

## A MEASURE WHICH IS SINGULAR AND UNIFORMLY LOCALLY UNIFORM

DAVID FREEDMAN AND JIM PITMAN

(Communicated by William D. Sudderth)

**ABSTRACT.** An example is given of a singular measure on  $[0,1]$  which is locally nearly uniform in the weak star topology. If this measure is used as a prior to estimate an unknown probability in coin tossing, the posterior is asymptotically normal.

### INTRODUCTION

Let  $\mu$  be a probability on  $[0,1]$ . If  $I$  is a subinterval of  $[0, 1]$ , let  $\mu_I$  be the probability on  $[0, 1]$  obtained from  $\mu$  as follows: restrict  $\mu$  to  $I$ ; renormalize so the mass is 1; map  $I$  affinely onto  $[0, 1]$ , preserving the order. The image of the restricted and renormalized measure is  $\mu_I$ . To illustrate the notation, if  $I = [a, b]$  then

$$\mu_I[0, y] = \frac{\mu[a, a + y(b - a)]}{\mu[a, b]}, \quad \text{for } y \in [0, 1].$$

Say that  $\mu$  is *locally uniform at  $x$*  when  $x \in I$  and  $|I| \rightarrow 0$  imply that  $\mu_I$  converges to Lebesgue measure, in the weak star topology. Here  $|I|$  is the length of  $I$ . In case the convergence of  $\mu_I$  to Lebesgue measure as  $|I| \rightarrow 0$  holds uniformly over all subintervals  $I$ , we call  $\mu$  *uniformly locally uniform*. Using Kolmogorov's distance between probability distributions, this property can be expressed as follows:

$$(1.1) \quad \lim_{\delta \rightarrow 0} \sup_{|I| \leq \delta} \sup_{0 \leq y \leq 1} |\mu_I[0, y] - y| = 0.$$

A uniformly locally uniform measure is obviously continuous.

If  $\mu$  is absolutely continuous with density  $f$ , then  $f(x) > 0$  and  $x$  is a Lebesgue point of  $f$  for  $\mu$ -almost all  $x$ ; and  $\mu$  is locally uniform at such  $x$ . See, for example, Dunford and Schwartz [1, pp. 217–218] or Saks [2, Chap. IV]. It is natural to conjecture the converse: if  $\mu$  is locally uniform at  $\mu$ -almost

Received by the editors July 18, 1988 and, in revised form, November 21, 1988.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 28A99, 42A55, 62F15.

*Key words and phrases.* Differentiation, Lebesgue points, Bayes estimates, Riesz product, singular measure, locally uniform measure, asymptotic normality of posterior distribution.

Research of the first author was partially supported by NSF grant DMS 86-01634.

Research of the second author was partially supported by NSF grant DMS 88-01808.

all  $x$ , it must be absolutely continuous. This is true, and easily proved, for local uniformity defined by convergence in variation norm. But for weak star convergence, this converse is false.

**1.2. Theorem.** *There exists a singular measure which is uniformly locally uniform.*

Consider using a measure  $\mu$  as a prior to estimate the unknown probability  $p$  that a coin lands heads in a sequence of tosses. Let  $\hat{p}$  be the fraction of heads among the first  $n$  tosses. As shown by Laplace [3], Bernstein [4], and von Mises [5], if the prior has a smooth density, the posterior is asymptotically normal. It may be conjectured that the converse holds: asymptotic normality of the posterior entails smoothness of the prior. But this too is false.

**1.3. Theorem.** *Suppose the prior  $\mu$  is uniformly locally uniform. For every  $\varepsilon > 0$ , the posterior distribution of  $p \rightarrow \sqrt{n}/(\hat{p}(1-\hat{p}))(p-\hat{p})$  converges weak star to standard normal as  $n \rightarrow \infty$ , uniformly in  $\hat{p} \in [\varepsilon, 1-\varepsilon]$ .*

If  $0 < p_0 < 1$ , and the data are generated by tossing a  $p_0$ -coin, then  $\hat{p} \rightarrow p_0$  almost surely, so the posterior is asymptotically normal almost surely. And the theorem applies to priors which may be singular, absolutely continuous, or mixed, provided they are uniformly locally uniform. The argument also extends to give a condition on  $\mu$  both necessary and sufficient for the conclusion of the theorem: for every  $\varepsilon > 0$ ,  $\mu$  is uniformly locally uniform within  $[\varepsilon, 1-\varepsilon]$ . But we omit the details. There is extensive literature on convergence of posterior distributions to normality in the stronger sense of total variation distance. See Le Cam [6 and 7] for a survey.

The balance of this paper is organized as follows: §2 gives some reformulations of uniform local uniformity; §3 presents the construction for Theorem 1.2; §4 proves Theorem 1.3; §5 provides some background on Riesz products; finally, history and acknowledgements in §6.

## 2. A CRITERION FOR UNIFORM LOCAL UNIFORMITY

We start this section with an elementary estimate. This just shows that if  $\mu$  puts approximately equal masses on equally spaced intervals, then  $\mu$  will be close to uniform in Kolmogorov distance. The elementary proof is omitted.

**2.1. Lemma.** *Let  $\mu$  be a probability on  $[0,1]$ . Let  $0 \leq a < 1$ ,  $0 < b < 1$ . Let  $N$  be a positive integer, with  $a + Nb \leq 1 < a + (N+1)b$ , so the  $N+1$  points*

$$a, \quad a+b, \quad a+2b, \quad \dots, \quad a+Nb$$

*partition  $[0,1]$  into  $N$  subintervals of equal length  $b$ , and two shorter (and possibly degenerate) end intervals  $[0, a)$  and  $[a + Nb, 1]$ . Suppose that for every one of these  $N+2$  intervals  $I'$ , and every one of the  $N$  intervals  $I$  of length  $b$ ,*

$$\frac{\mu(I')}{\mu(I)} \leq 1 + \varepsilon.$$

Then

$$\sup_{0 \leq y \leq 1} |\mu[0, y] - y| \leq \frac{4}{N}(1 + \varepsilon) + \varepsilon.$$

We now give a criterion for  $\mu$  to be uniformly locally uniform, in terms of the  $\mu$ -measure of adjacent binary intervals of order  $k$ , where  $k$  is a positive integer. There are pairs of intervals of the form

$$I = [j/2^k, (j + 1)/2^k), \quad I' = [(j + 1)/2^k, (j + 2)/2^k),$$

for  $j = 0, 1, \dots, s^k - 2$  (except that  $I'$  should include 1 if that is its right end point).

**2.2. Proposition.** *A measure  $\mu$  on  $[0, 1]$  is uniformly locally uniform if and only if*

$$(2.3) \quad \mu(I')/\mu(I) \rightarrow 1 \quad \text{as } k \rightarrow \infty$$

*uniformly over all pairs  $I$  and  $I'$  of adjacent binary intervals of order  $k$ .*

*Proof.* “Only if” is clear. For the converse, suppose (2.3) holds. Then (2.3) must also hold uniformly over all pairs  $I$  and  $I'$  of binary intervals of order  $k$  that are within distance  $A2^{-k}$  of each other, for each fixed  $A > 0$ . If  $J$  is any interval with  $2^{-i} \leq |J| < 2^{-i+1}$ , then  $J$  contains at least  $2^j - 2$  consecutive binary intervals of order  $k = i + j$ , and is contained in a union of at most  $2^{j+1} + 2$  such intervals. Now a routine argument using Lemma 2.1 shows  $\mu_J$  is close to uniform in Kolmogorov distance if  $|J|$  is sufficiently small.  $\square$

We conclude this section by stating two further conditions on a probability measure  $\mu$  on  $[0, 1]$ , each implied by uniform local uniformity. These will be used in §4. Their proofs are elementary and omitted.

**2.4. Condition.** *Let  $J$  and  $J'$  be two subintervals of an interval  $I \subset [0, 1]$ . For each  $\delta > 0$  and  $L < \infty$ ,*

$$\frac{\mu(J')/\mu(J)}{|J'|/|J|} \rightarrow 1 \quad \text{as } |I| \rightarrow 0,$$

*uniformly in  $J, J', I$ , provided*

$$\delta \leq |J'|/|J| \leq L.$$

**2.5. Condition.** *Let  $\mathbf{C}$  be a collection of uniformly bounded and uniformly equicontinuous functions on  $[0, 1]$ . For a subinterval  $I$  of  $[0, 1]$ , let  $A_I$  be the order preserving affine mapping of  $I$  onto  $[0, 1]$ . Then*

$$\frac{1}{\mu(I)} \int_I f \circ A_I d\mu \rightarrow \int_0^1 f(x) dx$$

*as  $|I| \rightarrow 0$ , uniformly in  $I$  and  $f \in \mathbf{C}$ .*

Condition 2.4 is in fact equivalent to uniform local uniformity of  $\mu$ . So is 2.5 for sufficiently rich  $C$ .

### 3. PROOF OF THEOREM 1.2

Let

$$(3.1) \quad f_N(x) = \prod_1^N [1 + a_n \cos(2\pi\lambda_n x)],$$

where  $0 \leq a_n \leq 1$ , and the  $\lambda_n$  are positive integers with integer ratios  $\lambda_{n+1}/\lambda_n \geq 3$ . Then  $f_N$  is a probability density function. A probability measure  $\mu$  on  $[0, 1]$ , called a *Riesz product*, can be defined as the weak star limit as  $N \rightarrow \infty$  of the probability with density  $f_N$ . Informally,

$$(3.2) \quad \mu = \prod_1^{\infty} [1 + a_n \cos(2\pi\lambda_n x)].$$

The existence of such Riesz product measures is shown by Fourier analysis. According to a result of Zygmund [8; or 9, p. 209], the Riesz product  $\mu$  is either absolutely continuous or singular with respect to Lebesgue measure, according to the convergence or divergence of  $\sum a_n^2$ . In either case,  $\mu$  is continuous. For the sake of completeness, in §5 we sketch arguments for the facts we use.

Because the assumptions imply that  $\lambda_1$  divides  $\lambda_n$  for every  $n$ ,

$$(3.3) \quad \mu \text{ has period } 1/\lambda_1.$$

That is, translating  $\mu$  by  $1/\lambda_1(\text{ mod } 1)$  leaves  $\mu$  invariant. In particular,

$$(3.4) \quad \mu[j/\lambda_1, (j+1)/\lambda_1) = 1/\lambda_1.$$

**3.5. Lemma.** Fix  $N$ . Suppose  $I$  and  $I'$  are intervals of the form

$$[j/\lambda_n, (j+1)/\lambda_n), \quad [j'/\lambda_n, (j'+1)/\lambda_n),$$

with  $n \leq N+1$  and  $1 \leq j < j' < \lambda_n$ . If

$$(3.6) \quad b < f_N(x')/f_N(x) < c \quad \text{for all } x \text{ in } I \text{ and } x' \text{ in } I',$$

then

$$(3.7) \quad b < \mu(I')/\mu(I) < c.$$

*Proof.* Let

$$\mu_N = \prod_{N+1}^{\infty} [1 + a_n \cos(2\pi\lambda_n x)].$$

Then  $\mu$  has density  $f_N$  with respect to  $\mu_N$ , so

$$\mu(J) = \int_J f_N d\mu_N \quad \text{for } J = I \text{ or } I'.$$

Applying (3.4) to  $\mu_N$  with  $\lambda_{N+1}$  in place of  $\lambda_1$  shows that  $\mu_N(I) = \mu_N(I')$ , because  $\lambda_n$  divides  $\lambda_{N+1}$ . So (3.6) implies (3.7).  $\square$

Fix now a sequence  $a_n$  with  $1/5 \geq a_n \rightarrow 0$  and  $\sum a_n^2 = \infty$ , say  $a_n = 1/(5\sqrt{n})$ . This makes the Riesz product  $\mu$  continuous and singular. If the  $\lambda_n$  increase rapidly enough, we can make  $\mu$  uniformly locally uniform. The  $\lambda_n$  will be of the form

$$(3.8) \quad \lambda_n = 2^{k(n)},$$

where  $k(1) < k(2) < \dots$  will be defined inductively. Let  $\mu$  be the resulting Riesz product. The inductive definition of  $k(n)$  will secure that (3.9) implies (3.10):

$$(3.9) \quad I \text{ and } I' \text{ are two adjacent binary intervals of order } k \text{ with} \\ k(n) \leq k \leq k(n+1),$$

$$(3.10) \quad 1 - 4a_n \leq \mu(I')/\mu(I) \leq 1 + 4a_n.$$

This makes  $\mu$  uniformly locally uniform by Proposition 2.2.

**Inductive construction.** Assume that  $1 = k(1) < \dots < k(n)$  have been defined, hence also  $\lambda_1, \dots, \lambda_n$  via (3.8), and the partial Riesz product  $f_n$  via (3.1). Use the strict positivity and uniform continuity of  $f_n$  to choose  $k(n+1) \geq k(n) + 2$  so large that

$$(3.11) \quad 1 - a_{n+1} < \frac{f_n(x')}{f_n(x)} < 1 + a_{n+1} \quad \text{provided } |x - x'| \leq 2^{-k(n+1)+1}.$$

This completes the induction. Let  $\mu$  be the Riesz product defined by (3.2). Take  $n \geq 2$ , so

$$f_n(x) = f_{n-1}(x)[1 + a_n \cos(2\pi\lambda_n x)].$$

Thus

$$(3.12) \quad (1 - a_n)f_{n-1}(x) \leq f_n(x) \leq (1 + a_n)f_{n-1}(x).$$

Assume  $I$  and  $I'$  are as in (3.9). If  $x \in I$  and  $x \in I'$ , then

$$|x - x'| \leq 2^{-k+1} \leq 2^{-k(n)+1}.$$

We can use (3.11) with  $n - 1$  instead of  $n$ . This and (3.12) give the bounds

$$\frac{(1 - a_n)^2}{(1 + a_n)} < \frac{f_n(x')}{f_n(x)} < \frac{(1 + a_n)^2}{(1 - a_n)}.$$

The same bounds for  $\mu(I')/\mu(I)$  follow from Lemma 3.5. This yields (3.10), since  $a_n \leq \frac{1}{5}$  by assumption. Conclusion: the measure  $\mu$  is uniformly locally uniform.

*Remarks.*

(i) Elementary estimates show that  $f_n$  is bounded between  $\exp(\pm 3\sqrt{n})$ , and its logarithmic derivative is bounded in absolute value by a constant times  $\lambda_n \exp(6\sqrt{n})$ . Thus, (3.11) will be achieved provided

$$2^{k(n)-k(n+1)} e^{6\sqrt{n}} / a_{n+1} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

In particular,  $k(n)$  of order  $n^\alpha$  for  $\alpha > \frac{3}{2}$  will do the job.

(ii) Given any sequence  $\delta_k$  decreasing to zero as  $k \rightarrow \infty$ , replacing  $2^{-k(n+1)+1}$  by  $2^{-k(n+1)+1} + \delta_{k(n+1)}$  in (3.11) gives a singular  $\mu$  such that  $\mu(I')/\mu(I) \rightarrow 1$  as  $k \rightarrow \infty$ , uniformly over all pairs  $I$  and  $I'$  of binary intervals of order  $k$  within distance  $\delta_k$  of each other. In other words,  $\mu$  is nearly uniform over intervals which shrink to zero arbitrarily slowly.

4. PROOF OF THEOREM 1.3

In  $n$  tosses of a  $p$ -coin, let  $\hat{p}$  be the proportion of heads. The probability of getting any particular string of  $j$  heads and  $n - j$  tails is

$$(4.1) \quad L_n(p|\hat{p}) = p^j(1 - p)^{n-j} = [p^{\hat{p}}(1 - p)^{1-\hat{p}}]^n.$$

Let  $\mu$  be a prior distribution for  $p$ . For a Borel set  $A \subset [0, 1]$ , let

$$(4.2) \quad \mu_n^\#(A|\hat{p}) = \int_A L_n(p|\hat{p})\mu(dp).$$

The posterior distribution of  $p$  is the probability  $\tilde{\mu}_n(\cdot|\hat{p})$  on  $[0, 1]$  defined by

$$(4.3) \quad \tilde{\mu}_n(A|\hat{p}) = \frac{\mu_n^\#(A|\hat{p})}{\mu_n^\#([0, 1]|\hat{p})}.$$

Write

$$\hat{\sigma}^2 = \hat{p}(1 - \hat{p})/n.$$

As is well known, the likelihood function  $p \rightarrow \log L_n(p|\hat{p})$  is strictly concave; its maximum occurs at  $p = \hat{p}$ ; and the function can be closely approximated by a Gaussian density with mean  $\hat{p}$  and variance  $\hat{\sigma}^2$ , times a suitable scale factor. To be precise, we control the error in this approximation by the following lemma. Sharper estimates are given in Diaconis and Freedman [10].

4.4. Lemma.

$$L_n(p|\hat{p}) = L_n(\hat{p}|\hat{p}) \exp \left\{ -\frac{1}{2} \frac{(p - \hat{p})^2}{\hat{\sigma}^2} F(p, \hat{p}) \right\}$$

where the factor  $F(p, \hat{p})$  does not depend on  $n$ . For  $\varepsilon \in (0, \frac{1}{2})$ ,

$$F(p, \hat{p}) \rightarrow 1 \quad \text{as } |p - \hat{p}| \rightarrow 0,$$

uniformly over all  $p, \hat{p} \in [\varepsilon, 1 - \varepsilon]$ . Finally

$$(4.6) \quad L_n(p|\hat{p}) \leq L_n(\hat{p}|\hat{p}) \exp \left\{ -\frac{n}{2} (p - \hat{p})^2 \right\}.$$

Proof. Let

$$(4.7) \quad H(p|\hat{p}) = \frac{1}{n} \log L_n(p|\hat{p}) = \hat{p} \log p + (1 - \hat{p}) \log(1 - p).$$

This does not depend on  $n$ . Now

$$-\frac{d^2}{dp^2} H(p|\hat{p}) = \frac{\hat{p}}{p^2} + \frac{1 - \hat{p}}{(1 - p)^2} = Q(p, \hat{p}), \quad \text{say.}$$

The function  $p \rightarrow H(p|\hat{p})$  attains its maximum at  $\hat{p}$ . By Taylor's theorem, there is an  $r$  between  $p$  and  $\hat{p}$  such that

$$H(p|\hat{p}) = H(\hat{p}|\hat{p}) - \frac{1}{2}(p - \hat{p})^2 Q(r, \hat{p}).$$

Then

$$\begin{aligned} L_n(p|\hat{p}) &= \exp\{nH(p|\hat{p})\} \\ &= L_n(\hat{p}|\hat{p}) \exp\left\{-\frac{1}{2} \frac{(p - \hat{p})^2}{\hat{\sigma}^2} \hat{p}(1 - \hat{p})Q(r, \hat{p})\right\}. \end{aligned}$$

Set  $F(p, \hat{p}) = \hat{p}(1 - \hat{p})Q(r, \hat{p})$ . Relation (4.5) follows:

$$|r - \hat{p}| \rightarrow 0 \quad \text{as } |p - \hat{p}| \rightarrow 0 \quad \text{and} \quad Q(\hat{p}, \hat{p}) = 1/\hat{p}(1 - \hat{p}).$$

Finally,  $Q(r, \hat{p}) \leq 1$ ; and this implies (4.6).  $\square$

**4.8. Lemma.** Assume that the prior  $\mu$  is uniformly locally uniform. Fix  $\varepsilon \in (0, \frac{1}{2})$  and  $0 < \delta < K < \infty$ . Then

$$\mu_n^\#([\hat{p} - a\hat{\sigma}, \hat{p} + b\hat{\sigma}]|\hat{p}) \approx L_n(\hat{p}|\hat{p})\mu(\hat{p}, \hat{p} + \hat{\sigma})\sqrt{2\pi}[\Phi(b) - \Phi(-a)].$$

Here  $\Phi$  is the standard normal distribution function. And the notation  $\approx$  means that the ratio of the two sides tends to 1 as  $n \rightarrow \infty$ , uniformly over  $\hat{p} \in [\varepsilon, 1 - \varepsilon]$  and  $0 \leq a, b \leq K$  with  $a + b \geq \delta > 0$ .

*Proof.* Let  $I_{ab} = [\hat{p} - a\hat{\sigma}, \hat{p} + b\hat{\sigma}]$ . Define a function  $f_{ab}$  on  $[0, 1]$  by

$$f_{ab}(x) = e^{-[(a+b)x-a]^2/2}.$$

If  $A$  is the affine mapping of  $I_{ab}$  onto  $[0, 1]$ , then

$$(4.9) \quad \exp\left\{-\frac{1}{2} \frac{(p - \hat{p})^2}{\hat{\sigma}^2}\right\} = f_{ab}(A(p)) \quad \text{for } p \in I_{ab}.$$

Clearly,

$$(4.10) \quad \text{the } f_{ab} \text{ are bounded and uniformly equicontinuous as } a, b \text{ range over } [0, K].$$

We compute as follows:

$$\begin{aligned} \mu_n^\#(I_{ab}|\hat{p}) &= \int_{I_{ab}} L_n(p|\hat{p})\mu(dp) \quad \text{by definition of } \mu^\# \\ &\approx L_n(\hat{p}|\hat{p}) \int_{I_{ab}} \exp\left\{-\frac{1}{2} \frac{(p - \hat{p})^2}{\hat{\sigma}^2}\right\} \mu(dp) \quad \text{by (4.5)} \\ &\approx L_n(\hat{p}|\hat{p})\mu(I_{ab}) \int_0^1 f_{ab}(x)dx \quad \text{by (4.9), (4.10) and 2.5} \\ &= L_n(\hat{p}|\hat{p}) \frac{\mu(I_{ab})}{a + b} \sqrt{2\pi}[\Phi(b) - \Phi(-a)] \quad \text{by change of scale.} \end{aligned}$$

Use uniform local uniformity once more, as expressed in 2.4, with  $J = I_{01}$ ,  $J' = I_{ab}$ , and  $L = 2K$   $\square$ .

In Theorem 1.3 the imagery is Bayesian:  $\hat{p}$  and  $\hat{\sigma}^2 = \hat{p}(1 - \hat{p})/n$  are “constants” computed from the data;  $p$  is random, with posterior distribution  $\tilde{\mu}_n(\cdot|\hat{p})$  given the data; the theorem is about the posterior distribution of  $p \rightarrow (p - \hat{p})/\hat{\sigma}$ ; and the convergence to normality as  $n \rightarrow \infty$  is uniformly over  $\hat{p} \in [\varepsilon, 1 - \varepsilon]$ . To complete the proof of the Theorem, it only remains to establish the following lemma.

4.11. **Lemma.** *Suppose  $\mu$  is uniformly locally uniform. Fix  $\varepsilon, \delta > 0$ . There exists a  $K = K(\varepsilon, \delta)$  such that*

$$\tilde{\mu}_n([\hat{p} - K\hat{\sigma}, \hat{p} + K\hat{\sigma}]|\hat{p}) \geq 1 - \delta \quad \text{for all } \hat{p} \in [\varepsilon, 1 - \varepsilon]$$

for all  $n$  greater than some  $n(K, \varepsilon, \delta)$ .

*Proof.* We will argue that the posterior probability of  $(\hat{p} + K\hat{\sigma}, 1]$  is negligible. The argument for  $[0, \hat{p} - K\hat{\sigma})$  is symmetric. For  $j = 0, 1, \dots$ , let

$$I_j = (\hat{p} + j\hat{\sigma}, \hat{p} + (j + 1)\hat{\sigma}].$$

Since  $\mu$  is uniformly locally uniform, Condition 2.4 implies that for all sufficiently large  $n$ ,

$$\mu(I_{j+1}) \leq 2\mu(I_j), \quad \text{for all } j \text{ and } \hat{p}.$$

Thus

$$(4.12) \quad \mu(I_j) \leq 2^j \mu(I_0), \quad \text{for } j \geq 1.$$

Now for  $\hat{p} \in [\varepsilon, 1 - \varepsilon]$ ,

$$\begin{aligned} \mu_n^\#(I_j|\hat{p}) &\leq \mu(I_j)L_n(\hat{p}|\hat{p})e^{-j^2\varepsilon(1-\varepsilon)/2} && \text{by (4.6)} \\ &\leq \mu(I_0)L_n(\hat{p}|\hat{p})2^j e^{-j^2\varepsilon(1-\varepsilon)/2} && \text{by (4.12)}. \end{aligned}$$

Summing this estimate over all  $j \leq K$  gives

$$\mu_n^\#((\hat{p} + K\hat{\sigma}, 1]|\hat{p}) \leq \mu(I_0)L_n(\hat{p}|\hat{p})t(K, \varepsilon),$$

where  $t(K, \varepsilon)$  is the tail of a convergent series, so  $t(K, \varepsilon) \rightarrow 0$  as  $K \rightarrow \infty$  for every  $\varepsilon$ . On the other hand, using (4.8),

$$\mu_n^\#([0, 1]|\hat{p}) \geq \mu_n^\#(I_0|\hat{p}) \approx \mu(I_0)L_n(\hat{p}|\hat{p})\sqrt{2\pi}[\Phi(1) - \Phi(0)].$$

Finally, use (4.3).  $\square$

### 5. EXISTENCE AND SINGULARITY OF THE RIESZ PRODUCT

This section provides quick proofs from the literature of the two features of the Riesz product which are essential for the construction in §3.

*Proof of existence of the Riesz product (3.1).* (Adapted from Katznelson [11], p. 107.) The condition  $\lambda_{j+1} \leq 3\lambda_j$  makes each integer  $m$  have at most one representation of the form  $\sum \xi_j \lambda_j$  where  $\xi_j = -1, 0, 1$ . If  $m$  admits such a representation, and  $|m| \leq \lambda_N$ , then  $\xi_j = 0$  for all  $j > N$ . Now fix  $m$  and

$N$  with  $|m| \leq \lambda_N$ , and consider  $n > N$ . The  $m$ th Fourier coefficient of the density  $f_n$  in (3.1) is

$$(5.1) \quad \int_0^1 e^{-2\pi i m x} f_n(x) dx = \prod_{\xi_j \neq 0} \frac{1}{2} a_j, \quad \text{if } m = \sum \xi_j \lambda_j \text{ with } \xi_j = -1, 0, 1,$$

$$= 0, \quad \text{if } m \text{ does not admit this representation.}$$

Thus, as  $n \rightarrow \infty$ , every Fourier coefficient of  $f_n$  is eventually constant. By a standard theorem, the probabilities with densities  $f_n$  must therefore converge weak star to a probability  $\mu$ , whose nonzero Fourier coefficients are defined by the right side of (5.1).  $\square$

*Proof of singularity of the Riesz product (3.1) in case  $\sum a_n^2 = \infty$ .* (Adapted from Peyrière [12]). We find from (5.1) that

$$(5.2) \quad \int_0^1 e^{\pm 2\pi i \lambda_n x} \mu(dx) = \frac{1}{2} a_n$$

and

$$(5.3) \quad \int_0^1 e^{\pm 2\pi i \lambda_n x \pm 2\pi i \lambda_{n'} x} \mu(dx) = \frac{1}{2} a_n \frac{1}{2} a_{n'}.$$

In particular, the functions

$$x \rightarrow e^{2\pi i \lambda_n x} - \frac{1}{2} a_n$$

are bounded and orthogonal in  $L^2(\mu)$ . Because  $\sum a_n^2 = \infty$ , there must exist a sequence  $c_n$  with

$$\sum c_n^2 < \infty, \quad c_n a_n \geq 0, \quad \sum c_n a_n = \infty.$$

In our application, with  $a_n = 1/(5\sqrt{n})$ ,  $c_n = 1/(\sqrt{n} \log n)$  will do. But (5.2), (5.3) and  $\sum c_n^2 < \infty$  imply the series of functions

$$\sum c_n (e^{2\pi i \lambda_n x} - a_n/2)$$

is an orthogonal convergent series in  $L^2(\mu)$ . On the other hand, the series

$$\sum c_n e^{2\pi i \lambda_n x}$$

is orthogonal and convergent in  $L^2$  (Lebesgue). By passing to subsequences, we can make the first series converge for  $\mu$ -almost all  $x$ , and the second for Lebesgue-almost all  $x$ . If  $\mu$  were not singular, there would be an  $x$  and a subsequence along which both series converged, hence also their difference would converge. But this is a contradiction, since their difference is  $\frac{1}{2} \sum c_n a_n = \infty$ .  $\square$

The idea behind this and similar arguments of Brown and Moran [13, 14] is extended by Brown [15], to give a general criterion for mutual singularity

of probability measures based on comparison of sequences of square integrable random variables with low correlation.

## 6. HISTORY AND ACKNOWLEDGMENTS

F. Riesz [16] introduced the product (3.2) to exhibit various possible behaviors of Fourier coefficients at infinity. In particular, by taking  $a_j = 1$  for all  $j$  and  $\lambda_j = 4^j$ , he gave the first example of a continuous singular measure with Fourier coefficients not vanishing at infinity. See Graham and McGehee [17, Chapter 7] for an extensive treatment of Riesz products in a more general setting.

In Freedman and Pitman [18], we presented a continuous, singular probability measure  $\mu$  on  $[0, 1]$  which is locally uniform at  $\mu$ -almost every point  $x$  in  $[0, 1]$ . This  $\mu$  was defined as the probability which makes the binary digits of  $x$  independent, the  $n$ th digit being a one with probability  $\frac{1}{2} + \varepsilon_n$  and a zero with probability  $\frac{1}{2} - \varepsilon_n$ , for a particular sequence  $\varepsilon_n$  with  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$  but  $\sum \varepsilon_n^2 = \infty$ . Singularity follows from the criterion of Kakutani [19]. By working with such coin-tossing measures  $\mu$  defined by various sequences  $\varepsilon_n$ , and introducing an element of smoothing into the construction, we subsequently became convinced that the exceptional set could be eliminated entirely, to yield a measure that was singular and uniformly locally uniform (SULU). But the construction was rather intricate. We are therefore very grateful to Russell Lyons, who suggested that such a measure might be created more easily as a Riesz product, and pointed us to the literature of these measures.

There is a close parallel between Riesz products and ordinary product measures of the coin-tossing kind mentioned above. For instance, it is easy to see that the coin-tossing measure for probabilities  $\frac{1}{2} \pm \varepsilon_n$  is the weak star limit as  $N \rightarrow \infty$  of the probability with density

$$(6.1) \quad f_N(x) = \prod_1^N [1 + 2\varepsilon_n r_n(x)],$$

where  $r_n(x)$  is the Rademacher function whose value is  $\pm 1$  according to the  $n$ th binary digit of  $x$ . Compare with the definition of the Riesz product via (3.1): cosines instead of Rademacher functions, and  $a_n$  instead of  $2\varepsilon_n$ . Zygmund's dichotomy for Riesz products used in §3 thus corresponds to Kakutani's dichotomy for coin-tossing: the measures are either singular or absolutely continuous. The Rademacher functions are easier to deal with in some respects, since they are independent under Lebesgue measure, whereas the cosines are only orthogonal. But the smoothness of the cosines make the Riesz product easier to manipulate for present purposes.

## REFERENCES

1. N. Dunford and J. T. Schwartz, *Linear operators. Part I: General Theory*, Wiley Interscience, New York, 1958.
2. S. Saks, *Theory of the integral*, (1937), reprinted in English by Dover, New York, 1964.
3. P. S. Laplace, *Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations*, Mémoires présentés à l'Académie des Sciences, Paris, 1809.
4. S. Bernstein, *Theory of probability* (in Russian), Moscow, 1917.
5. R. von Mises, *Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, (1931).
6. L. Le Cam, *Asymptotic methods in statistical decision theory*, Springer-Verlag, New York, (1986).
7. —, *On the Bernstein-von Mises theorem*, Technical report no. 57, Department of Statistics, University of California, Berkeley, (1986).
8. A. Zygmund, *On lacunary trigonometric series*, Trans. Amer. Math. Soc. **34** (1932), 435–446.
9. —, *Trigonometric Series, two volumes*, The University Press, Cambridge, 1959.
10. P. Diaconis and D. Freedman, *On the uniform consistency of Bayes estimates for multinomial probabilities*, Technical report no. 137, Department of Statistics, University of California, Berkeley, 1988.
11. Y. Katznelson, *An introduction to harmonic analysis*, Reprinted by Dover, New York, 1976.
12. J. Peyrière, *Sur les produits de Riesz*, C. R. Acad. Sci. Paris Sér. A-B **276** (1973), 1453–1455.
13. G. Brown and W. Moran, *On orthogonality of Riesz products*, Proc. Cambridge Philos. Soc. **76**, (1974) 173–181.
14. —, *Coin-tossing and powers of singular measures*, Math. Proc. Cambridge Philos. Soc. **77**, (1975), 349–364.
15. G. Brown, *Singular infinitely divisible distributions whose characteristic functions vanish at infinity*, Math. Proc. Cambridge Philos. Soc. **82**, (1977), 277–287.
16. F. Riesz, *Über die Fourierkoeffizienten einer stetigen Funktion von beschränkter Schwankung*, Math. Z. **18**, (1918), 312–315.
17. C. C. Graham and O. C. McGehee, *Essays in commutative harmonic analysis*, Springer-Verlag, New York, 1979.
18. D. Freedman and J. W. Pitman, *A singular measure which is locally uniform*, Technical report no. 163, Department of Statistics, University of California, Berkeley, 1988.
19. S. Kakutani, *On equivalence of infinite product measures*, Ann. Math **49** (1948), 214–244.

STATISTICS DEPARTMENT, UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA 94720