PACIFIC JOURNAL OF MATHEMATICS
Vol. 17, No. 1, 1966

GAUSSIAN MEASURES IN FUNCTION SPACE

L. A. SHEPP

Two Gaussian measures are either mutually singular or equivalent. This dichotomy was first discovered by Feldman and Hajek (independently). We give a simple, almost formal, proof of this result, based on the study of a certain pair of functionals of the two measures. In addition we show that two Gaussian measures with zero means and smooth Polya-type covariances (on an interval) are equivalent if and only if the right-hand slopes of the covariances at zero are equal.

The H and J functionals. Two probability measures μ_0 and μ_1 on a space $(\Omega, \mathscr{B})$ are called mutually singular ($\mu_0 \perp \mu_1$) if there is a set $B \in \mathscr{B}$ for which $\mu_0(B) = 0$ and $\mu_1(\Omega - B) = 0$. The measures are called mutually equivalent ($\mu_0 \sim \mu_1$) if they have the same zero sets, i.e., $\mu_0(B) = 0$ if and only if $\mu_1(B) = 0$.

Setting $\mu = \mu_0 + \mu_1$ we may define the Radon-Nikodym derivatives,

$$X_0 = d\mu_0/d\mu\ , \qquad X_1 = d\mu_1/d\mu\ . \tag{1.1}$$

LEMMA 1. (i) *$\mu_0 \perp \mu_1$ if and only if $X_0 \cdot X_1 = 0$ a.e. (μ).*
(ii) *$\mu_0 \sim \mu_1$ if and only if $X_0 \cdot X_1 > 0$ a.e. (μ).*

In (ii) suppose $\mu_0 \sim \mu_1$. If $E = \{X_0 = 0\}$ then $\mu_0(E) = \int_E X_0 d\mu = 0$. Thus $\mu_1(E) = 0$ and also $\mu(E) = \mu_0(E) + \mu_1(E) = 0$. Similarly $\mu\{X_1 = 0\} = 0$ and so $X_0 \cdot X_1 > 0$ a.e. (μ). The proofs of the remaining assertions are as easy.

We now define the functionals H (Hellinger [6], see also [9] and [10] and J (Jeffreys [8], see also [5]).

$$H = \int \sqrt{X_0 X_1}\, d\mu \tag{1.2}$$

$$J = \int (X_0 - X_1) \log (X_0/X_1) d\mu\ . \tag{1.3}$$

The integrand of J is of course to be taken as $+\infty$ if either X_0 or X_1 is zero but not both. As such it is well defined and nonnegative a.e. (μ) and so $0 \leqq J \leqq \infty$. By Schwarz's inequality, $0 \leqq H \leqq 1$. We remark that μ could have been chosen in (1.2) and (1.3) as any measure dominating both μ_0 and μ_1.

Received October 6, 1964, and in revised form November 20, 1964.

LEMMA 2. (i) $\mu_0 \perp \mu_1$ *if and only if* $H = 0$.
(ii) $\mu_0 \sim \mu_1$ *if* $J < \infty$.

Clearly $H = 0$ if and only if $X_0 \cdot X_1 = 0$ a.e. (μ). Similarly if $J < \infty$ then $X_0 \cdot X_1 > 0$ a.e. (μ) since the integrand of J must be finite a.e. (μ). An appeal to Lemma 1 completes the proof.

It follows from the lemma that if $J < \infty$ then $H > 0$. The converse of this assertion is not true in general but is true if the measures μ_0 and μ_1 are Gaussian. This is all that remains to prove the dichotomy theorem.

We shall now construct the general Gaussian measure. This paragraph is included in order to introduce notation and follows [2, p. 72]. If I is any set we take Ω to be the set of all real functions, X, on I and $\mathscr{B}$ to be the smallest σ-field on which each coordinate $X(t)$, $t \in I$ is measurable. If ρ is any real function on $I \times I$ for which the matrix $\rho^\pi = \rho^\pi_{ij} = \rho(t_i; t_j)$, $i = 1, \cdots, n$, $j = 1, \cdots, n$, has determinant $|\rho^\pi| \geqq 0$ for each finite set $\pi = \{t_1, \cdots, t_n\} \subset I$, $n = 1, 2, \cdots$, then ρ is called nonnegative definite. Given such a function ρ and a real function m on I we may define a Gaussian measure $\mu = \mu(\rho, m)$ as follows. If $\mathscr{B}(\pi)$ is the σ-field generated by $X(t)$, $t \in \pi$, then we define μ on generating sets of $\mathscr{B}(\pi)$ by

$$\mu\{X(t_1) \leqq a_1, \cdots, X(t_n) \leqq a_n\} = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} p^\pi(x_1, \cdots, x_n) dx_1 \cdots dx_n ,$$

where p^π is the Gaussian density,

$$(1.4) \qquad p^\pi(x) = (2\pi)^{-n/2} |\rho^\pi|^{-1/2} \exp\left(-\frac{1}{2}((\rho^\pi)^{-1}(x - m^\pi), x - m^\pi\right)$$

and $m^\pi = (m(t_1), \cdots, m(t_n))$, $x = (x_1, \cdots, x_n)$. This defines μ in a consistent way on each $\mathscr{B}(\pi)$. There is a unique extension of μ to a measure on $\mathscr{B}$ which is called the Gaussian measure with covariance ρ and mean m.

Suppose now that μ_i, $i = 0, 1$, are two Gaussian measures on $(\Omega, \mathscr{B})$. We define μ^π_i to be the restriction of μ_i to $\mathscr{B}(\pi)$, $i = 0, 1$, and $X^\pi_i = d\mu^\pi_i / d\mu$, $i = 0, 1$. It is easily checked that X^π_i, $\pi \subset I_0$ is a martingale net and by a theorem of Helms [7], $X^\pi_i \to X_i$, $i = 0, 1$, in $L'(\mu)$. We denote for each $\pi \subset I$, π finite,

$$(1.5) \qquad H(\pi) = \int \sqrt{X^\pi_0 X^\pi_1}\, d\mu$$

$$(1.6) \qquad J(\pi) = \int (X^\pi_0 - X^\pi_1) \log (X^\pi_0 / X^\pi_1) d\mu .$$

Using Jensen's inequality for conditional expectations, see [13], it is not difficult to prove that for $\pi_1 \subseteqq \pi_2$, $H(\pi_1) \geqq H(\pi_2) \geqq H$; $J(\pi_1) \leqq J(\pi_2) \leqq J$. Using the martingale convergence theorem we easily obtain with π running through all (finite) subsets of I,

$$(1.7) \qquad H = \inf_{\pi \subset I} H(\pi) , \qquad J = \sup_{\pi \subset I} J(\pi) .$$

Assuming for a moment that ρ_0 and ρ_1 are strictly positive-definite, i.e., $|\rho_0^\pi| > 0$, $|\rho_1^\pi| > 0$, for all $\pi \subset I$, we may evaluate $H(\pi)$ and $J(\pi)$. Let $m^\pi = m_0^\pi - m_1^\pi$ and $\rho = (\rho_0 + \rho_1)/2$, $\rho^\pi = (\rho_0^\pi + \rho_1^\pi)/2$. Then

$$(1.8) \qquad (H(\pi))^2 = \{|\rho_0^\pi|^{1/2}|\rho_1^\pi|^{1/2}/|\rho^\pi|\} \exp\left(-\frac{1}{2}((\rho^\pi)^{-1}m^\pi, m^\pi)\right) ,$$

$$(1.9) \qquad \begin{aligned} J(\pi) &= \frac{1}{2} tr\{(\rho_1^\pi - \rho_0^\pi)((\rho_0^\pi)^{-1} - (\rho_1^\pi)^{-1})\} \\ &\quad + \frac{1}{2}(((\rho_0^\pi)^{-1} + (\rho_1^\pi)^{-1})m^\pi, m^\pi) . \end{aligned}$$

Using (1.7) we shall show that if $H > 0$ then $J < \infty$.[1] As remarked before this will complete the proof of the dichotomy theorem.

Define the quantities

$$(1.10) \qquad D(\pi) = |\rho^\pi|^2/|\rho_0^\pi||\rho_1^\pi| , \qquad E(\pi) = ((\rho^\pi)^{-1}m^\pi, m^\pi) .$$

Then $(H(\pi))^{-4} = D(\pi)\exp E(\pi)$, and looking for a moment at the case $m = 0$ we obtain $D(\pi) \geqq 1$, since $E(\pi)$ then vanishes and $H(\pi) \leqq 1$. Since $D(\pi)$ does not depend on m we must have $D(\pi) \geqq 1$ in general. Since $(\rho^\pi)^{-1}$ is positive-definite, $\exp E(\pi) \geqq 1$ also, i.e.,

$$(1.11) \qquad D(\pi) \geqq 1 , \qquad \exp E(\pi) \geqq 1 .$$

Using (1.6) this shows that if $H > 0$ there is a number M independent of π such that

$$(1.12) \qquad D(\pi) \leqq M , \qquad E(\pi) \leqq M .$$

Now define the quantities

$$(1.13) \qquad T(\pi) = tr\{(\rho_1^\pi - \rho_0^\pi)((\rho_0^\pi)^{-1} - (\rho_1^\pi)^{-1})\}$$

$$(1.14) \qquad Q(\pi) = (\{(\rho_0^\pi)^{-1} + (\rho_1^\pi)^{-1}\}m^\pi, m^\pi)$$

and not that $2J(\pi) = T(\pi) + Q(\pi)$.

If $\lambda_1 > 0, \cdots, \lambda_n > 0$, are the eigenvalues of $\rho_0^\pi(\rho_1^\pi)^{-1}$ then

$$(1.15) \qquad T(\pi) = \Sigma(\lambda_j + \lambda_j^{-1} - 2) = \Sigma(\lambda_j - 1)^2/\lambda_j$$

[1] The author benefitted here from reading an unpublished manuscript of T. Kadota.

$$D(\pi) = |\rho^\pi(\rho_1^\pi)^{-1}|\,|\rho^\pi(\rho_0^\pi)^{-1}| = \prod_j \frac{\lambda_j + 1}{2}\frac{1 + \lambda_j^{-1}}{2} \tag{1.16}$$

and so

$$T(\pi) \leqq 4 \prod_j (1 + (\lambda_j - 1)^2/4\lambda_j) = 4D(\pi) \leqq 4M\,. \tag{1.17}$$

Thus $T(\pi)$ is uniformly bounded. To show that $Q(\pi)$ is also we simultaneously diagonalize the quadratic forms $E(\pi)$ and $Q(\pi)$. Let $\mu_1, \cdots, \mu_n$ be the eigenvalues of $\rho^\pi((\rho_0^\pi)^{-1} + (\rho_1^\pi)^{-1})$ and $x_1, \cdots, x_n$ corresponding eigenvectors (which can be properly chosen in the following even if the eigenvalues are not distinct). We normalize $x_1, \cdots, x_n$ so that $((\rho^\pi)^{-1}x_k, x_k) = 1, k = 1, \cdots, n$. Then

$$E(\pi) = \Sigma h_k^2\,, \qquad Q(\pi) = \Sigma \mu_k h_k^2 \tag{1.18}$$

where $m^\pi = \Sigma h_k x_k$ defines $h_k, k = 1, \cdots, n$. The eigenvalues μ can be written in terms of the eigenvalues λ. In fact since

$$\rho^\pi((\rho_0^\pi)^{-1} + (\rho_1^\pi)^{-1}) = I + \frac{1}{2}\rho_0^\pi(\rho_1^\pi)^{-1} + \frac{1}{2}(\rho_0^\pi(\rho_1^\pi)^{-1})^{-1} \tag{1.19}$$

we have

$$\mu = 1 + \frac{1}{2}\lambda + \frac{1}{2}\lambda^{-1} = 2(1 + (\lambda - 1)^2/4\lambda)$$

generically. Thus for each $k = 1, 2, \cdots, n$,

$$\frac{1}{2}\mu_k \leqq \prod_j (1 + (\lambda_j - 1)^2/4\lambda_j) = D(\pi) \leqq M \tag{1.20}$$

and so by (1.12) and (1.18),

$$Q(\pi) = \Sigma \mu_k h_k^2 \leqq 2ME(\pi) \leqq 2M^2\,. \tag{1.21}$$

Thus $J(\pi) = (1/2)T(\pi) + (1/2)Q(\pi) \leqq 2M + M^2$ is uniformly bounded and so $J < \infty$ if $H > 0$. This completes the proof in case $|\rho_0^\pi|\,|\rho_1^\pi| > 0$ for all π.

Returning now to the case when ρ_0 and ρ_1 are not both strictly positive-definite we may argue as follows. If there is a finite set π for which exactly one of $|\rho_0^\pi|$ and $|\rho_1^\pi|$ vanishes then $\mu_0 \perp \mu_1$ in a trivial way. In the opposite case we can choose a maximal set $I_0 \subset I$ with the property that ρ_0 and ρ_1 are strictly positive-definite on $I_0 \times I_0$. The proof above shows that either $\mu_0 \perp \mu_1$ or $\mu_0 \sim \mu_1$ relative to $\mathscr{B}(I_0)$. It is easy to show, however, that singularity or equivalence relative to $\mathscr{B}(I_0)$ is the same as that relative to $\mathscr{B} = \mathscr{B}(I)$. We have therefore proved the following theorem.

THEOREM. *If ρ_0 and ρ_1 are real-valued and nonnegative definite on $I \times I$ and m_0 and m_1 are real-valued functions on I then the Gaussian measures $\mu_0 = \mu_0(\rho_0, m_0)$ and $\mu_1 = \mu_1(\rho_1, m_1)$ are either mutually equivalent or mutually singular. They are mutually equivalent or mutually singular according as*

$$H > 0 \quad or \quad H = 0 \tag{1.22}$$

or equivalently according as

$$J < \infty \quad or \quad J = \infty\ , \tag{1.23}$$

where H and J are given by (1.5)–(1.7).

The criterion (1.23) was obtained by Hajek and (1.22) by Rao and Varadarajan and independently by the author. There are two special cases of (1.22) which deserve mention. If the mean m_0 and m_1 are zero then $H(\pi)$ does not involve inverting matrices. If $\rho_0 = \rho_1$ then the problem has been studied and solved by Grenander [4]. It is a corollary of (1.22) that these special cases include the general case as the following theorem of Rao and Varadarajan shows.

THEOREM. *Denoting $\rho = (\rho_0 + \rho_1)/2$, then*

$$\mu(\rho_0, m_0) \sim \mu(\rho_1, m_1) \tag{1.24}$$

if and only if

$$\mu(\rho_0, 0) \sim \mu(\rho_1, 0) \quad and \quad \mu(\rho, m_0) \sim \mu(\rho, m_1)\ . \tag{1.25}$$

To prove this we write $H = \inf H(\pi)$ where

$$H(\pi) = (D(\pi))^{-1/4}(\exp E(\pi))^{-1/4}\ . \tag{1.26}$$

Now $D(\pi) \geqq 1$ and $\exp E(\pi) \geqq 1$ by (1.11) and so $H > 0$ if and only if

$$\inf (D(\pi))^{-1/4} > 0 \quad and \quad \inf (\exp E(\pi))^{-1/4} > 0. \tag{1.27}$$

But $\inf D(\pi)^{-1/4} > 0$ if and only if $\mu(\rho_0, 0) \sim \mu(\rho_1, 0)$ since in the case $m = 0$, $H(\pi)$ is precisely $(D(\pi))^{-1/4}$ (since $D(\pi)$ does not depend on m in (1.8)). Similarly, $\inf (\exp E(\pi))^{-1/4} > 0$ if and only if $\mu(\rho, m_0) \sim \mu(\rho, m_1)$ since in the case $\rho_0 = \rho_1 = \rho$, $H(\pi)$ becomes $(\exp E(\pi))^{-1/4}$. Using the theorem and (1.22) in these two special cases we obtain (1.24).

The referee observed that we may take $\rho = \rho_0$ in the theorem. This simplifying observation follows easily from the fact that $\sim$ is an equivalence relation.

2. Some examples. We recall the Polya class, P, of positive-definite functions on $I \times I$, where I is the unit interval. A continuous function ρ is in P if $\rho(t; s)$ is a function only of $|t - s|$ (also denoted by ρ) and

(1) $\rho = \rho(t)$ is convex, $0 \leqq t \leqq 1$.

(2) $0 \leqq \rho(s) \leqq \rho(t), 0 \leqq t \leqq s \leqq 1$.

If $\rho_i \in P, i = 0, 1$, we may ask whether $\mu_0 = \mu(\rho_0, 0)$ and $\mu_1 = \mu(\rho_1, 0)$ are singular or equivalent. This question is partially answered by the following theorem.

THEOREM. *If* $\mu_i = \mu(\rho_i, 0)$ *where* $\rho_i \in P$, $i = 0, 1$, *and if in addition* ρ_i *has a bounded second derivative and is positive, then*

$$\mu_0 \sim \mu_1 \quad \textit{if and only if} \quad \rho_0'(0^+) = \rho_1'(0^+)\ . \tag{2.1}$$

REMARK. The theorem becomes false if one drops the assumption of bounded second derivatives as the case

$$\rho_0(u) = \begin{cases} 1 - u & 0 \leqq u \leqq \dfrac{1}{2} \\ (2 - u)/3 & \dfrac{1}{2} \leqq u \leqq 1 \end{cases}, \quad \rho_1(u) = e^{-u}$$

shows. Here one can detect the presence of the jump in the derivative of ρ_0 by techniques similar to those used by Baxter [1] for jumps at zero. Here $\mu_0 \perp \mu_1$ although $\rho_0'(0^+) = \rho_1'(0^+)$ and $\rho_i \in P, i = 0, 1$. It should be possible to eliminate the positivity condition, however.

Using the continuity of ρ_i it can be shown that it is enough to consider equi-spaced partitions π in evaluating $H(\pi)$. The determinants involved are then estimated by using the following theorem of H. O. Pollak and the author [11].

THEOREM. *If* $\rho \in P$ *has a uniformly bounded second derivative and is positive then*

$$|\rho^{\pi_n}| \sim \sim (2\,|\rho'(0^+)|/n)^n\ .$$

Here $\pi_n = \{0, 1/n, 2/n, \cdots, 1\}$ is the regular partition of I into n intervals. As usual $a_n \sim \sim b_n$ if $a_n = 0(b_n)$ and $b_n = 0(a_n)$.

One can prove more general theorems using these techniques but for simplicity of statement we have presented the results in this way. It is hoped that further techniques will be developed for estimating the determinants $|\rho^{\pi_n}|$ which will settle the dichotomy question in more general cases.

REFERENCES

1. G. Baxter, *A strong limit theorem for Gaussian processes*, Proc. Amer. Math. Soc. **7**, No. 3, (1956), 522-528.
2. J. L. Doob, *Stochastic Processes*, Dover (1953).
3. J. Feldman, *Equivalence and perpendicularity of Gaussian processes*, Pacific J. Math. **8**, No. 4, (1958), 699-708.
4. U. Grenander, *Stochastic processes and statistical inference*, Ark. Mat. **1**, No. 17, (1950), 195-277.
5. J. Hajek, *A property of J-divergences of marginal probability distributions*, Cz. Math. J. **8** (83), (1958), 460-462.
6. E. Hellinger, *Die Orthogonalinvarianten quadratischer Formen von unendlich vielen Variablen*, Dissertation, Gottingen, 1907.
7. L. L. Helms, *Mean Convergence of Martingales*, Trans. Amer. Math. Soc. **87**, (1958), 439-446.
8. H. Jeffreys, *Theory of Probability*, Oxford (1948).
9. S. Kakutani, *On equivalence of infinite product measures*, Ann. Math. **49**, (1948), 214-224.
10. C. Kraft, *Some Conditions for Consistency and Uniform Consistency of Statistical Procedures*, Univ. of Cal. publ. in Stat., **2**, No. 6, (1955), 125-142.
11. H. O. Pollak, and L. A. Shepp, *The estimation of some determinants of Toeplitz-type*, Proc. Amer. Soc. To appear.
12. C. R. Rao, and V. S. Varadarajan, *Discrimination of Gaussian processes*, Sankhya, Series A, **25**, (1963), 303-330.
13. L. A. Shepp, *The singularity of Gaussian Measures in function space*, Proc. Nat. Acad. of Sci. **52**, (1964), 430-433.
14. A. M. Yaglom, *On the equivalence and perpendicularity of two Gaussian probability measures in function space*, Proc. Sym. on Time Series Analysis, Brown University, 1962, ed. Rosenblatt, Wiley (1963).

BELL TELEPHONE LABORATORIES, INC.
MURRAY HILL, NEW JERSEY