DISTINGUISHING A SEQUENCE OF RANDOM VARIABLES FROM A TRANSLATE OF ITSELF

By L. A. Shepp

Bell Telephone Laboratories, Murray Hill, New Jersey

1. Introduction. Suppose $X = \{X_1, X_2, \cdots\}$ is a sequence of independent and identically distributed random variables and $a = \{a_1, a_2, \cdots\}$ is a numerical sequence, a_n representing the error in centering X_n. When are the sample paths of X and $X + a$ distinguishable?

We can distinguish X and $X + a$ with probability one if a is so big that $\sum a_n^2 = \infty$. If $\sum a_n^2 < \infty$ and X has finite information (see Equation (1)) then we cannot distinguish. Conversely if we cannot distinguish for all a with $\sum a_n^2 < \infty$ then X has finite information. For X with finite information we can distinguish if and only if $\sum a_n^2 = \infty$. The latter statement becomes false for any wider class than the finite information class.

Here X is said to have finite information ($I < \infty$) if the common distribution F has a positive (a.e.) and (locally) absolutely continuous density φ and

$$\int_{-\infty}^{\infty} (\varphi')^2/\varphi < \infty. \tag{1}$$

Fisher [2] called the quantity in (1) the information, or intrinsic accuracy. It is denoted by $I = I(F)$.

We briefly mention an application to a quantization problem. Following J. Feldman [1] one can produce examples of quasi-invariant (qi) distributions in l_2 : Construct a product measure λ on sequence space whose translates λ_a for $a \in l_2$ are all equivalent measures. Such a predistribution λ gives rise to a qi distribution on l_2 [1]. Theorem 1 gives, in particular, the exact class of such λ having identical one-dimensional marginals—namely those with finite information. Part of this result was obtained by Feldman who was concerned with more general situations. J. R. Klauder and J. McKenna have recently obtained very general classes of qi distributions in their work on continuous representations of l_2 [4].

Returning to the statistical setting, we say we can distinguish X and $X + a$ if there is a set E of sequences for which

$$P\{\Omega - E \mid X\} = P\{E \mid X + a\} = 0,$$

where Ω is the set of all sequences. This means that the measures μ and μ^a induced by X and $X + a$ respectively,

$$\mu(A) = P\{A \mid X\}, \qquad \mu^a(A) = P\{A \mid X + a\},$$

are singular ($\mu \perp \mu^a$).

Received 10 March 1965.

The measure μ is the product measure

$$\mu = \prod_{n=1}^{\infty} F\{dx_n\} \tag{2}$$

where F is the common distribution of $X_1, X_2, \cdots$. The measure μ^a is also a product measure

$$\mu^a = \prod_{n=1}^{\infty} F\{d(x_n - a_n)\}. \tag{3}$$

The measures μ and μ^a are equivalent ($\mu \sim \mu^a$) if they have the same null sets,

$$\mu(A) = 0 \quad \text{if and only if} \quad \mu^a(A) = 0 \quad \text{for all } A. \tag{4}$$

We call X and $X + a$ *totally* indistinguishable if (4) holds. This is stronger than merely being indistinguishable (nonsingular) and, roughly stated, means that for every observed sequence there is doubt as to whether it came from X or $X + a$.

Our results may now be formulated:

THEOREM 1. *Suppose $F = F\{dx\}$ is a probability distribution on R (real numbers), and μ and μ^a are the product measures defined in* (2) *and* (3) *above*:

(i) *If $\sum a_n^2 = \infty$ then $\mu \perp \mu^a$ (distinguishable).*

(ii) *Assume $I < \infty$: then $\mu \sim \mu^a$ (totally indistinguishable) if $\sum a_n^2 < \infty$, and $\mu \perp \mu^a$ if $\sum a_n^2 = \infty$.*

Moreover a converse to (ii) *holds:*

(iii) *If $\mu \sim \mu^a$ for all a with $\sum a_n^2 < \infty$ then $I < \infty$.*

(ii) was previously known for Gaussian F and shows that the Gaussian situation continues to hold for any F with finite information. It was unexpected that the Fisher information plays such a central role and it might be interesting to determine the class of F for which $\mu \sim \mu^a$ for $a \varepsilon l_p$ for values of $p \neq 2$. The proof of Theorem 1 leans heavily on an important result of S. Kakutani on equivalence of infinite product measures and on the machinery of the Hilbert space L_2 of square-integrable functions.

2. Singularity and equivalence of product measures. S. Kakutani [3] gave useful criteria for determining singularity or equivalence of infinite product measures. Suppose $\mu_1, \mu_2, \cdots, \nu_1, \nu_2, \cdots$ are probability measures and

$$\mu = \prod_{n=1}^{\infty} \mu_n, \qquad \nu = \prod_{n=1}^{\infty} \nu_n \tag{5}$$

are their product measures. Let H denote the Hellinger integral (9).

THEOREM 2. (Kakutani). *If μ and ν are given by* (5), *then*

$$H(\mu, \nu) = \prod_{n=1}^{\infty} H(\mu_n, \nu_n). \tag{6}$$

Moreover,

$$\mu \perp \nu \quad \textit{if and only if} \quad H(\mu, \nu) = 0. \tag{7}$$

If $\mu_n \sim \nu_n$, $n = 1, 2, \cdots$, then

$$\mu \sim \nu \quad \textit{if and only if} \quad H(\mu, \nu) > 0. \tag{8}$$

We recall that for any probability measures μ and ν the Hellinger integral $H(\mu, \nu)$ is defined by choosing any measure m for which $d\mu/d_m$, $d\nu/d_m$ are defined (for example $m = \mu + \nu$) and setting

$$H(\mu, \nu) = \int (d\mu/dm \cdot d\nu/dm)^{\frac{1}{2}}\, dm. \tag{9}$$

It is easy to check that $H(\mu, \nu)$ does not depend on m and $0 \leqq H(\mu, \nu) \leqq 1$ by Schwarz's inequality. Equation (6) allows us to calculate $H(\mu, \nu)$ from the component measures.

3. A separating sequence. To prove (i) we need to show that for every F, X and $X + a$ are distinguishable if $\sum a_n^2 = \infty$. We shall choose a sequence of numbers $\theta_1, \theta_2, \cdots$ which, in a sense, separates X and $X + a$. This will entail making $|F(\theta_n) - F(\theta_n - a_n)|$ large, where F is now used to denote the common distribution function.

LEMMA 1. *Suppose F is any distribution function and $\sum a_n^2 = \infty$. There is a sequence $\theta_1, \theta_2, \cdots$ for which*

$$\sum (F(\theta_n) - F(\theta_n - a_n))^2 = \infty. \tag{10}$$

Assuming the lemma for a moment let us define new sequences X' and $(X + a)'$. Given any sequence Y set

$$\begin{aligned} Y_n' &= 0 \quad \text{if } Y_n < \theta_n \\ &= 1 \quad \text{if } Y_n \geqq \theta_n . \end{aligned}$$

Suppose α_n (resp. β_n) is the measure induced on the two point space $\Omega = \{0, 1\}$ by X_n'(resp. $(X + a)_n'$),

$$\begin{aligned} \alpha_n(0) &= F(\theta_n) & \alpha_n(1) &= 1 - F(\theta_n) \\ \beta_n(0) &= F(\theta_n - a_n), & \beta_n(1) &= 1 - F(\theta_n - a_n). \end{aligned} \tag{11}$$

We have

$$H(\alpha_n, \beta_n) = [\alpha_n(0)\beta_n(0)]^{\frac{1}{2}} + [\alpha_n(1)\beta_n(1)]^{\frac{1}{2}}.$$

Using the elementary inequality

$$(xy)^{\frac{1}{2}} + [(1 - x)(1 - y)]^{\frac{1}{2}} \leqq [1 - (x - y)^2]^{\frac{1}{2}}, \quad 0 \leqq x \leqq 1,\ 0 \leqq y \leqq 1$$

with $x = \alpha_n(0)$, $y = \beta_n(0)$ we obtain using (10) and (11) that

$$\textstyle\prod_{n=1}^{\infty} H^2(\alpha_n, \beta_n) \leqq \prod_{n=1}^{\infty} (1 - (F(\theta_n) - F(\theta_n - a_n)^2) = 0.$$

It follows that $\prod_{n=1}^{\infty} \alpha_n$ and $\prod_{n=1}^{\infty} \beta_n$ are singular and so X' and $(X + a)'$ are distinguishable. It is clear that the original process X and $X + a$ must also be distinguishable. The proof of Lemma 1 rests on a second lemma.

LEMMA 2. *Suppose F is any distribution function. Define*

$$\delta = \inf\nolimits_{0<a<1} \sup\nolimits_{-\infty<\theta<\infty} [F(\theta) - F(\theta - a)]/a.$$

Then $\delta > 0$.

Suppose the lemma false so $\delta = 0$. Then for each $\eta > 0$ there is an a, $0 < a < 1$, for which

$$(12) \qquad F(\theta) - F(\theta - a) < a\eta \qquad \text{for all } \theta.$$

Choose numbers b and c with $F(c) - F(b) > 0$ and define the integer r by

$$(13) \qquad b + ra \leqq c < b + (r + 1)a$$

where a is such that (12) holds with $\eta = (F(c) - F(b))/(c - b + 1)$. We have

$$(14) \quad F(c) - F(b) \leqq F(b + (r + 1)a) - F(b) = \sum_{k=0}^{r} F(r + (k + 1)a) - F(r + ka).$$

Using (12) in (14) we get

$$(15) \qquad F(c) - F(b) < (r + 1)a\eta.$$

Now by (13), $(r + 1)a\eta = ra\eta + a\eta < (c - b + 1)\eta = F(c) - F(b)$, contradicting (15). This proves Lemma 2.

Now Lemma 1 is trivial if a_n does not tend to zero. Suppose then that $a_n \to 0$. This means that $|a_n| < 1$ eventually and Lemma 2 then shows that numbers θ_n exist so that $|F(\theta_n) - F(\theta_n - a_n)| \geqq (\delta/2)|a_n|$, for n sufficiently large. Since $\sum a_n^2 = \infty$ we see that the θ_n satisfy (10). This completes the proof of both Lemma 1 and (i) in Theorem 1.

4. The Fisher information. The information I can be written simply as

$$(16) \qquad I = 4\int (h')^2, \qquad h = \varphi^{\frac{1}{2}}.$$

Now φ is a density and so $h \,\varepsilon\, L_2$. Let $\hat{h}(u) = (2\pi)^{-\frac{1}{2}}\int e^{ixu}h(x)\,dx$ denote the transform of h. The formal transform of h' is $iu\hat{h}(u)$ and by Plancherel's theorem

$$(17) \qquad \int (h')^2 = \int u^2|\hat{h}(u)|^2\,du.$$

The correspondence between h' and $iu\hat{h}(u)$ is made precise by the following well-known lemma ([5], p. 92].

LEMMA 3. *Suppose $h \,\varepsilon\, L_2$. If $h' \,\varepsilon\, L_2$ then $u\hat{h}(u) \,\varepsilon\, L_2$. If $u\hat{h}(u) \,\varepsilon\, L_2$ then h is (almost everywhere) an absolutely continuous function and $h' \,\varepsilon\, L_2$. Moreover in this case (17) holds.*

Now suppose that $I < \infty$. By definition F then has a positive density and so the measures

$$\mu_n\{dx\} = F\{dx\}, \qquad \nu_n\{dx\} = F\{d(x - a_n)\}$$

are equivalent. We have

$$d\mu_n/dx = \varphi(x), \qquad d\nu_n/dx = \varphi(x - a_n)$$

and applying (9) with m = Lebesgue measure dx we obtain

$$(18) \qquad H(\mu_n, \nu_n) = \int [\varphi(x)\varphi(x - a_n)]^{\frac{1}{2}}\,dx.$$

Parseval's identity gives

$$\int [\varphi(x)\varphi(x - a_n)]^{\frac{1}{2}}\, dx = \int \cos a_n u |\hat{h}(u)|^2\, du$$

and from (6) we obtain

$$H(\mu, \mu^a) = \prod_{n=1}^{\infty} \left(1 - \int (1 - \cos a_n u)|\hat{h}(u)|^2\, du\right). \tag{19}$$

Using the inequality $1 - \cos t \leqq \frac{1}{2}t^2$ we obtain from (16) and (17)

$$\int (1 - \cos a_n u)|\hat{h}(u)|^2\, du \leqq \tfrac{1}{2}a_n^2 \int u^2 |\hat{h}(u)|^2\, du = \tfrac{1}{8}a_n^2 I.$$

It follows immediately that $H(\mu, \mu^a) > 0$ if $\sum a_n^2 < \infty$ and $I < \infty$. Using (8) this proves (ii). Feldman has already obtained this ([1], pp. 348–349).

Next we prove (iii). We are given that $\mu \sim \mu^a$ for all $a \varepsilon l_2$. In the next section we will show that F has a positive density. Assuming this, we have that $\mu_n \sim \nu_n$, $n = 1, 2, \cdots$. Using (8) we must have $H(\mu, \mu^a) > 0$ for all $a \varepsilon l_2$. Now (19) gives

$$\sum_{n=1}^{\infty} \int (1 - \cos a_n u)|\hat{h}(u)|^2\, du < \infty, \qquad \text{for all } a \varepsilon l_2.$$

Now $1 - \cos t \geqq \frac{1}{4}t^2$ for $|t| \leqq 1$ and so

$$\sum_{n=1}^{\infty} a_n^2 \int_{|u| \leqq |a_n|^{-1}} u^2 |\hat{h}(u)|^2\, du < \infty. \tag{20}$$

We next show $u\hat{h}(u) \varepsilon L_2$. Suppose this were false. Then

$$T(a) = \int_{|u| \leqq 1/a} u^2 |\hat{h}(u)|^2\, du$$

increases to infinity as a tends to zero.

Lemma 4. *If T is a function such that $T(a)$ increases to infinity as a decreases to zero then there is a sequence $a \varepsilon l_2$ for which $\sum a_n^2 T(a_n) = \infty$.*

Choose numbers t_n such that $0 < t_n \leqq (\frac{1}{2})^n$, $T(t_n) \geqq 2^n$, $n = 1, 2, \cdots$. Now choose a so that there are exactly r_n values of k for which $a_k = t_n$ where r_n is the unique integer satisfying

$$r_n 2^n t_n^2 < 1 \leqq (r_n + 1) 2^n t_n^2. \tag{21}$$

We then obtain $\sum a_n^2 = \sum r_n t_n^2 < \sum 2^{-n} = 1$ so $a \varepsilon l_2$. However, $\sum a_n^2 T(a_n) \geqq \sum (r_n + 1) t_n^2 2^n - \sum t_n^2 2^n = \infty$. This proves Lemma 4. Using a sequence a with the properties of the lemma we obtain a contradiction to (20). Thus $u\hat{h}(u) \varepsilon L_2$. Using Lemma 3 we see that $h = \varphi^{\frac{1}{2}}$ is absolutely continuous. It then follows, although the proof is not completely trivial, that φ itself is absolutely continuous. Using (16) and (17), $I < \infty$.

5. Translates of a linear measure. We are given that $\mu \sim \mu^a$ for $a \varepsilon l_2$ and we will show here that this means that F has a positive density. We must have $F\{dx\} \sim F\{d(x - a)\}$ for all a because marginals of equivalent product measures are necessarily equivalent. That is, F is equivalent to its translates. The fact that F has a positive density is already a consequence of this as the following lemma shows.

Lemma 5. *If F is a probability measure on the reals (R) which is equivalent to its translates then F has a positive density.*

For any set E, Fubini's theorem gives

$$\int_{a\varepsilon R} \left(\int_{x\varepsilon R} \chi_E(x - a)F\{dx\}\right) da = \int_{x\varepsilon R} \left(\int_{a\varepsilon R} \chi_E(x - a)\, da\right)F\{dx\}$$

where χ_E is the indicator of E, $\chi_E(u) = 1$ or 0 according as $u \varepsilon E$ or not. This is exactly

$$\int_{a\varepsilon R} F\{E + a\}\, da = \int_{x\varepsilon R} \lambda\{E\}F\{dx\} = \lambda\{E\}F\{R\}. \tag{22}$$

Now if $F\{E\} = 0$ then $\lambda\{E\} = 0$ because $F\{E + a\} = 0$ for all a. Conversely, if $\lambda\{E\} = 0$, $F\{E + a\} = 0$ for almost every a by (22). It follows that $F\{E + a\} = 0$ for every a, since F is equivalent to its translates. In particular $F\{E + a\} = 0$ for $a = 0$. We have proved that F is equivalent to Lebesgue measure. This means that F has a positive density φ and the lemma is proved.

We have already seen that φ is absolutely continuous and that (1) holds. This proves (iii) and finishes the proof of Theorem 1.

REFERENCES

[1] Feldman, J. (1961). Examples of non-Gaussian quasi-invariant distributions in Hilbert space. *Trans. Amer. Math. Soc.* **99** 342–349.

[2] Fisher, R. A. (1965). *Statistical Methods and Scientific Inference.* Oliver and Boyd, Edinburgh.

[3] Kakutani, S. (1948). On equivalence of infinite product measures. *Ann. Math.* **49** 214–224.

[4] Klauder, J. R. and McKenna, J. (1965). Continuous-representation theory. *J. Math. Phys.* **6** 68–87.

[5] Titchmarsh, E. C. (1937). *Introduction to the Theory of Fourier Integrals.* Oxford Univ. Press.