The Annals of Statistics
1982, Vol. 10, No. 1. 184–189

DE FINETTI'S THEOREM FOR SYMMETRIC LOCATION FAMILIES

BY DAVID FREEDMAN[1] AND PERSI DIACONIS[2]

University of California, Berkeley and Stanford University

Necessary and sufficient conditions are obtained for an exchangeable sequence of random variables to be a mixture of symmetric location families.

1. Introduction. This paper characterizes mixtures of symmetric location families. More specifically, let $X = (X_1, X_2, \cdots)$ be an exchangeable sequence of real-valued random variables. By de Finetti's theorem, X is a mixture of independent and identically distributed random variables. When does the representation take the special form of a mixture of distributions symmetric about a location parameter θ, where θ varies too?

More technically, let $\mathscr{S}$ be the set of distribution functions symmetric about 0. The object is to characterize processes X such that

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = \int_{\mathscr{S}} \int_R \prod_{i=1}^n F(x_i - \theta)\mu(dF, d\theta). \tag{1.1}$$

Here, μ is a probability on $\mathscr{S} \times R$, and the equation is to hold for all n and $x_1, \cdots, x_n$.

To state the theorem, let

$$T_m = \frac{1}{m}(X_2 + \cdots + X_{m+1}).$$

Then X will be called *location symmetric* if for every m, the distribution of $X_1 - T_m$ is symmetric. Informally, T_m is an estimate of θ; the difference between X_1 and the estimate is to be symmetric. Further, X will be called *conditionally location symmetric* if for every n, given $X_1, \cdots, X_n$, the process $X_{n+1}, X_{n+2}, \cdots$, is location symmetric. The following theorem will be proven in Section 2.

THEOREM 1. *Let* $X = (X_1, X_2, \cdots)$ *be a sequence of random variables. Then* (1.1) *holds if and only if* X *is exchangeable and conditionally location symmetric.*

Mixed distributions like (1.1) arise in Bayesian estimation of the location θ of a symmetric distribution of unspecified form. This is one Bayesian approach to robustness. For example, Box and Tiao (1962) consider F in a finite dimensional family of symmetric "power" distributions with parameters to control the scale and kurtosis. Fraser (1972) chooses the family of t-distributions with variable scale and degrees of freedom. Hogg (1972) considers the search for adaptive robust estimates from a Bayesian viewpoint. Dempster (1975) gives an extensive review of Bayesian approaches to robustness. A recent discussion is in Ramsay and Novick (1981). We have computed the posterior for a Dirichlet prior on F in Diaconis and Freedman (1981).

Section 2 also gives some other characterizations involving symmetry about an invariant consistent estimator of θ; Theorem 1 is different, in that the average is inconsistent for long-tailed error distributions.

Section 3 gives counterexamples. In particular, exchangeability and location symmetry

Received February 25, 1981; revised April 1981.

[1] Research partially supported by NSF Grant MCS-80-02535.

[2] Research partially supported by NSF Grant MCS77-16974.

AMS 1980 *subject classification*: 62C10, 62E10.

Key words and phrases. Exchangeability, robustness, characteristic functions.

do not imply (1.1): conditional location symmetry is needed. Section 4 considers independence of θ and F.

2. Characterizing symmetric location families. The "only if" part of Theorem 1 is almost obvious. The proof of the "if" part is accomplished by Lemmas 1 and 2.

LEMMA 1. *Let $X = (X_1, X_2, \cdots)$ be exchangeable and conditionally location symmetric. Then X is a mixture of location symmetric sequences of independent, identically distributed random variables.*

PROOF. The hypotheses imply that almost surely

$$P(X_1 - T_m \le x \mid X_{j+1}, \cdots, X_{j+k}) = P(X_1 - T_m \le -x \mid X_{j+1}, \cdots, X_{j+k})$$

for $m \le j$ and $k \ge 1$. Let $k \to \infty$ and then $j \to \infty$ to see that almost surely, given the tail σ-field, X is still location symmetric. On the other hand, a version of de Finetti's theorem asserts that almost surely, given the tail σ-field, $X_1, X_2, \cdots$, are independent and identically distributed. To push this argument through, a regular conditional distribution given the tail σ-field is needed, as in Diaconis and Freedman (1980, Appendix). □

LEMMA 2. *Let X_1, X_2, $\cdots$, be a location symmetric sequence of independent and identically distributed random variables. Then for some real number θ, the distribution of $X_i - \theta$ is symmetric about* 0.

PROOF. Let ϕ be the characteristic function of X_1. Choose $\varepsilon > 0$ so small that $\phi(t) \ne 0$ for $|t| \le \varepsilon$. For such t, there is a unique real valued continuous function $A(t)$ such that $A(0) = 0$ and

$$\phi(t) = e^{iA(t)}|\phi(t)|. \tag{2.1}$$

In particular, $t \to \log|\phi(t)| + iA(t)$ is a branch of $\log \phi(t)$. Of course, $A(-t) = -A(t)$ and $\log|\phi(-t)| = \log|\phi(t)|$. Location symmetry and independence imply that for any t and m,

$$\phi(t)\phi^m(-t/m) = \phi(-t)\phi^m(t/m). \tag{2.2}$$

Hence, if $|t| \le \varepsilon$, for our branch of the log,

$$\log \phi(t) + m \log \phi(-t/m) = \log \phi(-t) + m \log \phi(t/m).$$

Substitute the definition of $\log \phi$ in terms of A, and rearrange:

$$A(t) = mA\left(\frac{t}{m}\right).$$

Let $s = t/m$, and put $m = 2$ or 3: if $|s| \le \frac{1}{3}\varepsilon$ then

$$A(2s) = 2A(s) \quad \text{and} \quad A(3s) = 3A(s).$$

By induction, if j and k are signed integers with $2^j 3^k \le 1$ and $0 \le u \le \frac{1}{3}\varepsilon$ then

$$A(2^j 3^k u) = 2^j 3^k A(u).$$

Rational numbers of the form $2^j 3^k$ are dense in [0, 1] and A is continuous. Therefore, there is a real number θ such that

$$A(t) = \theta t \qquad \text{for } 0 \le t \le \frac{\varepsilon}{3}.$$

Likewise,

$$A(t) = \theta' t \qquad \text{for } -\frac{\varepsilon}{3} \le t \le 0.$$

Since $A(-t) = -A(t)$, it follows that $\theta' = \theta$. That is, A is linear on $[-\varepsilon/3, \varepsilon/3]$. By (2.1),

$$\phi(t) = e^{i\theta t}|\phi(t)| \qquad \text{for } |t| \le \frac{\varepsilon}{3}. \tag{2.3}$$

To complete the proof, let t be any real number. Choose m so large that $|t/m| \le \varepsilon/3$. By (2.2), and (2.3) with $\pm t/m$ in place of t

$$\phi(t)e^{-i\theta t}\left|\phi^m\left(\frac{t}{m}\right)\right| = \phi(-t)e^{+i\theta t}\left|\phi^m\left(\frac{t}{m}\right)\right|. \tag{2.4}$$

Set $\psi(t) = \phi(t)e^{-i\theta t}$, the characteristic function of $X_1 - \theta$. The factor $|\phi^m(t/m)|$ cancels in (2.4), because $\phi(t/m) \neq 0$. So ψ is real, and the distribution $X_1 - \theta$ is symmetric about 0. □

Other forms of the theorem will now be indicated. To begin with, T_m can be defined as $(X_1 + \cdots + X_m)/m$ rather than $(X_2 + \cdots X_{m+1})/m$; the argument is about the same. Also the mean can be replaced by other statistics, like the median or a trimmed mean. More generally, consider a sequence of measurable functions f_n from R^n to R. Say these are *location statistics* provided

$$f_n(x_1 + c, \cdots, x_n + c) = f_n(x_1, \cdots, x_n) + c \tag{2.5}$$

$$f_n(-x_1, \cdots, -x_n) = -f_n(x_1, \cdots, x_n) \tag{2.6}$$

and *consistent* provided $f_n(X_1, \cdots, X_n)$ converges a.e. to a constant, for any sequence X_1, $X_2, \cdots$, of independent, identically distributed random variables. If the latter have a distribution symmetric about 0, the limit must be 0 by (2.6); if the latter have a distribution symmetric about θ, the limit must be θ by (2.5).

Let $f = (f_1, f_2, \cdots)$ be a sequence of location statistics and $X = (X_1, X_2, \cdots)$ a sequence of random variables. Then X is *f-symmetric* provided the distribution of $X_1 - f_m(X_1, \cdots, X_m)$ is symmetric about 0, for all m. And X is *conditionally f-symmetric* provided that for every n, given $X_1, \cdots, X_n$, the sequence $X_{n+1}, X_{n+2}, \cdots$, is f-symmetric.

THEOREM 2. *Let $f = (f_1, f_2, \cdots)$ be a consistent sequence of location statistics, and $X = (X_1, X_2, \cdots)$ a sequence of random variables. Then* (1.1) *holds if and only if X is exchangeable and conditionally f-symmetric.*

PROOF. Again, the "only if" part is easy. For the "if" part, as before, given the tail σ-field the X-process is an f-symmetric sequence of independent, identically distributed sequences of random variables. (This uses only the equivariance of f.) Since f is consistent, X_1 must be symmetric about θ, the limit of $f_n(X_1, \cdots, X_n)$. □

3. Examples.

EXAMPLE 1. There is an exchangeable process X which is location symmetric, but not conditionally location symmetric. The representation (1.1) does not apply. Thus, conditional location symmetry must be assumed in Theorem 1.

Construction. Let $Z = (Z_1, Z_2, \cdots)$ be a sequence of independent random variables, with a common distribution unsymmetric about 0. Let $X = Z$ or $-Z$ with probability ½. Location symmetry is almost obvious. The uniqueness part of de Finetti's theorem shows that X cannot be a mixture of symmetric variables: (1.1) fails. □

Our first try at formulating Theorem 1 involved the following notion: X is *string symmetric* if the distribution of $a_1X_1 + \cdots + a_mX_m$ is symmetric about 0 for each $m \ge 1$ and each string $a_1, \cdots, a_m$ of real numbers with $a_1 + \cdots + a_m = 0$. And X is *conditionally string* symmetric if for each n, given $X_1, \cdots, X_n$, the sequence $X_{n+1}, X_{n+2}, \cdots$, is string symmetric. We found that (1.1) holds if and only if X is exchangeable and conditionally string symmetric.

On its face, location symmetry is a weaker condition than string symmetry: for each m, only one linear combination is involved, viz.

$$a_1 = 1, \qquad a_2 = \frac{1}{m-1}, \cdots, a_m = -\frac{1}{m-1}.$$

Of course, Lemma 2 shows that for sequences of independent and identically distributed random variables, the two conditions are equivalent.

We wondered whether it was enough to assume string symmetry for some fixed m, e.g., $m = 3$. The answer is no, as Example 2 shows. The following Lemma is needed. It gives an example of a characteristic function that is real in a neighborhood of zero, but not real everywhere. For a related construction, see Shepp, Slepian and Weiner (1980).

LEMMA 3. *For any* $A > 1$ *there is a random variable with mean* 0, *moments of arbitrarily high order, and a characteristic function which is real on* $[0, 1]$, *vanishes on* $[1, A] \cup [A + 1, \infty)$, *and is purely imaginary on* $[A, A + 1]$.

PROOF. The random variable will have a probability density of the form

$$f = c(f_1 + \delta f_2)$$

where the functions f_1 and f_2 are to be constructed, $f_1 \geq 0$ and f_2 is real; $\delta > 0$ will be chosen so small that $f_1 + \delta f_2 \geq 0$; then c can be chosen so the total mass is one. Let ^ stand for Fourier transform. Then the characteristic function $\phi = \hat{f}$ is

$$\phi = c(\hat{f}_1 + \delta \hat{f}_2).$$

Matters will be arranged so that $\hat{f}_1$ is real and vanishes off $[-1, 1]$; while $\hat{f}_2$ is purely imaginary, and vanishes off $[-A-1, -A] \cup [A, A+1]$.

To construct f_1, let

$$h(x) = \frac{\sin x}{x}.$$

Of course, the uniform density on $[-1, 1]$ has Fourier transform $h(t)$. Now let

$$H(x) = h(x/2^k)^{2^k}.$$

Then $H(t)$ is the characteristic function of

$$V = \frac{1}{2^k}(U_1 + \cdots + U_{2^k}),$$

the U's being independent and uniform on $[-1, 1]$. In particular, the probability density g of V is a quite smooth function supported on $[-1, 1]$. By taking an inverse Fourier transform, one sees that $\hat{H} = 2\pi g$ is a nonnegative real function vanishing off $[-1, 1]$. Finally, let

$$f_1(x) = H(x+1) + H(x-1).$$

For use later, verify the existence of a positive ε with

$$|x|^{2^k} f_1(x) \geq \varepsilon \qquad \text{for all } x \text{ with } |x| \geq 1. \tag{3.1}$$

The argument uses the periodicity of the sine function, and the irrationality of π; details are omitted. Clearly,

$$\hat{f}_1(t) = 2\hat{H}(t) \cos t$$

vanishes off $[-1, 1]$ as well.

To construct f_2, let $\psi(t)$ be a C_∞ purely imaginary function of the real argument t, vanishing except when $A < |t| < A + 1$, and satyisfying $\psi(-t) = -\psi(t)$. Let f_2 be the inverse Fourier transform of ψ. Then f_2 is real because ψ is odd, and integrating by parts j times shows in the usual way that

$$\sup_x |x|^j |f_2(x)| < \infty \qquad \text{for any } j \geq 1.$$

From this and (3.1), the existence of δ follows. Plainly, there are $2^k - 2$ moments. □

EXAMPLE 2. For each $m \geq 2$ and $N \geq 1$, there is a sequence $X_1, X_2, \cdots$, of independent identically distributed random variables such that:

i) X_1 has mean 0 and finite Nth moment
ii) X_1 is not symmetric
iii) if $a_1 + \cdots + a_m = 0$, then the distribution of $a_1X_1 + \cdots a_mX_m$ is symmetric about 0.

PROOF. Use Lemma 3, with $A > 2m$. Let $X_1, X_2, \cdots$, have the characteristic function ϕ constructed there. What must be shown is that $\sum_{j=1}^m t_j = 0$ entails

$$\prod_{j=1}^m \phi(t_j) = \prod_{j=1}^m \phi(-t_j). \tag{3.2}$$

The equality is trivial unless $|t_j| < 1$ or $A < |t_j| < A + 1$ for all j, so assume this to be the case. If $|t_j| < 1$, then $\phi(t_j) = \phi(-t_j)$; so it is enough to show that

$$\prod_{j\in S} \phi(t_j) = \prod_{j\in S} \phi(-t_j)$$

where S is the set of j's with $A < |t_j| < A + 1$. Now $\phi(-t_j) = -\phi(t_j)$ for $j \in S$ and it remains only to show that the cardinality of S is even.

Let J be the number of j's with $A < t_j < A + 1$, and K the number with $-A - 1 < t_j < -A$; so the cardinality of S is $J + K$. But $J = K$. For example, if $J > K$ then

$$\sum_{j\in S} t_j > JA - K(A + 1) \geq A - K \geq A - m,$$

but $j \in S$ entails $|t_j| < 1$ by assumption so that

$$|\sum_{j\notin S} t_j| < m.$$

Finally, $A > 2m$, so

$$\sum_j t_j > A - 2m > 0.$$

This contradiction completes the proof. □

The characteristic function constructed in Lemma 3 is also of interest in providing a counterexample to Theorem (5.31) of Kagan, Linnik and Rao (1973). Part (ii) of their theorem involves independent, identically distributed random variables having zero mean and finite variance, and states that X_1 is symmetric if and only if $E(X_1 + X_2 | X_1 - X_2) = 0$. As argued by Kagan, Linnik and Rao, the conditional expectation is zero if and only if the characteristic function ϕ of X_1 satisfies $\phi(t)\{\phi(-t)\}' = \phi(-t)\phi'(t)$. It is easy to see that the characteristic function constructed in lemma (3.2) satisfies this equation: if $|t| < 1$, then $\phi(t) = \phi(-t)$; if $A < |t| < A + 1$, then $\phi(t) = -\phi(-t)$; for other values of t, both sides vanish. By construction, the random variable corresponding to ϕ is not symmetric.

4. Independence. It is customary to take θ and F independent in (1.1). We do not know a neat condition on finite collections of the X_i for this to hold. In thinking about this problem we were led to ask if there was a function of $X_1, \cdots, X_n$ whose distribution depends only on F, not on θ. This turns out to be impossible, even if the shape of F is known up to a scale parameter. The following proposition is closely related to results of

Dantzig (1940) and Stein (1945) on fixed width confidence sets for a normal location parameter.

PROPOSITION 4.1. *Let X be a random vector in $\mathbb{R}^k$ which has an absolutely continuous distribution with density f. Let $Y_{\theta,\sigma} = \theta + \sigma X$. If g is a measurable function from $\mathbb{R}^k$ into the measurable space $(\mathscr{X}, b)$ such that the distribution of $g(Y_{\theta,\sigma})$ only depends on θ, then g is constant a.e.*

PROOF. It is enough to show that for every measurable set A, if $P(\theta + \sigma X \in A)$ depends only on θ, then this probability is constant. Now

$$|P(\theta + \sigma X \in A) - P(\sigma X \in A)| \leq \sup_A \left| P\left(\frac{\theta}{\sigma} + X \in A\right) - P(X \in A) \right|$$

$$= \frac{1}{2} \int \left| f\left(x - \frac{\theta}{\sigma}\right) - f(x) \right| dx.$$

The right side of the inequality becomes arbitrarily small as σ tends to infinity, because translations are continuous in L^1. □

REFERENCES

BERMAN, S. M. (1962). An extension of the arc sine law. *Ann. Math. Statist.* **33** 681–684.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Mass.

DANTZIG, G. B. (1940). On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ. *Ann. Math. Statist.* **11** 186–192.

DEMPSTER, A. P. (1975). A Subjective Look at Robustness. Research Report S-33 Harvard University.

DIACONIS, P. and FREEDMAN, D. (1980). De Finetti's theorem for Markov chains. *Ann. Probab.* **8** 115–130.

DIACONIS, P. and FREEDMAN, D. (1981). An example of inconsistent Bayes estimates. Unpublished manuscript.

FRASER, D. A. S. (1976). Necessary analysis and adaptive inference. *J. Amer. Statist. Assoc.* **71** 99–113.

HOGG, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *J. Amer. Statist. Assoc.* **69** 909–922.

KAGAN, A. M., LINNIK, YU. V., and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York.

RAMSAY, J. O. and NOVICK, M. R. (1981). PLU robust Bayesian decision theory: point estimation. *J. Amer. Statist. Assoc.* **75** 901–907.

SHEPP, L. A., SLEPIAN, D. and WEINER, A. D. (1980). On prediction of moving-average processes. *Bell System Tech. J.*

STEIN, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

STATISTICS DEPARTMENT
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720

DEPARTMENT STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CA 94305