De FINETTI'S THEOREM FOR SYMMETRIC LOCATION FAMILIES

BY

DAVID FREEDMAN and PERSI DIACONIS

TECHNICAL REPORT NO. 165
MARCH 1981

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
De FINETTI'S THEOREM FOR SYMMETRIC LOCATION FAMILIES

BY

DAVID FREEDMAN and PERSI DIACONIS

TECHNICAL REPORT NO. 165
MARCH 1981

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
De Finetti's Theorem for Symmetric Location Families

by

David Freedman¹
University of California, Berkeley

and

Persi Diaconis²
Stanford University

Abstract

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

AMS/MOS classification number:

Key words and phrases:

¹Research partially supported by NSF Grant MCS-80-02535.
²Research partially supported by NSF Grant MCS77-16974.
De Finetti's Theorem for Symmetric Location Families

by

David Freedman
University of California, Berkeley

and

Persi Diaconis
Stanford University

1. Introduction

This paper characterizes mixtures of symmetric location families. More specifically, let \( X = (X_1, X_2, \ldots) \) 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 \( \theta \), where \( \theta \) varies too.

More technically, let \( \mathcal{G} \) be the set of distribution functions symmetric about 0. The object is to characterize processes \( X \) such that

\[
\Pr\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = \int_{\mathcal{G}} \int_{\mathbb{R}} \prod_{i=1}^{n} F(x_i - \theta) \, \mu(dF, d\theta). \tag{1}
\]

Here, \( \mu \) is a probability on \( \mathcal{G} \times \mathbb{R} \), and the equation is to hold for all \( n \) and \( x_1, \ldots, x_n \).

To state the theorem, let \( T_m = \frac{1}{m} (X_2 + \ldots + X_{m+1}) \). Then \( X \) will be called location symmetric if for every \( m \), the distribution of \( X_1 - T_m \)
is symmetric. And $X$ will be called conditionally location symmetric about 0. If for every $n$, given $X_1, \ldots, X_n$, the process $X_{n+1}, X_{n+2}, \ldots$, is location symmetric.

Informally, $X$ is conditionally location symmetric if no matter what the past is, the future is symmetric about an estimate of $\theta$. The following theorem will be proven in Section 2.

(1.2) **Theorem.** Let $X = (X_1, X_2, \ldots)$ 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) arise in Bayesian estimation of the location $\theta$, of a symmetric distribution $F$, of unspecified form - the Bayesian approach to robustness. Current Bayesian approaches put independent priors on $\theta$ and $F$. 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). All of the cited references assume that $\theta$ and $F$ are independent under the prior. We do not know a neat condition on $X$ for this to hold.

Section 2 also gives some other characterizations involving symmetry about an invariant consistent estimator of $\theta$; theorem (1.2) is different, in that the average is inconsistent for long-tailed error distributions.
Section 3 gives counterexamples. In particular, exchangeability and location symmetry do not imply (1.1): conditional location symmetry is needed.

2. Characterizing Symmetric Location Families

The "only if" part of theorem (1.2) is almost obvious. The proof of the "if" part is accomplished by lemma (2.1) and (2.2).

(2.1) Lemma. Let \( X = (X_1, X_2, \ldots) \) 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

\[
\Pr\{X_1 - T_m \leq x | X_{j+1}, \ldots, X_{j+k}\} = \Pr\{X_1 - T_m \geq -x | X_{j+1}, \ldots, X_{j+k}\}
\]

for \( m \leq j \) and \( k \geq 1 \). Let \( k \to \infty \) and then \( j \to \infty \) to see that almost surely, given the tail \( \sigma \)-field, \( X \) is still location symmetric.

On the other hand, a version of de Finetti's theorem asserts that almost surely, given the tail \( \sigma \)-field, \( X_1, X_2, \ldots \), are independent and identically distributed. To push this argument through, a regular conditional distribution given the tail \( \sigma \)-field is needed, as in Diaconis and Freedman (1980, Appendix).

(2.2) Lemma. Let \( X_1, X_2, \ldots \), be a location symmetric sequence of independent and identically distributed random variables. Then
for some real number \( \theta \), the distribution of \( X_1 - \theta \) is symmetric about 0.

**Proof.** Let \( \phi \) be the characteristic function of \( X_1 \). Choose \( \varepsilon > 0 \) so small that \( \phi(t) \neq 0 \) for \( |t| \leq \varepsilon \). For such \( t \), there is a unique real valued continuous function \( A(t) \) such that \( A(0) = 0 \) and

\[
\phi(t) = e^{iA(t)} |\phi(t)|.
\]

In particular, \( t \to \log |\phi(t)| + iA(t) \) is a branch of \( \log \phi(t) \), and for this branch \( |t| \leq \varepsilon \) and \( n \geq 1 \) imply

\[
\log[\phi(t) \cdot \phi^n(-t/n)] = \log \phi(t) + n \log \phi(-t/n).
\]

Of course, \( A(-t) = -A(t) \) and \( \log |\phi(-t)| = \log |\phi(t)| \). Location symmetry and independence imply that for any \( t \) and \( m \),

\[
\phi(t) \phi^m(-t/m) = \phi(-t) \phi^m(t/m).
\]

Use this and (2.4): if \( |t| \leq \varepsilon \),

\[
\log \phi(t) + m \log \phi(-t/n) = \log \phi(-t) + m \log \phi(t/n).
\]

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

\[
A(t) = m A(t/m).
\]

Let \( s = n t/m \), and put \( m = 2 \) or 3: if \( |s| \leq \frac{1}{3} \varepsilon \) then

\[
A(2s) = 2A(s) \text{ and } A(3s) = 3A(s).
\]
By induction, if \( j \) and \( k \) are signed integers with \( 2^j 3^k \leq 1 \) and \( 0 \leq u \leq \frac{1}{3} \varepsilon \) 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 \( \theta \) such that
\[
A(t) = \theta t \quad \text{for} \quad 0 \leq t \leq \frac{1}{3} \varepsilon.
\]

Likewise,
\[
A(t) = \theta' t \quad \text{for} \quad -\frac{1}{3} \varepsilon \leq t \leq 0.
\]

Since \( A(-t) = -A(t) \), it follows that \( \theta' = \theta \). That is, \( A \) is linear on \([-\frac{1}{3} \varepsilon, \frac{1}{3} \varepsilon]\). By (2.3),
\[
(2.6) \quad \phi(t) = e^{i\theta t} |\phi(t)| \quad \text{for} \quad |t| \leq \frac{1}{3} \varepsilon.
\]

To complete the proof, let \( t \) be any real number. Choose \( m \) so large that \( |t/m| \leq \frac{1}{3} \varepsilon \). By (2.5), and (2.6) with \( \pm t/m \) in place of \( t \).

\[
(2.7) \quad \phi(t) e^{-i\theta t} |\phi^m(t/m)| = \phi(-t) e^{i\theta t} |\phi^m(t/m)|.
\]

Set \( \Psi(t) = \phi(t) e^{-i\theta t} \), the characteristic function of \( X_1 - \theta \). The factor \( |\phi^m(t/m)| \) cancels in (2.7), because \( \phi(t/m) \neq 0 \). So \( \Psi \) is real, and the distribution \( X_1 - \theta \) is symmetric about \( 0 \). \( \square \)
Other forms of the theorem will now be indicated. To begin with, \( T_m \) can be defined as \( \frac{1}{m}(X_1 + \ldots + X_m) \) rather than \( \frac{1}{m}(X_2 + \ldots + X_{m+1}) \); 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 \( \mathbb{R}^n \) to \( \mathbb{R} \). Say these are location statistics provided

\[
(2.8a) \quad f_n(x_1 + c, \ldots, x_n + c) = f_n(x_1, \ldots, x_n) + c
\]

\[
(2.8b) \quad f_n(-x_1, \ldots, -x_n) = -f_n(x_1, \ldots, x_n)
\]

and consistent provided \( f_n(x_1, \ldots, x_n) \) converges, a.e., to a constant, for any sequence \( X_1, X_2, \ldots \), of independent, identically distributed random variables. If the latter have a distribution symmetric about \( 0 \), the limit must be \( 0 \), by (2.8b); if the latter have a distribution symmetric about \( \theta \), the limit must be \( \theta \), by (2.8a).

Let \( f = (f_1, f_2, \ldots) \) be a sequence of location statistics and \( X = (X_1, X_2, \ldots) \) a sequence of random variables. Then \( X \) is \( f \)-symmetric provided the distribution of \( X \_1 - f_m(X_1, \ldots, X_m) \) is symmetric about \( 0 \), for all \( m \). And \( X \) is conditionally \( f \)-symmetric provided that for every \( n \), given \( X_1, \ldots, X_n \), the sequence \( X_{n+1}, X_{n+2}, \ldots \), is \( f \)-symmetric.

(2.9) Theorem. Let \( f = (f_1, f_2, \ldots) \) be a consistent sequence of location statistics, and \( X = (X_1, X_2, \ldots) \) 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 $\sigma$-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 0, the limit of $f_n(X_1, \ldots, X_n)$.

3. Examples.

(3.1) Example. 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 (1.2).

Construction. Let $Z = (Z_1, Z_2, \ldots)$ be a sequence of independent random variables, with a common distribution unsymmetric about 0. Let $X = Z$ or $-Z$ with probability $1/2$. 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 (1.2) involved the following notion: $X$ is string symmetric if the distribution of $a_1X_1 + \ldots + a_mX_m$ is symmetric about 0 for each $m \geq a$ and each string $a_1, \ldots, a_m$ of real numbers with $a_1 + \ldots + a_m = 0$. And $X$ is conditionally string symmetric if for each $n$, given $X_1, \ldots, X_n$, the sequence $X_{n+1}, \ldots, X_{n+2}, \ldots$ 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 construction is involved, viz.
\[ a_1 = 1, \quad a_2 = -\frac{1}{m_1}, \ldots, \quad a_m = -\frac{1}{m-1}. \]

Of course, (2.2) 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 (3.4) 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).

\textbf{(3.2) Lemma.} 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 \( [A,A+1] \).

\textbf{Proof.} The random variable will have a probability density of the form

\[ f = c(f_1 + \delta f_2) \]

where the function \( f_1 \) 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 \( \hat{f} \) 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 \(\varepsilon\) with

\[
|x|^2^k f_1(x) \geq \varepsilon \quad \text{for all } x \text{ with } |x| \geq 1.
\]

(3.3)

The argument uses the periodicity of the sine function, and the irrationality of \(\pi\); 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 \(C_\infty\) purely imaginary function of the real argument \(t\), vanishing except when \(A < |t| < A+1\), and satisfying \(\psi(-t) = -\psi(t)\). Let \(f_2\) be the inverse Fourier transform
of $\psi$. Then $f_2$ is real because $\psi$ is odd, and integrating by parts $j$ times shows in the usual way that

$$\sup_x |x|^j |f_2(x)| < \infty$$

for any $j \geq 1$.

From this and (3.3), the existence of $\delta$ follows. Plainly, there are $2^k - 2$ moments.

(3.4) Example. For each $m \geq 2$ and $N \geq 1$, there is a sequence $X_1, X_2, \ldots$, of independent identically distributed random variables such that:

i) $X_1$ has mean 0 and finite $N$th moments

ii) $X_1$ is not symmetric

iii) if $a_1 + \ldots + a_m = 0$, then the distribution of $a_1X_1 + \ldots + a_mX_m$

is symmetric about 0.

Proof. Use lemma (3.2), with $A > 2m$. Let $X_1, X_2, \ldots$, have the characteristic function $\phi$ constructed there. What must be shown is that $\sum_{j=1}^m t_j = 0$ entails

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

(3.4)

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 \( S = J+K \). But \( J = K \). For if, e.g., \( J > K \), then

\[
\sum_{j \in S} t_j > JA - K(A+1) \\
\geq A - K \\
\geq A - m ;
\]

but

\[
\left| \sum_{j \not\in S} t_j \right| < m
\]

because \( j \in S \) entails \( |t_j| < 1 \) by assumption; finally,

\[
\sum_j t_j > A-2m > 0
\]

because \( A > 2m \). This contradiction completes the proof.

The characteristic function constructed in lemma (3.2) 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 \( \phi \) 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: \( |t| < 1 \), 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 $\phi$ is not symmetric.

References


