For a continuously indexed process the above theorem is not immediately applicable since the index is not a countable (not even) linearly ordered set. However for many second order processes it can be reduced to the preceding case as shown below. The idea here is to replace the whole process by (linear) combinations of a fixed suitably chosen countable set of random variables for which Theorem 1 is applicable. [The general case will be treated after this one.] It is based on the Karhunen-Loève representation which we now describe.

Thus let $\{X(t), t \in I \subset \mathbb{R}\}$ be a second order (scalar) process with mean function $t \mapsto m(t) = E(X(t))$ and a continuous covariance function $(s,t) \mapsto r(s,t) = E[(X(s)-m(s))\overline{(X(t)-m(t))}]$. Then $r(s,t)$ is positive definite, and if I is a compact interval, assumed hereafter, by the classical Mercer theorem (cf., e.g., Riesz and Sz.-Nagy [1], p.245) since $\int_I \int_I |r(s,t)|^2\, ds\, dt < \infty$, it can be represented by a uniformly convergent series: ('bar' for complex conjugate)

$$r(s,t) = \sum_{i=1}^{\infty} \frac{\psi_i(s)\overline{\psi}_i(t)}{\lambda_i};\ \lambda_i > 0, \tag{14}$$

where the $\psi_i(\cdot)$ are continuous functions satisfying the integral equation

$$\psi(t) = \lambda \int_I r(s,t)\psi(s)\, ds, \tag{15}$$

and $\sum_{i=1}^{\infty} \frac{1}{\lambda_i} < \infty$. Here the λ_i are the eigenvalues (counted according to their multiplicity) and ψ_i are the corresponding eigenfunctions of the "kernel" r, and $\{\psi_n, n \geq 1\}$ forms a complete orthonormal set in the Lebesgue space $L^2(I)$, with Lebesgue measure, satisfying (15). This classical result and its relation to the Hilbert-Schmidt theory of symmetric kernels is nicely treated in the above reference, and their properties are needed here. First we consider the case that $E(X(t)) = m(t) = 0$ so that the $X(t)$ are centered. Now define the random variables

$$\xi_n = \sqrt{\lambda_n} \int_I X(t)\bar{\psi}_n(t)\, dt,\ n \geq 1, \tag{16}$$

where the integral is obtained using Fubini's theorem, since $X(t,\omega)$ is jointly measurable ($r(\cdot,\cdot)$ being jointly continuous) in (t,ω). [Alternatively it may be regarded as a Bochner integral.] In any case, we get

$$E(\xi_n\bar{\xi}_m) = \sqrt{\lambda_m\lambda_n} \int_I \int_I E(X(s)\bar{X}(t))\psi_n(s)\bar{\psi}_m(t)\, ds\, dt$$

$$= \sqrt{\lambda_n \lambda_m} \int_I [\int_I r(s,t)\psi_n(s)\, ds]\bar{\psi}_m(t)\, dt$$
$$= \sqrt{\frac{\lambda_m}{\lambda_n}} \int_I \psi_n(t)\bar{\psi}_m(t)\, dt, \text{ by (15)},$$
$$= \delta_{nm}\sqrt{\frac{\lambda_m}{\lambda_n}}.$$

It follows (on expanding inner products) that $X_n(t) = \sum_{k=1}^{n} \xi_k \frac{\psi_k(t)}{\sqrt{\lambda_k}} \to X(t)$ in $L^2(P)$, by (14), where $X(t) = \sum_{n=1}^{\infty} \xi_n \frac{\psi_n(t)}{\sqrt{\lambda_n}}$, and conversely if $X(t)$ is given by this series, converging in mean, then $E(X(s)\overline{X}(t)) = \lim_n E(X_n(s)\overline{X}_n(t)) = \sum_{n=1}^{\infty} \frac{\psi_n(s)\bar{\psi}_n(t)}{\lambda_n}$ holds. If $E(X(t)) = m(t) \neq 0$, then the above argument applied to $Y(t) = X(t) - m(t)$ establishes the following classical *Karhunen-Loève* representation:

4. Proposition. *If $\{X(t), t \in I\}$ is a second order process with $E(X(t)) = m(t)$, and a continuous covariance function $r(\cdot,\cdot)$ on a compact interval I, then*

$$X(t) = m(t) + \sum_{n=1}^{\infty} \xi_n \frac{\bar{\psi}_n(t)}{\sqrt{\lambda_n}}, \quad t \in I, \tag{17}$$

holds uniformly in t, and the convergence is in $L^2(P)$ where the $\lambda_n > 0$ and ψ_n are the eigenvalues and the corresponding (complete orthonormal in $L^2(I)$) eigenfunctions of the kernel r, satisfying (15), and hence the ξ_n are orthonormal in $L_0^2(P)$, given by (16).

Let $\mathcal{F} = \sigma(X(t) \in I)$, $\mathcal{F}_\infty = \sigma(\xi_n, n \geq 1)$ be the σ-algebras generated by the random variables shown. Since each $X(t)$ is a linear combination of the ξ_n, by (17), it follows that the $X(t)$ are $\mathcal{F}_\infty$-measurable for each $t \in I$ so that $\mathcal{F} \subset \mathcal{F}_\infty$. On the other hand each ξ_n is $\mathcal{F}$-measurable, by (16), for $n \geq 1$, so that $\mathcal{F}_\infty \subset \mathcal{F}$ and hence $\mathcal{F} = \mathcal{F}_\infty \subset \Sigma$. If $\tilde{P} = P|\mathcal{F}$, it is then determined by the $X(t)$ as well as the ξ_n. Thus *using (17), we can transfer the testing problem for measures P and Q, (or $\tilde{P}, \tilde{Q}$ on $\mathcal{F} = \mathcal{F}_\infty$) to the* **sequence** *$\{\xi_n, n \geq 1\}$, to find the likelihood function $f_\infty = \frac{d\tilde{Q}^c}{d\tilde{P}}$* by the approximation procedure of Theorem 1 with $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$. Consequently, if $f_n = \frac{d\tilde{Q}_n^c}{d\tilde{P}_n}$ where $\tilde{P}_n = \tilde{P}|\mathcal{F}_n$, $\tilde{Q}_n = \tilde{Q}|\mathcal{F}_n$, then $f_n \to f_\infty$ a.e. as in Theorem 1, and f_∞ is the desired likelihood function.

This method will now be illustrated to gain an insight into the type of calculations needed for some test problems.

5. Example. Let $\{X(t), t \in [0,1]\}$ be a Gaussian process with mean function 0 and covariance function $r_b (b \neq 0)$, given by

$$r_b(s,t) = \begin{cases} \frac{\cosh bs \cosh b(1-t)}{b \sinh b}, & \text{for } s \leq t \\ \frac{\cosh bt \cosh b(1-s)}{b \sinh b}. & \text{for } t \leq s. \end{cases} \tag{18}$$

That this defines a covariance function follows from the well-known fact that any function of the form $(s,t) \mapsto r(s,t) = u(\min(s,t))v(\max(s,t))$ is a covariance function on $T \times T$ if $u, v \geq 0$ and $\frac{u}{v}$ is strictly increasing on $T \subset \mathbb{R}$. (Cf., e.g., Rao [25], p.340; and one can also verify this by computing the matrix $R_n = (r(s_i, s_j), 1 \leq i,j \leq n)$, $n \geq 1$, and showing its determinant $\det R_n > 0$ which in this case has a simple pattern.)

Now the problem is to test the hypotheses $H_0 : b = b_0 > 0$, vs $H_1 : b \in [b_0 + \varepsilon, B]$, where $\varepsilon > 0$ is given, making the hypotheses distinguishable. (The same procedure applies if $b_0 < 0$, and then $H_1 : b \in [B_1, b_0 - \varepsilon]$ but $b_0 = 0$ is excluded.) To employ Proposition 4, it is necessary to find the eigenvalues λ_n^i and the corresponding eigenfunctions ψ_n^i relative to the hypotheses $H_i, i = 0, 1$. This may be done as follows.

Consider the integral equation with r_b as its symmetric kernel:

$$\psi(t) = \lambda \int_0^1 r_b(s,t)\psi(s)\,ds. \tag{19}$$

Substituting (18) here and differentiating, it is seen that (19) is equivalent to an ordinary second order linear differential equation with suitable boundary conditions at 0 and 1:

$$\psi''(t) = (b^2 - \lambda)\psi(t);\ \psi'(0) = \psi'(1) = 0. \tag{19'}$$

Solving this equation, it is immediately found that

$$\begin{aligned} \lambda_n &= n^2\pi^2 + b^2, \quad n = 0, 1, 2, \dots, \\ \psi_0(t) = 1,\ \psi_n(t) &= \sqrt{2}\cos n\pi t, \ \ n = 1, 2, \dots, \end{aligned} \tag{20}$$

Define the *coordinate (or observable)* random variables

$$Z_n = \int_0^1 X(t)\psi_n(t)\,dt.$$

Then the Z_n are orthogonal (hence independent here) Gaussian random variables with $E(Z_n) = 0$ and $E(Z_n^2) = \frac{1}{\lambda_n}$. Note that in this particular case the eigenfunctions do not depend on b, and only the λ_n do. Hence writing λ_n^i for λ_n of (20) under the hypotheses $H_i, i = 0, 1$, we can calculate the likelihood functions f_n on $\mathcal{F}_n = \sigma(Z_1, \cdots, Z_n)$ (by setting $Z_n(\omega) = z_n$) as:

$$f_n(\omega) = [\Pi_{i=1}^n \frac{\lambda_i^1}{\lambda_i^0}]^{\frac{1}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^n z_i^2(\lambda_i^1 - \lambda_i^0)\}$$

Then by the preceding work $f_n \to f_\infty$ a.e.$[\tilde{P}]$. Since the series

$$\sum_{n=1}^\infty E_i(Z_n^2)(\lambda_n^1 - \lambda_n^0) = b_i^2 \sum_{n=1}^\infty \frac{1}{\lambda_n^i} < \infty,$$

and (E_i = expectation, Var_i = Variance under H_i)

$$\sum_{n=1}^\infty Var_i\, Z_n^2\, (\lambda_n^1 - \lambda_n^0)^2 = 2b_i \sum_{n=1}^\infty \frac{1}{(\lambda_n^i)^2} < \infty,$$

the series $\sum_{n=1}^\infty Z_n^2(\lambda_n^1 - \lambda_n^0)$ converges with probability one (by a standard result in probability theory), under both $\tilde{P}$ and $\tilde{Q}$ [they are equivalent measures], and similarly $\Pi_{n=1}^\infty \frac{\lambda_n^1}{\lambda_n^0} = \lim_{n\to\infty} \frac{\Pi_{i=1}^n(1+\frac{b_1^2}{(n\pi)^2})}{\Pi_{i=1}^n(1+\frac{b_0^2}{(n\pi)^2})} = \frac{a(b_1)}{a(b_0)} > 0$, (say) exists. Thus

$$f_\infty = [\frac{a(b_1)}{a(b_0)}]^{\frac{1}{2}} \exp\{-\frac{1}{2}\sum_{n=1}^\infty Z_n^2(\lambda_n^1 - \lambda_n^0)\}$$

exists a.e. Then the critical region A_0^k of (1) is given (on taking logs) by:

$$A_0^k = \left\{\omega : \sum_{n=1}^\infty Z_n^2(\omega) \le \frac{k}{(b_1 - b_0)}\right\}, \tag{21}$$

for a suitable $k_0 > 0$, since $\lambda_n^1 - \lambda_n^0 = b_1 - b_0 > 0$. The same result holds for all $b_1 > b_0 > 0$ so that the set A_0^k is a (one-sided) uniformly most powerful critical region (and the inequality in (21) should be reversed if $b_1 < b_0$ for a similar conclusion).

A related problem, (due to Grenander [1]) which is generalized in the next chapter, will be discussed here since it serves as a motivation for that work. With the notation of the above example, let $\{X(t), t \in [0,1]\}$ be a Gaussian process with mean zero and a continuous covariance function $r(\cdot,\cdot)$. Suppose that $H_0 : r(s,t) = r_0(s,t)$ vs $H_1 : r(s,t) = \sigma^2 r_0(s,t), \sigma \neq 1$. If λ_n^i and ψ_n^i are as in (19) for r then it is clear that $\lambda_n^2 = \sigma^2\lambda_n^1$ and if $Z_n^i = \int_0^1 X(t)\psi_n^i\,dt$ are the observable coordinates of the process, then $Z_n^i \sim N(0, \lambda_n^i)$ are independent and the corresponding likelihood ratio becomes:

$$f_n(\omega) = (\frac{1}{\sigma^2})^{\frac{n}{2}} \exp\{-\frac{1}{2}\sum_{k=1}^{n} \lambda_k^1 z_k^2 (\frac{1}{\sigma^2} - 1)\}.$$

By the preceding work $f_n \to f_\infty$ a.e. $[\tilde{P}]$, and since $\sum_{k=1}^{\infty} \lambda_k^1 Z_k^2$ converges with probability one, $f_\infty = 0(\infty)$ according as $\sigma^2 > 1(< 1)$ and in either case the corresponding probabilities are mutually singular so that H_0 and H_1 can be distinguished with probability one. If a transformation $T_\sigma : X \mapsto \sigma X$ is considered instead, then the same conclusions obtain since $r(s,t) = \sigma^2 r_0(s,t)$. In case the process is BM so that $r(s,t) = \min(s,t)$ the corresponding result was established by Cameron and Martin [2] from a different point of view. On the other hand Example 5 above shows that nontrivial likelihood functions can be obtained for the equivalence of the measures at least when covariances are triangular, but not scalar multiples of each other. An interesting generalization of this result for distinct triangular covariances will be considered, especially affine linear transformations of the BM, in the next chapter.

Several other examples, each demanding a special nontrivial treatment, have been detailed in Grenander [1],[2], which will greatly assist the reader's appreciation of the subject. The preceding elegant argument can be utilized, via Proposition 4, only when the eigenvalues and eigenfunctions λ_n, ψ_n of the covariance kernel can be explicitly calculated. That is relatively easy when the problem is converted into a differential equation with suitable (two point) *boundary conditions*, such as those given in (19'). This equivalence is a classical result in the Hilbert-Schmidt theory of linear homogeneous integral equations. In general, however, one has to obtain these λ_n, ψ_n by other means and it is not easy. [See, e.g., Riesz and Sz.-Nagy [1], Sections 95 and 96.] We encounter these situations even in relatively simple cases as seen in the important Ornstein-Uhlenbeck (or O.U.) process. We now consider this process because of its many applications. *It will be used in other illustrations as well*, again following Grenander [1,2].

6. An O.U. Process example. The O.U. process $\{X(t), t \in [a,b]\}$ is real Gaussian with mean $m(t)$ and covariance $r(s,t) = \sigma^2 \exp[-\beta|s - t|]$, $\beta > 0$, $\sigma > 0$. Since $r(\cdot,\cdot)$ is a continuous positive definite symmetric kernel, (this follows from the fact that $t \mapsto e^{-\beta|t|}$ is the characteristic function of a Cauchy distribution with parameter $\beta > 0$), we can consider as before ($a = 0, b = 1, \sigma = 1$, taken for simplicity) the integral equation:

$$\psi(t) = \lambda \int_0^1 e^{-\beta|s-t|} \psi(s)\, ds$$

$$= \lambda[e^{-\beta t} \int_0^t e^{\beta s} \psi(s)\, ds + e^{\beta t} \int_t^1 e^{-\beta s} \psi(s)\, ds],$$

so that differentiating relative to t one gets ("$'$" for $\frac{d}{dt}$):

$$\psi'(t) = \lambda\beta[-e^{-\beta t} \int_0^t e^{\beta s} \psi(s)\, ds + e^{\beta t} \int_t^1 e^{-\beta s} \psi(s)\, ds].$$

The boundary conditions are somewhat unfamiliar, and one sees that, from the above two expressions for ψ, ψ', the following as the two point boundary conditions:

$$\psi(0) - \frac{1}{\beta}\psi'(0) = 0 = \psi(1) + \frac{1}{\beta}\psi'(1). \tag{22}$$

Then the previous procedure converts the integral equation into the differential equation:

$$\psi''(t) + 2\beta(\lambda - \frac{\beta}{2})\psi(t) = 0. \tag{23}$$

Let $\alpha^2 = \beta(2\lambda - \beta) > 0$ so that $\lambda = \frac{\alpha^2+\beta^2}{2\beta} > 0$. (The case $2\lambda \leq \beta$ leads to $\lambda \leq 0$ which is inadmissible for the positive definite kernel r.) Then the solution of (23) becomes ($i = \sqrt{-1}$)

$$\psi(t) = c_1 e^{i\alpha t} + c_2 e^{-i\alpha t}, \tag{24}$$

and substitution of this in (22) gives the pair of equations:

$$\begin{aligned} c_1(1 - \frac{i\alpha}{\beta}) + c_2(1 + \frac{i\alpha}{\beta}) &= 0 \\ c_1 e^{i\alpha}(1 + \frac{i\alpha}{\beta}) + c_2 e^{-i\alpha}(1 - \frac{i\alpha}{\beta}) &= 0. \end{aligned} \tag{25}$$

For a nontrivial solution of this equation in c_1, c_2 the determinant of the coefficients must vanish which implies that

$$e^{i\alpha t} = \frac{(\beta - i\alpha)^2}{(\beta + i\alpha)^2} = \frac{(\beta - i\alpha)^4}{(\beta^2 + \alpha^2)^2}. \tag{26}$$

Consider the real part of this equation:

$$\cos 2\alpha = \frac{\alpha^4 + \beta^4 - 6\alpha^2\beta^2}{(\alpha^2 + \beta^2)^2}. \tag{27}$$

(The imaginary part gives the sine function which is obtainable from this equation immediately, and need not be discussed.) Let α_n be the zeros of this (transcendental) equation which, written as an infinite power series, shows that there are infinitely many roots of which only the positive ones are of interest here. Then the eigenvalues of the real $r(\cdot,\cdot)$ are $\lambda_n = \frac{\alpha_n^2+\beta^2}{2\beta} > 0$, and the corresponding eigenfunctions ψ_n are real and orthogonal, given from (24)–(27)as:

$$\psi_n(t) = (c_1 + c_2) \cos \alpha_n t = c \cos \alpha_n t \text{ (say)}.$$

(Because of the previous observation, an explicit form of these functions cannot be written.) Then c is chosen to normalize ψ_n, and suppose this is done. As before one defines $Z_n = \int_0^1 X(t)\psi_n(t)\,dt$. Then the Z_n are independent Gaussian random variables. To test the hypothesis $H_0 : m(t) = 0$ vs $H_1 : m(t) \neq 0$, the observable (or coordinate) Z_n have means zero under H_0, and $a_n = \int_0^1 m(t)\psi_n(t)\,dt$ under H_1 with the same variance under both hypotheses, var $Z_n = \frac{1}{\lambda_n}$. The finite dimensional likelihood ratio is given by

$$f_n(\omega) = \exp[-\frac{1}{2}\sum_{i=1}^{n} \lambda_i a_i^2 + \sum_{i=1}^{n} \lambda_i a_i Z_i(\omega)].$$

If $\sum_{i=1}^{\infty} \lambda_i a_i^2 < \infty$, then by a classical Kolmogorov theorem the series $\sum_{i=1}^{\infty} \lambda_i a_i Z_i$ converges with probability one, and $f = \lim_{n\to\infty} f_n$ exists in the same sense. Now if $g_n(t) = -\sum_{i=1}^{n} \lambda_i a_i \psi_i(t)$, then $g_n \to g$ in $L^2([0,1])$, and one gets in a similar manner

$$\begin{aligned} f_n &= \exp[-\int_0^1 g_n(t)\frac{X(t)-m(t)}{2}\,dt] \\ &\to \exp[-\frac{1}{2}\int_0^1 g(t)(X(t)-m(t))\,dt] = f(= \frac{dQ}{dP}). \qquad (28) \end{aligned}$$

Thus if $\sum_{i=1}^{\infty} \lambda_i a_i^2 < \infty$, then $Q \sim P$ and if this condition fails, $\sum_{i=1}^{\infty} \lambda_i a_i Z_n$ diverges with probability one (by the same Kolmogorov theorem) so that $Q \perp P$. It may be noted that since by Mercer's theorem $\sum_{i=1}^{\infty} \frac{1}{\lambda_i} < \infty$, the case that $m(t) = a \neq 0$, a constant leads to the singular case that $Q \perp P$, and the hypotheses can be distinguished with probability one based on a single realization. (This fact is a special case of Theorem V.1.1 to be established later, cf., also Section VII.2.)

The elegant method, illustrated in both problems above, can be employed in practical cases only if λ_n, ψ_n are explicitly calculated. But in the case of the O.U. process it is seen that our equation (27) cannot be solved for its roots easily, and hence (28) is not effectively used. However, it is possible to find (a slightly weaker) alternative procedure that still gives a reasonably satisfactory solution.

An alternative procedure. Consider the O.U. process $\{X(t), t \in [0,1]\}$ as before with mean and covariance functions under $H_0 : m(t) = 0$, and under $H_1 : m(t) \neq 0$ but satisfies a uniform Lipschitz condition of order one and with the same covariance $r(s,t) = \exp[-\beta|s-t|]$. Recall that a function g satisfies a uniform Lipschitz condition of order $\alpha > 0$ if there is an absolute constant $C > 0$ such that $|g(s) - g(t)| \leq C|s-t|^{\alpha}$ and $C = 0$ if $g = a$, constant. Let $t_1^{(n)} < \ldots < t_n^{(n)}$ be a partition π_n of $[0,1]$ at the n^{th} stage, the π_n being ordered by refinement, so that as $|\pi_n| \to 0$ the partition points form a dense set of the unit interval. For instance the binary subdivision is adequate. Since the covariance function and the means under both hypotheses are continuous and the process is Gaussian, using the form of r, one can verify that the process has continuous sample paths with probability one. Indeed this was proved by Doob ([1], pages 304-5) who observed that the O.U. process $X = \{X(t), t \in \mathbb{R}\}$ has the property that if

$$Y(t) = \sqrt{t}X(\frac{1}{2\beta}\log t), t > 0, \tag{29}$$

then (after a computation which is left to the reader as an exercise)

$$E(Y(s+t) - Y(s)) = 0; \; E(|Y(s+t) - Y(s)|^2) = \sigma_0^2 t, \tag{30}$$

and for $s_1 < s_2 \leq t_1 < t_2$, the increments $Y(s_2) - Y(s_1)$ and $Y(t_2) - Y(t_1)$ are uncorrelated (hence independent) since $Y = \{Y(t), t \in \mathbb{R}\}$ is Gaussian. Thus the Y-process is a Brownian motion, and the latter is well-known to have a.a. continuous sample paths, already proved by Wiener in the 1920s. (See also the sketch after Theorem 2.6 below.)

For a partition π_n, as above, let $X_i = X(t_i^{(n)}), \rho_i = \exp[-\beta(t_{i+1}^{(n)} - t_i^{(n)})$, and $m_i = m(t_i^{(n)})$, for fixed n. Then $\{X_1, \ldots, X_n\}$ are jointly normal random variables of the O.U. process, and the n-dimensional density $f_{\pi_n}^m$ with mean m, can be written as:

$$f_{\pi_n}^m(x_1, \cdots, x_n) = \left[\Pi_{i=1}^{n-1}(2\pi(1-\rho_i^2)\right]^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x_1 - m_1)^2 - \frac{1}{2}\sum_{i=1}^{n-1}\frac{[x_{i+1} - m_i - \rho_i(x_i - m_i)]^2}{1-\rho_i^2}\}. \tag{31}$$

Hence the "log likelihood function" of the hypotheses $H_0 : m = 0$, vs $H_1 : m \neq 0$ with $X_i(\omega) = x_i$ is given for each such partition by:

$$\log\left(\frac{f_{\pi_n}^m}{f_{\pi_n}^0}\right)(\omega) = m_1 x_1 - \frac{m_1^2}{2} + \sum_{i=1}^{n-1}\frac{x_{i+1} - \rho_i x_i}{1+\rho_i} \cdot \frac{m_{i+1} - \rho_i m_i}{1-\rho_i}$$

$$-\frac{1}{2}\sum_{i=1}^{n-1}\frac{1-\rho_i}{1+\rho_i}\left(\frac{m_{i+1}-\rho_i m_i}{1-\rho_i}\right)^2. \tag{32}$$

The right side terms are Riemann sums which converge in $L^1([0,1])$-mean (hence in probability) as $|\pi_n| \to 0$, to the function $f(\cdot)$ given by:

$$\begin{aligned}\log f = {}& \frac{X(0)}{2}(m(0)-c) + \frac{X(1)}{2}(m(1)+c) - \frac{m^2(1)}{2} \\ & + \frac{1}{2}\int_0^1 X(t)(\beta m(t) - c + c\beta)\,dt - \frac{\beta}{4}\int_0^1 m^2(t)\,dt \\ & - \frac{c\beta}{2}\int_0^1 m(t)\,dt - \frac{c^2\beta}{4},\end{aligned} \tag{33}$$

where $c \geq 0$ is the (smallest) constant in the Lipschitz condition on m. Consequently, from (33) the critical region for the hypotheses is given by

$$\begin{aligned}A_0^k = \{\omega : {}& X(0,\omega)[m(0)-c] + X(1,\omega)(m(1)+c) + \\ & \int_0^1 X(t,\omega)(\beta m(t) - c + c\beta)\,dt > k\},\end{aligned} \tag{34}$$

where k is chosen to satisfy the size of the region, $P(A_0^k) = \alpha$. When $m = a$, a constant, then $c = 0$ in the above computation. If $H_1 : m = a > 0$, then (34) becomes

$$A_0^{k_1} = \{\omega : X(0,\omega) + X(1,\omega) + \beta\int_0^1 X(t,\omega)\,dt \geq k_1\}, \tag{34'}$$

for a suitable k_1, which is a (one sided) uniformly most powerful critical region (the inequality being reversed if $a < 0$).

The above procedure shows that the $\{\pi_n, n \geq 1\}$ are only partially ordered and the $f_{\pi_n} = \frac{dQ_{\pi_n}}{dP_{\pi_n}}$ converge in the mean but not pointwise. As counter examples show one has only such a weaker statement when a complete ordering is not available. To include these problems one seeks just convergence in probability for these likelihood ratios. This is covered, even with mean convergence, by employing a general technique based on Hellinger integrals which we now describe.

If P and Q are a pair of probability measures on a measurable space (Ω, Σ), and μ is a dominating (σ-finite) measure for both (specifically take $\mu = P + Q$), then the *Hellinger "distance"* between P and Q is defined by:

$$H(P,Q) = \int_\Omega \sqrt{dP\,dQ} = \int_\Omega \sqrt{\frac{dP}{d\mu}\cdot\frac{dQ}{d\mu}}\,d\mu$$

$$= (f, g) = \int_\Omega fg\, d\mu, \tag{35}$$

where f^2, g^2 are the Radon-Nikodým derivatives of P and Q respectively relative to μ. If $\tilde{\mu}$ is another such dominating measure then $\mu \ll \tilde{\mu}$ so that one has

$$\begin{aligned} H(P, Q) &= \int_\Omega fg\, d\mu = \int_\Omega fg \frac{d\mu}{d\tilde{\mu}}\, d\tilde{\mu} \\ &= \int_\Omega \sqrt{\frac{dP}{d\tilde{\mu}} \frac{dQ}{d\tilde{\mu}}}\, d\tilde{\mu} = \int_\Omega \sqrt{dP\, dQ}, \end{aligned}$$

and hence $H(P, Q)$ does not depend on the particular dominating μ or $\tilde{\mu}$. The last expression above is the *Hellinger integral*. Note that $H(P, Q)$ is not a true distance since it does not satisfy the triangle inequality. However, $0 \le H(P, Q) = H(Q, P) \le 1$, the last being a consequence of the CBS- inequality. Moreover, $H(P, Q) = 0$ iff $Q \perp P$ and $H(P, Q) = 1$ iff $P = Q$ (by the equality conditions in the CBS-inequality and since $\min(f, g) = 0$ if $P \perp Q$). But a true distance between P, Q, say $\rho(P, Q)$, is obtained from (35) by considering the $L^2(\mu)$-metric:

$$\begin{aligned} \rho(P, Q) &= \|f - g\|_{2,\mu} = \sqrt{(f, f) + (g, g) - 2(f, g)} \\ &= \sqrt{2(1 - H(P, Q))}, \end{aligned} \tag{36}$$

and this can be used to translate the Hellinger distance to the $L^2(\mu)$-metric. The expression $H(P, Q)$ will be useful in deciding the singularity or equivalence (or non-singularity) of P, Q in many problems, with (35) and (36), complementing the result of Theorem 1 in obtaining the likelihood ratio $\frac{dQ^c}{dP}$, strengthening the alternative method employed in Example 6 above.

7. Theorem. *Let $\{(\Omega, \Sigma_\alpha, \Sigma, {}^{P}_{Q}), \alpha \in I\}$ be a probability model for testing a hypothesis and its alternative, where I is a directed index set with a partial ordering denoted by '$\le$' and $\Sigma_\alpha \subset \Sigma_\beta \subset \Sigma$ for $\alpha \le \beta$ in I, the Σ_α being σ-algebras. If $P_\alpha = P|\Sigma_\alpha$, $Q_\alpha = Q|\Sigma_\alpha$ and (for simplicity) $\Sigma = \sigma(\cup_\alpha \Sigma_\alpha)$, then the corresponding Hellinger distances satisfy the limit relation $H(P, Q) = \lim_\alpha H(P_\alpha, Q_\alpha)$, so that $P \perp Q$ iff $H(P, Q) = 0$ which trivially holds when $H(P_\alpha, Q_\alpha) = 0$ for some $\alpha \in I$.*

Proof. We first establish the result, based on an elementary argument due to Brody [1], as stated, and then show how it can be reformulated when P_α, Q_α are image measures of P, Q on finite dimensional spaces such as $\mathbb{R}^\alpha \cong \mathbb{R}^n$. This will enable a direct application of the theorem.