Most of section 5.2 on Admissible Translates of General Gaussian Processes in Stochastic Processes: Inference Theory by M.M. Rao

BY

3. Example. (a) Recall that an O.U. process is a stationary Gaussian process with mean zero and covariance given by

$$r\,(s,t) = \sigma^2\,e^{-\beta\,|s-t|} \tag{1}$$

for $0 \leq s,t \leq a$ where $\beta > 0, \sigma^2 > 0$ are constants. If $\beta_i, \sigma_i^2, i = 1,2$ relate to a pair of O.U. processes, so that

$$r_i = u_i\,v_i \tag{2}$$

with

$$u_i(t) = \sigma_i^2\,e^{\beta_i t} \tag{3}$$

$$v_i(t) = e^{-\beta_i t} \tag{4}$$

then the conditions given in the theorem show that the measures $P_0^{r_i}, i = 1,2$ are equivalent iff

$$\sigma_1^2\,\beta_1 = \sigma_2^2\,\beta_2 = K \tag{5}$$

(say). The likelihood ratio is then obtained from (2) as follows where D_1, D_2 and the integrand, denoted f, are found to be:

$$D_1^2 = \frac{\beta_2}{\beta_1} \tag{6}$$

$$D_2 = \frac{2}{K}\,(\beta_1 - \beta_2) \tag{7}$$

and

$$f = -\frac{(\beta_2 - \beta_1)}{K}\exp\left[-(\beta_1 + \beta_2)\,t\right] \tag{8}$$

Substituting these in (2) and using integration by parts one gets

$$\frac{d\,P_0^{r2}}{d\,P_0^{r1}}(X) = \sqrt{\frac{\beta_2}{\beta_1}}\, e^{-\frac{(\beta_2-\beta_1)(X^2(0)+X^2(a))+(\beta_2^2-\beta_1^2)\int_0^a X^2(t)\,dt}{2K}} \tag{9}$$

This example is also discussed by Striebel [1]. [There is a small numerical discrepancy here.] See also Duttweiler and Kailath [1],[2], Kailath and Weinert [1], and Kailath, Geesy and Weinert [1], for related work.

(b) An even simpler example is noted for the equivalence of the BM and the Brownian bridge both have means zero and covariances

$$r_1\,(s,t) = \min\,(s,t) \tag{10}$$

and

$$r_2\,(s,t) = u\,(s \wedge t)\, v\,(s \vee t) \tag{11}$$

$\forall 0 \leq s, t \leq a < 1$ where $u(s) = s, v(s) = 1 - s$. In this case

$$D_1^2 = \frac{1}{1-a} \tag{12}$$

$$D_2 = 0 \tag{13}$$

$$f(t) = -1 \tag{14}$$

and since $r_i\,(0,0) = 0, i = 1, 2$, so that $X(0) = 0$, a.e. under both measures, the likelihood ratio is given by:

$$\frac{d\,P_0^{r2}}{d\,P_0^{r1}} = \frac{e^{-\frac{x^2(a)}{1-a}}}{\sqrt{1-a}} \tag{15}$$

The fact that the covariance function r_i of the O.U. process is stationary, so that it has the spectral (or Fourier) representation:

$$\begin{aligned} r_i\,(s,t) &= \sigma_i^2\, e^{-\beta_i\,|s-t|} \\ &= \frac{\sigma_i^2\,\beta_i}{\pi} \int_{-\infty}^{\infty} \frac{e^{-ix\,|s-t|}}{\beta_i^2 + x^2}\, d\,x \end{aligned} \tag{16}$$

(if $\sigma_i^2 = 1$, then it is the characteristic function of the Cauchy density, which thus appears here as the spectral density), motivates a study of obtaining conditions on the spectral functions for the equivalence and singularity of the measures (instead of the covariances). We discuss this aspect of the problem briefly.

The additional knowledge of stationarity calls for a sharper result on the dichotomy problem for the resulting likelihood ratios. The following is representative of the situation and it is due to Gladyshev [1] which we include for a comparison.

Remark 1. Recall that if $\{X_t, t \in \mathbb{R}\}$ is a (weakly) stationary process with a continuous covariance function r, then it admits the representation (by Bochner's theorem on positive definite functions)

$$r(s,t) = r(s-t) = \int_{\mathbb{R}} e^{iu(s-t)}\, dF(u) \tag{17}$$

, where F is the spectral distribution (a bounded Borel measure) and $f = \frac{dF}{du}$ is the spectral density whenever it exists.

Theorem 2. *Let* $(\Omega, \Sigma, \frac{P}{Q})$ *be Gaussian measures with zero means and continuous stationary covariances* r_k *having spectral densities* $f_k, k = 1, 2$ *for a process* $\{X_t, t \in \mathbb{R}\}$*. If the* f_k *satisfy the growth condition*

$$f_k(u) = c_k |u|^{\alpha_k} + O(|u|^{\alpha_k - 2}) \;\; \text{where } k = 1, 2 \text{ and } c_1 \cdot c_2 \neq 0 \tag{18}$$

the α_k *being real and* $\lim_{|u| \to \infty} \frac{f_1}{f_2}(u) \neq 1$*, then* $P \perp Q$*.*

It should be observed that the orthogonality of measures P, Q may obtain even when $\lim_{|u| \to \infty} \frac{f_1}{f_2}(u) = 1$ and thus for a presence of the di- chotomy, further conditions are necessary. We omit this specialization here and also the proof of the theorem referring the reader to Gladyshev's paper where the result is established by first extending Baxter's proposition and applying it to the present situation. A detailed study of the dichotomy problem for stationary Gaussian processes is given in considerable detail in Rozanov's [1] memoir.

It is clear from these computations, when the covariances of Gaussian processes are different, the work cannot be materially simplified. In fact the approximations here lead to a stochastic integral. If covariance functions are more general than the triangular ones, then it is necessary to use more sophisticated results from abstract analysis, as already seen in Theorem 1.12. To understand this structure better, we now include some additional results and the corresponding likelihood ratios. There is also an integral form of triangular covariances covering a large class of Gaussian processes, as will appear from the work below. It signifies an aspect analogous to that of Section IV.4.

Observe that if $P_{m_i}^{r_i}, i=1,2$ are equivalent measures, then by the chain rule for the RN-derivatives, one has for a.a. (ω):

$$\begin{aligned} \frac{d\,P_{m_2}^{r_2}}{d\,P_{m_1}^{r_1}}(\omega) &= \frac{d\,P_{m_2}^{r_2}}{d\,P_{m_1}^{r_2}}(\omega)\frac{d\,P_{m_1}^{r_2}}{d\,P_{m_1}^{r_1}}(\omega) \\ &= \frac{d\,P_{m_2-m_1}^{r_2}}{d\,P_0^{r_2}}(\omega)\frac{d\,P_0^{r_2}}{d\,P_0^{r_1}}(\omega) \end{aligned} \tag{19}$$

Thus $P_{m_2}^{r_2}$ is equivalent to $P_{m_1}^{r_1}$ iff both $P_{m_2}^{r_2}$ is equivalent to $P_{m_2}^{r_2}$ and $P_{m_1}^{r_2}$ is equivalent to $P_{m_1}^{r_1}$. By Proposition 1.2 we deduce that m_1, m_2 are admissible means of $P_0^{r_2}$ so that $\delta = m_2 - m_1$ is also one, by linearity of that set. Consequently $P_\delta^{r_2}$ is equivalent to $P_0^{r_1}$. This fact may be stated for a convenient reference as follows:

5. Proposition. *If $P_{m_i}^{r_i}, i=1,2,$ are Gaussian measures with means m_i and covariances r_i, then $P_{m_2}^{r_2} \equiv P_{m_1}^{r_1}$ iff $P_\delta^{r_2} \equiv P_0^{r_1}$ where $\delta = m_2 - m_1$.*

Since by Proposition 1.2 the conditions on the equivalence (or ex- istence) of the first factor on the right side of (15) are known, it is now necessary to find similar conditions for $P_0^{r_2} \equiv P_0^{r_1}$. This is a far deeper problem and considerable insight is obtained by use of abstract analysis via Aronszajn space technology. [Mathematically this is on the level of the analysis of the Behrens-Fisher problem that we discussed in Chapter II based on Linnik's penetrating study.] Its use and effective- ness in the present work was brought out by Parzen [1] and refined by Neveu [2], using a different notation. We follow this technique to elu- cidate its role in our theory and eventually obtain the likelihood ratio, in Theorem 12 below.

In order to present an equivalent version of the above proposition, with hypotheses on the covariance functions, it will be useful to intro- duce an "entropy function" for measuring the distinctness of probabilities, in place of the Hellinger "distance" used in the proof of Theorem 1.1. It is borrowed from information theory and highlights the covari- ance functions more directly than the earlier one. Thus if P, Q are probability measures on (Ω, Σ) and μ is a dominating (σ-finite) measure on Σ, for example $\frac{P+Q}{2}$ so

$$f = \frac{d\,P}{d\,\mu} \text{ and } g = \frac{d\,Q}{d\,\mu} \tag{20}$$

let I be the information functional defined by:

$$\begin{aligned} I &= I(P,Q) \\ &= \int_\Omega (g-f)\log\frac{g}{f}\,d\,\mu \\ &= \int_\Omega \log\frac{d\,Q}{d\,P}\,d\,P + \int_\Omega \log\frac{d\,P}{d\,Q}\,d\,Q \\ & I(Q,P) \end{aligned} \tag{21}$$

which thus does not depend on μ, and similarly if $\mathcal{F}_\alpha \subset \Sigma$ is a σ- algebra, and $P_\alpha, Q_\alpha, \mu_\alpha$ are the restrictions of P, Q, μ on $\mathcal{F}_\alpha$, let f_α, g_α be the corresponding densities and I_α the resulting number given by (16). If $\{\mathcal{F}_\alpha, \alpha \in J\}$ is a directed or filtering set of σ-algebras from Σ (J being a directed set and $\mathcal{F}_\alpha \subset \mathcal{F}_\beta$ for $\alpha < \beta$ in J) then is seen by the conditional Jensen's inequality applied to the convex functions $\varphi_1(x) = x \log x$ and $\varphi_2(x) = \log \frac{1}{x}, x > 0$, that $I_\alpha \leq I_\beta$ if $\alpha < \beta$; whence it is a monotone nondecreasing functional and one can verify that $\lim_\alpha I_\alpha = I \leq \infty$, where Σ is replaced by $\sigma(\cup_{\alpha \in J} \mathcal{F}_\alpha)$. [The necessary computation uses a martingale convergence theorem, since $\{u_\alpha = \frac{d\,Q_\alpha}{d\,P_\alpha}, \mathcal{F}_\alpha, \alpha \in J\}$, forms a martingale on (Ω, Σ, P), the details being standard and are found, e.g., in Rao [12],p.213.] Also it follows from definition that $I(P, Q) < \infty$ implies $Q \equiv P$ just as $H(P, Q) = 1$. But by Theorem 1.1, $H(P, Q) > 0 \Rightarrow H(P, Q) = 1$ for Gaussian measures, and similarly it is shown that, in this case, $I(P, Q) < \infty$ iff $Q \equiv P$, and $Q \perp P$ iff $I_\alpha(P, Q) = I(P_\alpha, Q_\alpha) = \infty$ for some $\alpha \in J$ (or $I(P, Q) = \infty$). This is also detailed in the above reference, and will be used without reproducing the algebra. Thus Theorem 1.1 can also be proved using this information functional I, as was originally done by Hájek [1].

If now $\alpha = (t_1, \ldots, t_n)$ is a finite set of points of the index set T for a segment of the observed process $\{X_t, t \in T\}$, let $X = (X_{t_1}, \cdots, X_{t_n})$, and $r_{jn} = [E_{P^{r_j}}(X_{t_i} X_{t_k}), 1 \leq i, k \leq n]$, $j = 1, 2$, be the $n \times n$ covariance matrices. Then the (elementary) finite dimensional likelihood ratio is given for $x = X(\omega)$ by

$$p_\alpha(x) = \frac{d\,Q_\alpha}{d\,P_\alpha}(x) = \sqrt{\frac{|r_{1n}|}{|r_{2n}|}} \exp\{-\frac{1}{2}[x'(r_{2n}^{-1} - r_{1n}^{-1})\,x']\} \tag{22}$$

where x' is the transpose of the row vector x and $|r_{jn}|$ is the determinant of the corresponding matrix. In the proof of Theorem 1.1 we simplified this expression by simultaneously diagonalizing the matrices r_{jn}, and here we proceed differently by using the I_α-functional with (17) to keep track with the covariances. Thus (16) with (17) becomes:

$$\begin{aligned} I_\alpha = I_\alpha(P, Q) \; &= E_{P_0^{r_2}}(\log p_\alpha(X)) - E_{P_0^{r_1}}(\log p_\alpha(X)) \\ &= \frac{1}{2}\,(E_{P_0^{r_2}} - E_{P_0^{r_1}})\,(X\,(r_{1n}^{-1} - r_{2n}^{-1})\,X') \\ &= \frac{1}{2}\,(E_{P_0^{r_2}} - E_{P_0^{r_1}})(\mathrm{tr}[X\,(r_{1n}^{-1} - r_{2n}^{-1})\,X']) \\ &= \frac{1}{2}\,t\,r\,[(E_{P_0^{r_2}} - E_{P_0^{r_2 n}})\,(\mathrm{r}_{1n}^{-1} - \mathrm{r}_{2n}^{-1})\,X\,X'] \\ &= \frac{1}{2}\,t\,r\,[\mathrm{r}_{2n}\mathrm{r}_{1n}^{-1} + \mathrm{r}_{1n}\mathrm{r}_{2n}^{-1} - 2\,i\,d] \end{aligned} \tag{23}$$

where $i\,d$ is the identity matrix, $t\,r$ is trace, $E_{P_i^{r_i}}$ denotes the expectation relative to the measure P^{r_i}, and where we used the standard computation for the expectations of covariances in Gaussian integrals. The crucial discovery here, due to Parzen [1], is that the expression in the last line of (18) can be identified as an element of the tensor product of the (finite dimensional) RKHSs $\mathcal{H}_{r1n}$ and $\mathcal{H}_{r2n}$. We now consider the desired product space.

If $\mathcal{H}_{r_i}$ is the RKHS for the covariance kernel r_i, then their tensor product is defined as

$$(r_1 \otimes r_2)(s_1, s_2, t_1, t_2) = r_1\,(s_1, t_1)\, r_2\,(s_2, t_2) \forall s_i, t_i \in T, i = 1, 2 \tag{24}$$

so that $r_1 \otimes r_2$ is again a covariance kernel (by Schur's lemma). Let $g_i \in \mathcal{H}_i, i = 1, 2$ and $g = g_1 \otimes g_2$ defined on $T \otimes T$ as $g(t_1, t_2) = g_1(t_1)\, g_2(t_2)$, and $\|g\|_{\mathcal{H}_1 \otimes \mathcal{H}_2} = \|g_1\|_{\mathcal{H}_1} \|g_2\|_{\mathcal{H}_2}$ derived from the inner product

$$\begin{aligned}(g, r_1 \otimes r_2(\,\cdot, \cdot, t_1, t_2)) &= \langle g_1, r_1(\,\cdot, t_1)\rangle \langle g_2, r_2(\,\cdot, t_2)\rangle \\ &= g_1(t_1)\, g_2(t_2) \\ &= g(t_1, t_2)\end{aligned} \tag{25}$$

The space $\mathcal{H}_{r_1} \otimes \mathcal{H}_{r_2}$ is the closure under the norm $\|\cdot\|_{r_1 \otimes r_2}$ of finite linear combinations of elements of the form g, and is the tensor product of the spaces $\mathcal{H}_{r_i}, i = 1, 2$ which is again a Hilbert space. It is verified easily that $\mathcal{H}_{r_1 \otimes r_2} = \mathcal{H}_{r_1} \otimes \mathcal{H}_{r_2}$.

As an example, for the triangular kernel $r\,(s, t) = \sigma^2 \min\,(s, t), 0 \leq s, t \leq 1$ the space $\mathcal{H}_r \subset C_0([0, 1])$ of absolutely continuous functions vanishing at the origin, with square integrable derivatives, i.e., $f \in \mathcal{H}_r$ iff (with $\dot{f} = \frac{d}{d\,u} f(u)$)

$$f(t) = \int_0^t \dot{f}(u)\ d\,u = \int_0^1 \dot{f}(u)\, \chi_{[0,t]}(u)\ d\,u \tag{26}$$

and the inner product

$$\langle f, g\rangle = \frac{\int_0^1 \dot{f}(u) \dot{g}(u)\ d\,u}{\sigma^2} \tag{27}$$

is the RKHS for this kernel r. Since $\frac{\partial\, r}{\partial\, s}(s, t) = \sigma^2\, \chi_{[0,t]}(s)$ one gets $r\,(\cdot, t) \in \mathcal{H}_r$ for each t and $(f, r(\cdot, t)) = f(t)$. This is the classical Wiener space. Here, replacing σ by $\sigma_i > 0, i = 1$, 2 and calling the resulting kernels r_i one gets after a simple computation for (18):

$$I_\alpha(P, Q) = \frac{n}{2} \left(\frac{\sigma_1}{\sigma_2} - \frac{\sigma_2}{\sigma_1}\right)^2 \tag{28}$$

which is unbounded as α varies on all finite subsets of [0,1], if $\sigma_1 \neq \sigma_2$. Resuming the general discussion, it follows from the above construc- tion that for $\alpha < \alpha'$ (i.e., $\alpha \subset \alpha'$) are finite sets, then $\mathcal{H}_{r1\alpha} \subset \mathcal{H}_{r1\alpha'}$ and $\{\mathcal{H}_{r1\alpha} \otimes r_{2\alpha}, \alpha \in J\}$ forms an increasingly nested set of subspaces of $\mathcal{H}_{r1} \otimes r_2$. Moreover, $\{\|r_1 - r_2\|_{\mathcal{H}_{r1\alpha} \otimes r_{2\alpha}}\}$ forms a monotone increas- ing net and hence has a limit (finite or not) by the general RKHS theory (cf., Aronszajn [1], Theorem I on p. 362). Consequently,

$$\begin{aligned}I(P, Q) &= \lim_\alpha I_\alpha(P, Q) \\ &= \lim_\alpha \|r_{1\alpha} - r_{2\alpha}\|_{\mathcal{H}_{r1\alpha} \otimes r_{2\alpha}}\end{aligned} \tag{29}$$

exists, finite or not. It is finite iff P and Q are equivalent, and singular if $I(P,Q)=\infty$. Thus for equivalence, one must have $r_1-r_2\in\mathcal{H}_{r1}\otimes r_2$. The converse implication is established by showing that the non finiteness of the pre- ceding limit implies that there is a set $A\in\cap_\alpha\mathcal{F}_\alpha$ on which P has arbitrarily small value and Q has a value close to unity so that $P\perp Q$ on Σ. In particular we have by (20) the well-known result that the scale different BM processes (i.e., $\sigma_1\neq\sigma_2$) always determine singular measures. This discussion allows us to state Proposition 5 in the fol- lowing more convenient but equivalent form involving conditions only on the means and covariances:

6. Theorem. *The Gaussian probability measures $P_{m_i}^{r_i}, i=1,2$ on (Ω,Σ) with means m_i and covariances $r_i, i=1,2$ are equivalent iff (a) $\delta=m_1-m_2\in\mathcal{H}_{r1}$, (b) $\mathcal{H}_{r1}=\mathcal{H}_{r2}$, and (c) $r_1-r_2\in\mathcal{H}_{r1}\otimes r_2(=\mathcal{H}_{r1}\otimes\mathcal{H}_{r2})$, the equalities between the spaces denoting isometric isomorphisms.*

The spaces $\mathcal{H}_r$ can be given a more interesting representation if $T\subset\mathbb{R}$ is a compact interval and r is (left or right) continuous on $T\times T$ so that $\mathcal{H}_r$ is separable, since then r may be shown to admit a (generalized) triangular form. This is utilized to obtain a sharper form of the preceding theorem which automatically includes the triangular covariances treated earlier. The result has methodological interest and reveals the structure of the problem vividly, in addition to unifying many of the previous formulations. We present the work for any r for which $\mathcal{H}_r$ is separable.

7. Theorem. *Let $P_{m_i}^{r_i}, i=1,2$, be a pair of Gaussian measures on $(\mathbb{R}^T,\mathcal{B}_T)$ with means m_i and covariances r_i. Then they are equivalent iff there exists a (σ-finite) measure space (Ω,Σ,ν) and an $R\in L^2(\Omega\times\Omega,\Sigma\otimes\Sigma,\nu\otimes\nu)$ satisfying the conditions:*

(i) $R(\omega,\omega')=R(\omega',\omega)$ for a.a. (ω,ω') and for the Hilbert-Schmidt operator $A\colon L^2(\nu)\to L^2(\nu)$ defined as

$$A\,f=\int_\Omega R(\cdot,\omega')\,f(\omega')\,d\,\nu(\omega') \tag{30}$$

-1 does not belong to the spectrum $\sigma(A)$ of A;

$$-1\notin\sigma(A) \tag{31}$$

(ii) and

$$(r_1-r_2)\,(u,v)=\int_\Omega\int_\Omega\Psi(u,\omega)\,\Psi(v,\omega')\,R(\omega,\omega')\,d\,\nu\,d\,\nu \tag{32}$$

with r_2 having a representation as

$$r_2(u,v)=\int_\Omega\Psi(u,\omega)\,\Psi(v,\omega)\,d\,\nu(\omega) \tag{33}$$

relative to the family $\{\Psi(u,\cdot),u\in T\}\subset L^2(\nu)$; and

(iii) there is a $g \in L^2(\nu)$ such that

$$(m_1 - m_2)(u) = \int_\Omega \Psi(u,\omega)\, g(\omega)\, d\,\nu(\omega) \tag{34}$$

[The family $\{\Psi(u,\cdot), u \in T\}$ need not be unique, but each such collection determining r_2 in (ii) has the same cardinality and satisfies (iii).]

The proof, to be given after Proposition 8, is helped by the following auxiliary decomposition of r_2 which is more general than the Mercer series representation (the latter demands continuity of the kernel ev- erywhere), and explains the structure of $\mathcal{H}_r$ using only the separability hypothesis. We present the result here since it has independent interest. The procedure is motivated by the work of Cramér [5], with references to his earlier contributions, and of Hida [1] both of which use the clas- sical Hellinger-Hahn expansion instead. [But a reader may skip the following discussion, and proceed to the statement of Proposition 8 for the necessary facts that are used in the proof of the theorem. However, the result will also be found useful for linear prediction considered in Chapter VIII.]

Thus let $T \subset \mathbb{R}$ be as given, $K: T \times T \to \mathbb{C}$ be a positive definite kernel, and $\mathcal{H}_K$ be its RKHS with the inner product defined in (4) of Section 1. Then $\{K(\,\cdot, t), t \in T\}$ is dense in that space. Let $\mathcal{H}_t = \overline{sp}\{K(\,\cdot, s), s \leq t\} \subset \mathcal{H}_{t'}$ for $t \leq t'$. It may and will be supposed that $\mathcal{H}_t = \cup_{s<t}\mathcal{H}_s, t \in T$ and $\mathcal{H}^0 = \cap_{t\in T}\mathcal{H}_t = \{0\}$ by replacing the collection $\{\mathcal{H}_t, t \in T\}$ with $\mathcal{H}_s^* = \cup_{s<t}\mathcal{H}_s$ where $\mathcal{H}_s \ominus \mathcal{H}^0$ if necessary. The real restriction is the assumption that $\mathcal{H}_K$ is separable which we now make and which is always satisfied if K is left (or right) continuous. [Without this, the following argument has to be refined using a more advanced result, as indicated below.]

Let $\pi_t: \mathcal{H}_K \to \mathcal{H}_t$ denote the (unique) orthogonal projection onto $\mathcal{H}_t$. Then $\{\pi_t, t \in T\}$ forms a 'resolution of the identity', i.e., a left continuous mutually commuting family of projections tending to 0 (id) as $t \to -\infty(\infty)$. The desired (classical) Hellinger-Hahn theorem states: There exists a denumerable family of elements $\{f_n, n \geq 1\}$ (not neces- sarily uniquely but of the same cardinality) such that $f_n(t) = \pi_t f_n, t \in T$, has

(i) orthogonal increments,

(ii) $\langle f_n(t), f_m(t)\rangle = \delta_{mn}\|f_n(t)\|^2$, and

(iii) $\{f_n(t_2) - f_n(t_1), [t_1,t_2] \subset \Delta = [a,b] \subset T, n \geq 1\}$ is a complete orthonormal set in $\Delta\mathcal{H}_K = \mathcal{H}_b \ominus \mathcal{H}_a$.

Moreover for each n, the additive function μ_n defined by

$$\mu_n([t_1,t_2])(=\mu_n(t_2) - \mu_n(t_1)) = \|f_n(t_2) - f_n(t_1)\|^2 \tag{35}$$

is uniquely extendible to a Borel measure on $\mathbb{R}$, and one has for each $x \in \mathcal{H}_K$:

$$\psi_n^x(t) = \frac{d\,\langle x, f_n(\,\cdot\,)\rangle}{d\,\mu_n}(t), \quad a.\,e.[\mu_n], \tag{36}$$

$$\|x\|^2 = \sum_n \int_T |\psi_n^x(t)|^2\, d\,\mu_n(t) \tag{37}$$

Taking $x = K(\cdot, t) \in \mathcal{H}_K$, for any fixed t in (21) let the corresponding ψ_n^x be denoted by $\psi_n(t, \cdot), t \in T$, and define $\mu = \sum_n \frac{1}{2^{|n|}} \mu_n$. Then μ is a Borel measure on T and $\mu_n \ll \mu, n \geq 1$. If $L^2(T, \mu)$ is the resulting Lebesgue space on T, then $\tilde{\psi}_n(t, \cdot) = \psi_n(t, \cdot)(\frac{d\mu_n}{d\mu})^{\frac{1}{2}}(\cdot) \in L^2(T, \mu)$ and the second equation of (21) gives:

$$\begin{aligned}\sum_n \int_T |\psi_n(s,t)|^2 \, d\mu_n(t) &= \sum_n \int_T |\tilde{\psi}_n(s,t)|^2 \, d\mu(t) \\ &= \int_T \sum_n \langle \tilde{\psi}_n(s,x), \tilde{\psi}_n(s,x) \rangle \, d\mu(x) \\ &= \langle K(\cdot, s), K(\cdot, s) \rangle \\ &= K(s,s)\end{aligned} \tag{38}$$

This representation is of interest here. With this (using polarization) one can express the kernel as:

$$\int_T \sum_n \langle \tilde{\psi}_n(s,x), \tilde{\psi}_n(t,x) \rangle \, d\mu(x) = K(s,t) \tag{39}$$

For simplicity let $\Psi(s, \cdot) = (\tilde{\psi}_1(s, \cdot), \tilde{\psi}_2(s, \cdot), \dots)$ be an infinite vector so that (23) can be simply written, using the inner product notation of the sequence space $(\ell^2, \langle \cdot, \cdot \rangle)$, to get

$$\int_T \langle \Psi(s,x), \Psi(t,x) \rangle_{\ell^2} \, d\mu(x) = K(s,t) \tag{40}$$

It is known (and easy to verify) that such a function space can be identified as $L^2(T, \mu; \ell^2)$ of vector valued (here ℓ^2-valued) functions on T with norm $\sqrt{K(s,s)}$. Now let $\mathcal{F} = \overline{sp}\{\Psi(t, \cdot), t \in T\} \subset L^2(T, \mu; \ell^2)$, and

$$\mathcal{H}_K = \{g \in \mathbb{C}^T : |g(t)| = |\int_T \langle \Psi(t,x), u(x) \rangle_{\ell^2} \, d\mu(x)| < \infty, u \in \mathcal{F}\} \tag{41}$$

with $\|g\|^2 = \int_T \langle u(x), u(x) \rangle_{\ell^2} \, d\mu(x)$. Also

$$(g, K(\cdot, t)) = \int_T \langle \Psi(t,x), u(x) \rangle_{\ell^2} \, d\mu(x) \tag{42}$$

then it follows that $K(\cdot, t) \in \mathcal{H}_K$ for $t \in T$ by (23'). So that $\mathcal{H}_K$ is an RKHS for the kernel K. But the general theory of Aronszajn's [1] implies that K defines uniquely such a space so that $\mathcal{H}_K = \tilde{\mathcal{H}}_K$, and the mapping $\mathcal{H}_K \to \mathcal{F}$ is an isometric isomorphism. This is the desired representation for the proof of the theorem.

It may be noted that if K is continuous on a compact interval T, then by Mercer's theorem it can be expanded as:

$$K(s,t) = \sum_{n=1}^{\infty} \frac{\psi_n(s) \, \bar{\psi}_n(t)}{\lambda_n} \tag{43}$$

the series converging uniformly and absolutely. If we consider μ as a measure concentrated on $\mathbb{N}$ with $\mu(\{n\}) = \frac{1}{\lambda_n}$ so that (24) becomes

$$K(s,t) = \int_{\mathbb{N}} \langle \psi(s,n), \bar{\psi}(t,n) \rangle_{\ell^2} \, d\,\mu(n) \tag{44}$$

then it denotes a special case of (23'). We present the general statement established above for reference as follows.

8. Proposition. *Let $K: T \times T \to \mathbb{C}$ be a covariance function such that the associated RKHS $\mathcal{H}_K$ is separable where $T \subset \mathbb{R}$. Then there exists a family of vector functions $\Psi(t, \cdot) = \{\psi_n(t,\cdot), n \geq 1\}, t \in T$ and a Borel measure μ on T such that $\Psi(t,\cdot) \in L^2(T, \mu; \ell^2)$ in terms of which K is representable as:*

$$K(s,t) = \int_T \langle \Psi(s,x), \Psi(t,x) \rangle_{\ell^2} \, d\,\mu(x) \tag{45}$$

The vector functions $\Psi(s,\cdot), s \in T$ and the measure μ may not be unique, but all such $(\Psi(t, \cdot), \mu)$ determine K and $\mathcal{H}_K$ uniquely and the cardinality of the components determining Ψ remains the same.

Important remarks: 1. It may be observed that if $\Psi(t,\cdot)$ is a scalar, then we have $K(s, t) = \int_{\mathbb{R}} \Psi(s,x) \bar{\Psi}(t,x) \, d\,\mu(x)$, which includes the tra- ditional triangular covariance with μ absolutely continuous relative to the Lebesgue measure.

2. The following notational simplification of (25) can be made. Let $\Omega = \mathbb{R} \times \mathbb{Z}, \Sigma = \mathcal{B} \otimes \mathcal{P}$ where $\mathcal{P}$ is the power set of the integers $\mathbb{Z}$, and let $\nu = \mu \otimes \alpha$ where α is the counting measure. Then $\tilde{\Psi}(t,\lambda) = \{\psi_n(t,x), n \in \mathbb{Z}\} = \Psi(t,\lambda,n), (\lambda,n) \in \Omega$. Hence

$$\|\Psi(t,\cdot)\|_{2,\nu}^2 = \int_\Omega \tilde{\Psi}(t,\omega) \bar{\tilde{\Psi}}(t,\omega) \, d\,\nu(\omega) \tag{46}$$

and

$$K(s,t) = \int_\Omega \tilde{\Psi}(s,\omega) \bar{\tilde{\Psi}}(t,\omega) \, d\,\nu(\omega) \forall s,t \in T \tag{47}$$

$$\mathcal{H}_K = \{g: g(t) = \int_\Omega \tilde{\Psi}(t,\omega) \, \bar{u}(\omega) \, d\,\nu(\omega) \forall u \in \mathcal{F} \subset L^2(\Omega,\nu)\} \tag{48}$$

where $\mathcal{F} = \overline{s\,p}\{\tilde{\Psi}(t,\cdot), t \in T\}$ and

$$\|g\|^2 = \int_\Omega |u(\omega)|^2 \, d\,\nu(\omega) \tag{49}$$

Thus $\tau: g \mapsto u$ is an isometry between $\mathcal{H}_K$ and $\mathcal{F}$. This form of (25') is convenient for the proof of the theorem below.

3. An interesting application of the form (26) is the following on a general characterization of admissible means, complementing the work of Section 1. Thus let a function $f: T \to \mathbb{C}$ be termed an (generally) admissible mean relative to a positive definite kernel $K: T \times T \to \mathbb{C}$ if it is the mean of a second order process $\{X_t, t \in T\}$ on some probability space (Ω, Σ, P) with covariance K. Since one can always take the process to be Gaussian with a given positive definite kernel as its covariance and zero mean by Kolmogorov's existence theorem (cf. Theorem I.1.1), this is the same as saying that if P_f is the measure on Σ of the X_t-process and if P is the measure of the $Y_t = X_t - f$-process, then $P_f \ll P$. Thus it is another way of stating the concept introduced before. Also

$$K_1(s,t) = f(s)\bar{f}(t) \tag{50}$$

evidently defines a (degenerate) posi- tive definite function, and we have

$$E\,(Y_s \bar{Y_t}) = K(s,t) - K_1(s,t) \tag{51}$$

. Then f is an admissible mean of P, i.e., iff $f \in M_P = \mathcal{H}_K$ by Proposition 1.2. But by (26) above, $\mathcal{H}_K$ can be realized as: $f \in \mathcal{H}_K$ iff, with K represented by (25'), f can be represented as the integral

$$f(t) = \int_\Omega \tilde{\Psi}(t,\omega)\, u(\omega)\, d\,\nu(\omega) \forall u \in \mathcal{F} \tag{52}$$

with $\|f\|_K = \|u\|_2$. But $u \in \mathcal{H}_{K_1}$ always, and from (25)-(26) where now $\mu(\{\lambda\}) = 1$ concentrating at one point and $u = 1$ so that $\|f\|_{K_1} = [\int_\Omega 1\, d\,\mu]^{\frac{1}{2}} = 1$. Since $K - K_1$ is positive definite iff $\mathcal{H}_1 \subset \mathcal{H}$ by Aron- szajn's theory [1], pp.354-5, Theorems I-II), $\|f\|_K \leq \|f\|_{K_1} = 1$. Thus we have shown that f is admissible relative to a covariance kernel K iff $\|f\|_K \leq 1$. In this form Ylvisaker [1] proved this by a somewhat different argument. Note that if K is also a continuous stationary covariance so that

$$K(s,t) = \int_{\mathbb{R}} e^{i(s-t)u}\, d\,G(u) \tag{53}$$

for a unique bounded nondecreasing G, then f is admissible iff with

$$\Psi(s,\lambda) = e^{is\lambda} \tag{54}$$

in (25) so that (26) gives $f \in \mathcal{H}_K$ iff

$$f(t) = \int_{\mathbb{R}} \Psi(s,\lambda)\, u(\lambda)\, d\,G(\lambda) \tag{55}$$

This means f is the Fourier transform of

$$dF(\lambda) = u(\lambda)\, dG(\lambda) \tag{56}$$

, F being a function of bounded variation. This sharp form of the result was obtained directly by Balakrishnan [1]. Note that (25) uses a series representation of r and hence is not good enough for the Fourier representation in the RKHS setup. In particular if

$$r(s,t) = \int_{\mathbb{R}}\int_{\mathbb{R}} e^{isx-ity}\, dF(x,y) \tag{57}$$

for harmonizable processes for admissibility of f only that $\|f\|_R \leq 1$ is concluded but its corresponding representation in the stationary case is not given, and a different argument for such specializations is needed. [As Ylvisaker notes, the RKHS argument uses only the Aronszajn theorem which does not depend on any property of T and it can be any point set. This is the generality, but then the structure of f could not be made more precise since the additional information on r is not utilized. However, the corresponding result can be obtained, using different techniques (cf. Rao [23] and Exercise 6.9 for more detail).]

4. The measure μ and the functions $\psi_n(t, \cdot)$ are obtained in the general case by the Hellinger-Hahn theory when $\mathcal{H}_K$ is separable. If the last condition is dropped, one has to consider a more advanced analysis based on the general spectral theory of a normal operator in a Hilbert space due to Plessner and Rokhlin [1]. The details of this in the context of processes are not yet available. So it will not be discussed further, but it points out the essential need to invoke deep mathematical tools even in such "naturally simple" problems.

Proof of Theorem 7. Using the notation introduced for the above result, we first observe that each element $V \in \mathcal{F} \otimes \mathcal{F}$ corresponds uniquely to a Hilbert-Schmidt (HS)-operator U on $\mathcal{F}$ (cf., Schatten [1], pp.35-36), and moreover each such U is representable by a kernel $K_0 \in L^2(T \times T, \mu \otimes \mu; \ell^2) \cap (\mathcal{F} \otimes \mathcal{F})$ where $\mathcal{F} = \tau(\mathcal{H}_K)$. In fact if $F \in \mathcal{H}_{K \otimes K}$, then there is a (not necessarily positive definite) unique K_0 such that

$$F(u,v) = \int_T\int_T (\tilde{\Psi}(u,x), K_0(x,y)\tilde{\Psi}(v,y))_{\ell^2}\, d\mu(x)\, d\mu(y) \tag{58}$$

and K_0 is hermitian if F is. All of this is an easy consequence of the theorem in Schatten referred to above. Putting $\Omega = T \times \mathbb{Z}, \nu = \mu \otimes \alpha$, as in the above remark ($\alpha =$ counting measure), (27) is expressed as:

$$F(u,v) = \int_\Omega\int_\Omega \tilde{\Psi}(u,\omega)\overline{\tilde{\Psi}(v,\omega')}G(\omega,\omega')\, d\nu(\omega)\, d\nu(\omega') \tag{59}$$

for a $G \in L^2(\nu \otimes \nu)$ in the new notation ($G = K_0$). With these identifications, we show that the conditions of Theorem 6 are equivalent to those of Theorem 7, which will establish the result.

Let $F = r_2 - r_1$ and A be the (integral) operator corresponding to F acting on $L^2(\nu)$. Then by the above recalled theorem, A associates with it a kernel $G(=R$ of the theorem). To verify the equivalence of the present conditions with those of Theorem 6, consider the isometric isomorphism $\tau\colon \mathcal{H}_{r_1} \to \mathcal{F}$ defined after (25') in which K is taken as r_1. Then $\tau(\delta) = g \in \mathcal{F}$ and this is equivalent to (iii). Also $r_2 - r_1 \in \mathcal{H}_{r_1 \otimes r_2}$ is equivalent to showing that $(\tau \otimes \tau)\,(r_2 - r_1) \in \mathcal{F} \otimes \mathcal{F}$, since $r_2 - r_1$ is hermitian. Then by (27'), there is an $R \in L^2\,(\nu \otimes \nu)$ in fact is in the subset $(\tau \otimes \tau)(\mathcal{H}_{r_1 \otimes r_2})$. This is (c) which is thus equivalent to the argument is reversible, condition (b) of Theorem 6 and (i) here are equivalent, so that all the conditions of the present result are equivalent to those of Theorem 6, as desired. □

Note. If the covariance is already given to be in a generalized triangular form, i.e., of the type (25'), the result holds and no multiplicity or Hellinger-Hahn theory is needed. Indeed such a direct application was made by Park [1] if r_1 is a triangular covariance on (T, μ) where $T \subset \mathbb{R}^n$ and μ is the Lebesgue measure. In our case r_1 is also a general covariance but $T \subset \mathbb{R}$ so that Hellinger-Hahn representation has to be (and was) invoked and μ is a σ-finite Borel measure. We present a specialization for BM as an example. If $T \subset \mathbb{R}^n$, a result corresponding to Proposition 8 is not immediately available, and an assumption of triangular (or "factorizable") covariance, i.e., the representation (25'), seems to be desirable so that Theorem 6 can still be employed.

Taking r_1 as the covariance of the BM (so it is triangular

$$r_1(u, v) = \int_0^1 \chi_{[0,u]}(t)\, \chi_{[0,v]}(t)\, d\,t \tag{60}$$

we can present conditions for equivalence of an arbitrary Gaussian measure $P_0^{r_2}$ with $P_0^{r_1}$ of the BM, first obtained by Shepp [1] using a different method, as follows. [Here $\Omega = [0, 1]$ and $d\,\nu = d\,x$, the Lebesgue measure with $K = r_1$ in (25').]

9. Corollary. *Let P_0^r correspond to the BM and P_m^s be an arbitrary Gaussian measure, both on (Ω, Σ) $\Omega = \mathbb{R}^{[0,1]}$. Then they are mutually equivalent iff there is a hermitian $R \in L^2([0,1]^2, d\,x\,d\,y)$ such that for $0 \le u, v \le 1$ $(r(u, v) = \min(u, v))$:*

(i). $s(u, v) = r(u, v) + \int_0^u \int_0^v R(x, y)\, d\,x\, d\,y$

(ii). *if A is determined by r on $L^2([0,1], d\,x)$ then $-1 \notin \sigma(A)$, and*

(iii). *there is a $g \in L^2([0,1], d\,x)$ such that $m(t) = \int_0^t g(u)\, d\,u$.*

Note that from (i) and (iii) it follows that s, m are differentiable and in fact $R(u, v) = \frac{\partial^2 s}{\partial u\, \partial v}(u, v)$ and $g(u) = \frac{d\,m}{d\,u}(u)$, a.e. Thus in the case of BM, the equivalence condition on mean and covariances can be given explicitly. The above conditions are a specialization of those in Theorem 7, but were discovered by Shepp [1] by a different procedure without using the RKHS techniques. A direct RKHS proof of this case of Shepp's was immediately followed by Kailath [1].

Since Theorem 7 shows that the kernels r_1, r_2 determine HS opera- tors and when their difference belongs to $\mathcal{H}_{r_1 \otimes r_2}$ so that $r_1 - r_2$ defines a similar operator, it is of interest to find conditions for the equivalence in terms of the latter transformations. The following is such a result, and it slightly extends a theorem due to Pitcher [7] (cf., also Root [1], p.302). Our demonstration is again based on the RKHS technique and follows from the preceding work.

10. Theorem. *Let $P_0^{r_i}, i=1,2,$ be a pair of Gaussian measures on*

...to be continued...