but with mean zero. The theorem will then follow from the following three lemmas.

LEMMA 1. *Either*

$$(\kappa - zr')_T \quad \textit{is equivalent to} \quad (\kappa)_T \quad \textit{for all} \quad z$$

or

$$(\kappa - zr')_T \quad \textit{is orthogonal to} \quad (\kappa)_T \quad \textit{for all} \quad z \neq 0\,.$$

LEMMA 2. *If $(\kappa - zr')_T$ is orthogonal to $(\kappa)_T$ then Z is measurable $\bar{\mathscr{A}}_T$.*

LEMMA 3. *If $(\kappa - zr')_T$ is equivalent to $(\kappa)_T$ then the conditional probabilities $P^{vw}(\cdot \mid \mathscr{A}_T)$ and $P^{hw}(\cdot \mid \mathscr{A}_T)$ are different.*

The only-if part of the theorem is an immediate consequence of the lemmas.

PROOF OF LEMMA 1. The lemma follows from the well-known fact that normal processes with the same covariance function and different mean functions are either orthogonal or equivalent. This can be formulated rather nicely using the terminology of Reproducing Kernel Hilbert Spaces (RKHS). Even though we do not need the Hilbert space here the following notation borrowed from Parzen (1959) is useful.

Let $T^n = \{t_1, \cdots, t_n\}$, $n = 1, 2, \cdots$ be an increasing sequence of finite subsets of T such that $\bigcup_1^\infty T^n$ is dense in T. Write

$$R_n = (R(t_i, t_j))_{i,j=1,\cdots,n}$$

for the covariance matrix of $\kappa(t_1), \cdots, \kappa(t_n)$ and note that R_n^{-1} exists by assumption. Also recall the notation $(f)_n = (f(t_1), \cdots, f(t_n))'$ for the restriction of a nonrandom or random function f to the set T^n, and write

(7) $$(f, g)_n = \sum_{i,j=1}^n f(t_i)(R_n^{-1})_{ij} g(t_j)$$

for the scalar product of f and g on T^n with respect to R_n^{-1}.

If $(\kappa)_T$ is a normal process on T with mean zero and covariance function

$(R)_{T\times T}$ and if $(m)_T$ is any function on T then it holds that either $(\kappa + m)_T$ is orthogonal to $(\kappa)_T$ or $(\kappa + m)_T$ is equivalent to $(\kappa)_T$. The criterion for the two alternatives is simple. It can be shown (cf. Parzen, Theorem 6E) that

$$(m, m)_n = \sum_{i,j=1}^{n} m(t_i)(R_n^{-1})_{ij} m(t_j)$$

is never decreasing so that

$$\lim_{n\to\infty} (m, m)_n = (m, m)_T \leqq \infty$$

always exists. Then orthogonality between $(\kappa + m)_T$ and $(\kappa)_T$ holds if and only if $(m, m)_T = \infty$ while equivalence holds if and only if $(m, m)_T < \infty$; see Parzen, Theorems 6E and 9A.

We now apply this to the mean value function $-z(r')_T$. It directly follows that if $(\kappa - zr')_T$ is orthogonal to $(\kappa)_T$ for some $z \neq 0$ then

$$\lim_{n\to\infty} (-zr', -zr')_n = z^2 \lim_{n\to\infty} (r', r')_n = \infty$$

so that orthogonality holds for all $z \neq 0$. Therefore we have

(8) $(\kappa - zr')_T$ is orthogonal to $(\kappa)_T$ for all $z \neq 0$ if and only if $(r', r')_T = \lim_{n\to\infty} (r', r')_n = \infty$.

On the other hand, under equivalence the probability measure $P_{(\kappa - zr')_T}$ induced by $(\kappa - zr')_T$ has a density, or Radon–Nikodym derivative, with respect to the measure $P_{(\kappa)_T}$ induced by $(\kappa)_T$. This density can be constructed from the quotient between the finite-dimensional densities of $\kappa(t_k) - zr'(t_k)$ and $\kappa(t_k)$ respectively for $k = 1, \cdots, n$; $n = 1, 2, \cdots$:

$$\begin{aligned}\frac{f_{(\kappa - zr')_n}(x)}{f_{(\kappa)_n}(x)} &= \frac{\exp\{-\frac{1}{2}(x + zr')_n' R_n^{-1}(x + zr')_n\}}{\exp\{-\frac{1}{2}(x)_n' R_n^{-1}(x)_n\}} \\ &= \exp\left\{-z \cdot (x)_n' R_n^{-1}(x)_n - \frac{z^2}{2}(r')_n' R_n^{-1}(r')_n\right\} \\ &= \exp\left\{-z \cdot (x, r')_n - \frac{z^2}{2}(r', r')_n\right\},\end{aligned} \tag{9}$$

where we used the notation (7). Since we have assumed equivalence the limit $\lim_{n\to\infty} (r', r')_n = (r', r')_T$ exists finite, and so does $\lim_{n\to\infty} (x, r')_n = (x, r')_T$ for almost all $x\,(P_{(\kappa)_T})$. Therefore the likelihood ratio (9) has a limit $\exp\{-z\cdot(x, r')_T - (z^2/2)(r', r')_T\}$ which is then the desired Radon–Nikodym derivative. We summarize (cf. Parzen, Theorem 9A):

(10) $(\kappa - zr')_T$ is equivalent to $(\kappa)_T$ for all z if and only if $(r', r')_T = \lim_{n\to\infty} (r', r')_n < \infty$; the density is then given by

$$\frac{dP_{(\kappa - zr')_T}}{dP_{(\kappa)_T}}(x) = \exp\left\{-z \cdot (x, r')_T - \frac{z^2}{2}(r', r')_T\right\}.$$

Proof of Lemma 2. We have to show that if $\lim_{n\to\infty} (r', r')_n = \infty$ so that $(\kappa - zr')_T$ is orthogonal to $(\kappa)_T$ for all $z \neq 0$, then $Z = \xi'(0)$ is (almost surely) a function of the observed variables $(\xi)_T$. This is actually an estimation problem: we want to estimate the parameter Z in the process $(\kappa - Zr')_T$ perfectly, i.e. without error. To achieve this we construct an estimator as a limit (as $n \to \infty$) of the maximum-likelihood estimator

$$z_n^* = -\frac{(x, r')_n}{(r', r')_n} = -\frac{1}{(r', r')_n}(x)_n' R_n^{-1}(x)_n$$

based on the likelihood ratio (9) for the observations $(x)_n = (x(t_1), \cdots, x(t_n))'$. Since κ has mean zero and the covariance function R and is independent of Z, the conditional mean and variance of z_n^* given that $Z = z_0$ are

$$E[z_n^* \mid Z = z_0] = -\frac{(E[\kappa - z_0 r'], r')_n}{(r', r')_n} = z_0 \cdot \frac{(r', r')_n}{(r', r')_n} = z_0 ,$$

$$V[z_n^* \mid Z = z_0] = \frac{1}{(r', r')_n^2} \cdot (r')_n' R_n^{-1} R_n R_n^{-1}(r')_n = \frac{1}{(r', r')_n}$$

so that z_n^* is unbiased as an estimator of Z. Since $(r', r')_n \to \infty$ as $n \to \infty$ we can conclude that z_n^* tends in quadratic mean to Z as $n \to \infty$.

What we actually want to show is that z_n^* tends almost surely to Z. This now follows from a martingale argument; considered only at outcomes such that $Z = z_0$, the sequence $z_1^*, z_2^*, \cdots$ is a reverse martingale, i.e.

$$E[z_{n-1}^* \mid z_n^*, z_{n+1}^*, \cdots] = z_n^* . \tag{11}$$

To prove this, it suffices to show that

$$C[z_j^*, z_k^* \mid Z = z_0] = V[z_k^* \mid Z = z_0] \qquad \text{if } 1 \leqq j \leqq k , \tag{12}$$

because this implies (11), the z_k^* having a joint normal distribution (still under the condition that $Z = z_0$). Now

$$\begin{aligned} C[z_j^*, z_k^* \mid Z = z_0] &= \frac{1}{(r', r')_j (r', r')_k} \cdot (r')_j' R_j^{-1} E[(\kappa)_j \cdot (\kappa)_k'] R_k^{-1}(r')_k \\ &= \frac{1}{(r', r')_j (r', r')_k} (r', r')_{\min(j,k)} \\ &= \frac{1}{(r', r')_{\max(j,k)}} \\ &= V[z_{\max(j,k)}^* \mid Z = z_0] \end{aligned}$$

so that (12) holds.

We can now apply the martingale convergence theorem to the reverse martingale $z_1^*, z_2^*, \cdots$ (Doob (1953) Theorem 4.2, Section VII) and conclude that z_n^* converges almost surely $(P_{(\kappa)_T})$ to a random variable which then must be constant and equal to z_0.

Thus, as a function on the subspace $\{Z = z_0\}$ of the space Ω of all outcomes, the estimator z_n^* tends almost surely $(P_{(\kappa)_T})$ to z_0. Let Ω_1 be the subspace of Ω

on which z_n^* converges to Z and define the random variable g on Ω as

$$g(\omega) = \lim_{n\to\infty} z_n^* \qquad \text{for } \omega \in \Omega_1 ,$$
$$= 0 \qquad \text{for } \omega \in \Omega_1^c .$$

Then g is measurable with respect to $\mathscr{A}_T = \sigma\{(\xi)_T\}$; this follows since z_n^* is a function of the observed variables $\xi(t_1), \cdots, \xi(t_n)$ and thus measurable $\mathscr{A}_T$. Because $\lim z_n^*$ is equal to Z on Ω_1 it now follows that Z is measurable $\bar{\mathscr{A}}_T$:

$$\{\omega; Z \leqq z\} \subset \{\omega; g(\omega) \leqq z\} \cup \Omega_1^c$$
$$\supset \{\omega; g(\omega) \leqq z\} \cap \Omega_1 ,$$

where, by Fubini's theorem, for $i = vw, hw$,

$$P^i(\Omega_1^c) = \int_z P(z_n^* \nrightarrow z \mid Z = z)\, dP_Z^i(z) = 0 .$$

Thus $\{\omega; Z \leqq z\}$ differs from the $\mathscr{A}_T$-set $\{\omega; g(\omega) \leqq z\}$ only by a null set, and therefore Z is measurable with respect to $\bar{\mathscr{A}}_T$.

PROOF OF LEMMA 3. We have to show that if $\lim_{n\to\infty} (r', r')_n = (r', r')_T < \infty$ so that $(\kappa - zr')_T$ is equivalent to $(\kappa)_T$ for all z, then the conditional probability measures $P^{vw}(\cdot \mid \mathscr{A}_T)$ and $P^{hw}(\cdot \mid \mathscr{A}_T)$ are different. We prove this by showing that the conditional distribution of $Z = \xi'(0)$ given $\mathscr{A}_T$ is different under the vw- and hw-measures, since this of course implies that those two conditional measures are different. Actually we can compute a conditional density for Z with respect to Lebesgue measure given the observed process $(\xi)_T = (\kappa - Zr')_T$, by manipulating the density for $(\kappa - zr')_T$ which we derived in the proof of Lemma 1; see formula (10).

We start with the densities f_Z^{vw} and f_Z^{hw} for Z as they were defined by (2) and with the density (10),

$$\text{(10)} \qquad \frac{dP_{(\kappa - zr')_T}}{dP_{(\kappa)_T}}(x) = \exp\left\{-z \cdot (x, r')_T - \frac{z^2}{2}(r', r')_T\right\} .$$

Letting z vary and weighing (10) with the densities for Z we first see that $(\xi)_T = (\kappa - Zr')_T$ is equivalent to $(\kappa)_T$ and that it has the density

$$\text{(13)} \qquad \frac{dP^i_{(\kappa - Zr')_T}}{dP_{(\kappa)_T}}(x) = \int_z f_Z^i(z) \exp\left\{-z \cdot (x, r')_T - \frac{z^2}{2}(r', r')_T\right\} dz$$

($i = vw, hw$) with respect to $P_{(\kappa)_T}$. We also see that the distribution of the two-dimensional random element $(Z, (\kappa - Zr')_T)$ is absolutely continuous with respect to the product measure $dz \cdot dP_{(\kappa)_T}$ and that the density is

$$\text{(14)} \qquad \frac{dP^i_{(Z,(\kappa - Zr')_T)}}{dz \cdot dP_{(\kappa)_T}}(z, x) = f_Z^i(z) \cdot \frac{dP_{(\kappa - zr')_T}}{dP_{(\kappa)_T}}(x)$$
$$= f_Z^i(z) \cdot \exp\left\{-z \cdot (x, r')_T - \frac{z^2}{2}(r', r')_T\right\} .$$

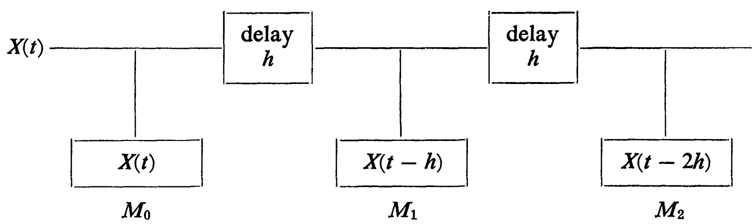
The quotient between (14) and (13) will then give a density for the conditional distribution of Z given $(\kappa - Zr')_T$, i.e. given the observed process $(\xi)_T$:

$$f^i_{Z|(\kappa - Zr')_T = (x)_T}(z) = c_i f_Z^{\,i}(z) \exp\left\{-z \cdot (x, r')_T - \frac{z^2}{2}(r', r')_T\right\} \tag{15}$$

where $c_i = c_i(x)$ is a normalizing constant. Now we are finished: the conditional distribution of Z given $(\xi)_T$ depends on whether we started with the vw-measure P^{vw} or with the hw-measure P^{hw} and this implies the conclusion of the lemma.

4. An example. As the theorem shows, the hw-conditioned model process gives better prediction from a crossing point than the generally used vw-conditioned in those cases where the derivative at the crossing is unknown. A simple and natural situation in which this occurs is the following.

In a series of memories $M_0, M_1, M_2, \cdots$ are continuously stored the actual values at time t of the signal $\{X(t)\}$ and of the delayed signals $\{X(t - kh)\}$ for $k = 1, \cdots, n$ $(n \leqq \infty)$:



Each time t_ν when a zero (or, in a more general setting, a prescribed value u) appears in the memory M_0 some action is taken, including prediction of future values based on the stored values $X(t_\nu), X(t_\nu - h), \cdots$. We may call this "prediction under panic"; the value u is the trigger value.

Formulated in the language of the main theorem prediction of $X(t_\nu + t)$ given $X(t_\nu - kh)$, $k = 0, 1, \cdots, n$ is equivalent to prediction of the model process $\xi(t) = \kappa(t) - Zr'(t)$ given the σ-algebra

$$\mathscr{A}_{T^h} = \sigma\{\xi(s), s \in T^h\}$$

where

$$T^h = \{-kh, k = 0, 1, \cdots, n\}, \qquad (n \leqq \infty).$$

Conventional Wiener-prediction of $X(t_\nu + t)$ using likelihood ratios or Radon–Nikodym derivatives then corresponds to the use of the vw-measure P^{vw} for the model process. Since the derivative Z is not measurable $\mathscr{A}_T$ (except for special processes) the main theorem implies that the hw-measure P^{hw} gives better prediction in the long run.

Naturally, a small sampling distance h makes Z nearly a function of the observed variables and then the conditioned vw- and hw-measures can be expected to give almost the same predictor. Specifically, since the set of variables $\{\xi(s), s \in T^h\}$ for small h generates almost the same σ-algebra as $\{\xi(s), s \leqq 0\}$, and since Z is measurable with respect to this latter σ-algebra, $P^{vw}(\cdot \mid \mathscr{A}_{T^h})$ and $P^{hw}(\cdot \mid \mathscr{A}_{T^h})$

should tend to the same limit when $h \downarrow 0$. A simple proof shows this if h is of the form 2^{-m} and $n = \infty$.

COROLLARY. *If $\mathscr{A}_m$ is the σ-algebra generated by $\{\xi(k/2^m), k = 0, 1, \cdots\}$ then the limits*

$$\lim_{m\to\infty} P^{vw}(B \mid \mathscr{A}_m)$$

and

$$\lim_{m\to\infty} P^{hw}(B \mid \mathscr{A}_m)$$

exist almost surely and are equal to $P(B \mid \mathscr{A})$, the joint (for vw- and hw-measures) conditional probability given the σ-algebra $\mathscr{A}$ generated by $\{\xi(s), s \leqq 0\}$.

PROOF. Since $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \cdots$ is an increasing sequence of σ-algebras with the limit $\bigcup_1^\infty \mathscr{A}_m = \mathscr{A}$, it follows from the martingale convergence theorem (cf. Doob, Theorem 4.3, Section VII) that, for $i = vw, hw$,

$$P^i(B \mid \mathscr{A}_m) = E^i[I_B \mid \mathscr{A}_m] \to E^i[I_B \mid \mathscr{A}] = P^i(B \mid \mathscr{A})$$

(almost surely) which is independent of i by the theorem.

REFERENCES

[1] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

[2] CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York.

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

[4] KAC, M. and SLEPIAN, D. (1959). Large excursions of Gaussian processes. *Ann. Math. Statist.* **30** 1215–1228.

[5] LINDGREN, G. (1970). Some properties of a normal process near a local maximum. *Ann. Math. Statist.* **41** 1870–1883.

[6] PARZEN, E. (1959). Statistical inference on time series by Hilbert space methods. I. Technical Report No. 23, Appl. Math. Statist. Lab., Stanford Univ. Also in *Time Series Analysis Papers*. Holden-Day, San Francisco (1967), 251–382.

[7] SLEPIAN, D. (1962). On the zeros of Gaussian noise. *Time Series Analysis*, ed. M. Rosenblatt. Wiley, New York, 104–115.

DEPARTMENT OF MATHEMATICAL STATISTICS
UNIVERSITY OF UMEÅ
S-901 87 UMEÅ
SWEDEN