BULL. AUSTRAL. MATH. SOC.
VOL. 25 (1982), 327-335.

62F03, 62F12
(62M99)

# SOME RESULTS CONCERNING SYMMETRIC DISTRIBUTIONS

SÁNDOR CSÖRGŐ AND C.R. HEATHCOTE

The purpose of this note is to establish results of a technical nature concerning a stochastic process that appears to be useful in the study of certain problems in statistical inference. These problems concern a test for symmetry, a method for obtaining an adaptive estimator of the centre of symmetry, and the detection of outliers with respect to the normal distribution. Details of the applications will be presented elsewhere.

## 1. Statement of the problem

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed real random variables with common non-degenerate distribution function $F$ and characteristic function

$$c(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) = u(t) + iv(t) .$$

Our point of departure is the following simple fact:

THEOREM 1. *$F$ is symmetric about the constant $\theta$ if and only if*

$$v(t)/u(t) = \tan \theta t \tag{1.1}$$

*holds for all* $t \neq \theta^{-1}\pi(k+\frac{1}{2})$, $k = 0, \pm 1, \pm 2, \ldots$ .

Received 30 October 1981. This work was done whilst the first author was a Visiting Fellow, Department of Statistics, The Faculties, Australian National University.

Proof. Write

$$c(t) = \{\exp(it\theta)\}\{u_\theta(t)+iv_\theta(t)\}$$

with $u_\theta(t) = E\cos\{t(X-\theta)\}$ , $v_\theta(t) = E\sin\{t(X-\theta)\}$ . If $X$ is symmetric about $\theta$ then $X-\theta$ is symmetric about zero, in which case $v_\theta(t) \equiv 0$ . Hence $u(t) = u_\theta(t)\cos t\theta$ , $v(t) = u_\theta(t)\sin t\theta$ and we obtain (1.1). Conversely, (1.1) implies

$$\begin{aligned} 0 &= v(t)\cos t\theta - u(t)\sin t\theta \\ &= E[\sin tX\cos t\theta - \cos tX\sin t\theta] \\ &= E\sin\{t(X-\theta)\} = v_\theta(t) . \end{aligned}$$

That is, $X-\theta$ is symmetric about zero and hence $X$ is symmetric about $\theta$ .

This means that the function

$$\theta(t) = t^{-1}\arctan\{v(t)/u(t)\} , \quad t \neq 0 , \tag{1.2}$$

is constant, $\theta$ say, if and only if the distribution is symmetric about $\theta$ (we take the principal value). Therefore, given an independent identically distributed sample of $n$ , its empirical counterpart

$$\tilde{\theta}_n(t) = t^{-1}\arctan\{V_n(t)/U_n(t)\} , \quad t \neq 0 , \tag{1.3}$$

could play a useful role in deciding whether or not the sample comes from a symmetric distribution.

Here $U_n(t)$, $V_n(t)$ are the real and imaginary parts respectively of the empirical characteristic function

$$C_n(t) = n^{-1}\sum_{j=1}^{n} e^{itX_j} = U_n(t) + iV_n(t) .$$

Let $S$ be any compact set of the real line not containing the origin and on which $u(t)$ does not vanish. Typically $S$ sill be of the form

$$S = [-b, -a] \cup [a, b]$$

where $a > 0$ is arbitrarily small and $b > a$ is sufficiently small. A consequence of Theorem 2.1 of Feuerverger and Mureika [5] and the uniform continuity of the arctan function is that, as $n \to \infty$ ,

https://doi.org/10.1017/S0004972700005402 Published online by Cambridge University Press

$$\sup_{t\in S} |\tilde{\theta}_n(t)-\theta(t)| \to 0 \quad \text{almost surely.} \tag{1.4}$$

That is, $\tilde{\theta}_n(t)$ is a strong uniform estimator of $\theta(t)$. Hence the constant or non-constant behaviour of $\tilde{\theta}_n(t)$ is indicative of symmetry, or lack of it, in the underlying distribution. If the distribution is indeed symmetric about some point then, for large samples, $\tilde{\theta}_n(t)$ is approximately constant and the value of this constant is an estimate of the centre of symmetry.

Next, suppose the sample members $X_1, X_2, \ldots, X_n$ are symmetrically distributed about some point $\theta$ which is estimated by minimising

$$n^{-1} \sum_{j=1}^{n} t^{-2}[1-\cos\{t(X_j-\theta)\}] , \quad |t| \le b , \quad t \neq 0 .$$

The resulting estimator $\theta_n^*(t)$ satisfies

$$n^{-1} \sum_{j=1}^{n} t^{-1} \sin\{t(X_j-\theta)\} = 0 , \quad |t| \le b , \quad t \neq 0 .$$

As $t \to 0$ this equation becomes $n^{-1} \sum_{1}^{n} (X_j-\theta) = 0$ and

$\theta_n^*(0) = n^{-1} \sum_{1}^{n} X_i = \overline{X}_n$, the sample mean. Also

$$n^{-1} \sum_{j=1}^{n} \{\sin tX_j \cos t\theta_n^*(t) - \cos tX_j \sin t\theta_n^*(t)\} = 0 .$$

Hence

$$\tan t\theta_n^*(t) = V_n(t)/U_n(t)$$

and $\theta_n^*(t)$ is the same as $\tilde{\theta}_n(t)$ of (1.3). We then have an interpretation of $\tilde{\theta}_n(t)$ as the quantity minimising a version of the sample circular variance, a point briefly discussed in the regression context by Heathcote [6]. It will be shown in Section 3 that the so-called variance function associated with $\theta_n^*(t)$ provides a means for detecting outliers from normality.

https://doi.org/10.1017/S0004972700005402 Published online by Cambridge University Press

## 2. The process $T_n(t) = n^{\frac{1}{2}}\{\tilde{\theta}_n(t)-\theta(t)\}$

We first require a rate result for $\{\tilde{\theta}_n(t),\ t \in S\}$. Let

$$\phi(t) = [1-u(t)]^{\frac{1}{2}} ,$$

$$m(y) = \lambda\{t \in (-\tfrac{1}{2}, \tfrac{1}{2}) : \phi(t) < y\} ,$$

where $\lambda$ denotes Lebesgue measure, and introduce the nondecreasing rearrangement of $\phi$ as

$$\overline{\phi}(h) = \sup[y : m(y) < h] .$$

THEOREM 2. *If*

$$\int_0 \frac{\overline{\phi}(h)}{h(\log 1/h)^{\frac{1}{2}}}\, dh < \infty \tag{2.1}$$

*then*

$$\limsup_{n\to\infty} \left(\frac{n}{\log\log n}\right)^{\frac{1}{2}} \sup_{t\in S} |\tilde{\theta}_n(t)-\theta(t)| \le k \ , \ \textit{almost surely,}$$

*where* $k > 0$ *is a finite constant.*

Proof. A one term Taylor expansion for the arctan function gives

$$\tilde{\theta}_n(t) - \theta(t) = A_n(t)t^{-1}/\left[1+\{v(t)/u(t)\}^2\right] - \psi(t)A_n^3(t)t^{-1}/\left[1+\{\psi(t)A_n(t)\}^2\right]^2$$

where

$$A_n(t) = \frac{V_n(t)}{U_n(t)} - \frac{v(t)}{u(t)}$$

and $0 \le \psi(t) \le 1$. A simple algebraic manipulation and the first law of the iterated logarithm in Theorem 9.1 of Csörgő [3] implies that

$$\sup_{t\in S} |A_n(t)| = O\left[\left[\frac{\log\log n}{n}\right]^{\frac{1}{2}}\right] \ , \text{ almost surely.} \tag{2.2}$$

The desired result follows since $t$ is bounded away from zero in $S$.

Observe that condition (2.1) is generally not satisfied if it is only true that

https://doi.org/10.1017/S0004972700005402 Published online by Cambridge University Press

$$\int_{-\infty}^{\infty} \log^{+}(|x|)dt(x) < \infty ,$$

but is satisfied if for arbitrary $\varepsilon > 0$ we have

$$\int_{-\infty}^{\infty} \{\log^{+}(|x|)\}^{1+\varepsilon}d(x) < \infty .$$

Thus (2.1) holds in most situations of practical interest. A finer analysis is given in Csörgő [2].

Our main objective is to establish the asymptotic normality of the process

$$T_n(t) = n^{\frac{1}{2}}\{\tilde{\theta}_n(t)-\theta(t)\} , \quad t \in S .$$

If the underlying distribution is symmetric about $\theta$, that is $\theta(t) \equiv \theta$, then this result will enable us to show the asymptotic normality of an adaptive estimator $\tilde{\theta}_n(t_n)$ for $\theta$ within the class $\{\tilde{\theta}_n(t), t \in S\}$.

Let $C(S)$ be the Banach space of continuous functions defined on $S$, endowed with the supremum norm, and consider a Gaussian process $\{T(t), t \in S\}$ with zero mean and covariance

$$ET(t)T(s) = \frac{h(t,s)}{ts|c(t)|^2|c(s)|^2}$$

with

$$h(t, s) = [u(t-s)\{u(t)u(s)+v(t)v(s)\}+u(t+s)\{v(t)v(s)-u(t)u(s)\}$$
$$+v(t-s)\{v(t)u(s)-u(t)v(s)\}-v(t+s)\{u(t)v(s)+v(t)u(s)\}]/2 .$$

THEOREM 3. *$T_n(\cdot)$ converges weakly in $C(S)$ to $T(\cdot)$ if and only if (2.1) holds.*

Proof. Commencing exactly as in the proof of Theorem 2, we find

$$T_n(t) = n^{\frac{1}{2}}B_n(t) + R_n(t) ,$$

where

$$B_n(t) = \frac{u(t)\{V_n(t)u(t)-U_n(t)v(t)\}}{tU_n(t)\{u^2(t)+v^2(t)\}}$$

https://doi.org/10.1017/S0004972700005402 Published online by Cambridge University Press

and, on applying (2.2),

$$\sup_{t\in S} |R_n(t)| = O|n^{-1}(\log\log n)^{3/2}| \text{ , almost surely.}$$

Since $u(t)/U_n(t)$ converges almost surely uniformly to 1 on $S$, the weak limit of $T_n$ in $C(S)$, if it exists, coincides with that of

$$D_n(t) = \{v(t)\ \mathrm{Re}\ Y_n(t) - u(t)\ \mathrm{Im}\ Y_n(t)\}/\{t|c(t)|^2\}$$

where

$$Y_n(t) = n^{\frac{1}{2}}\{C_n(t)-c(t)\} .$$

From Theorem 1 of Marcus [7] (*cf.* also Theorem 3.1 of Csörgö [3]), $D_n(\cdot)$ converges weakly to $T(\cdot)$ if and only if (2.1) holds, and this yields the desired result.

We define the *variance function* $\sigma^2(t)$ associated with $\tilde{\theta}_n(t)$ to be the variance of the limit process

$$(2.3)\quad \sigma^2(t) = ET^2(t) = \frac{u^2(t)\{1-u(2t)\}-2u(t)v(t)v(2t)+v^2(t)\{1+u(2t)\}}{2t^2\{u^2(t)+v^2(t)\}^2}$$

$$= \mathrm{var}\left\{\frac{u(t)\sin tX-v(t)\cos tX}{t\{u^2(t)+v^2(t)\}}\right\}$$

$$= (2t^2)^{-1}g\{u(t),\ v(t),\ u(2t),\ v(2t)\} .$$

This variance function coincides exactly with that of the functional least squares estimator of Chambers and Heathcote [1], Heathcote [6].

## 3. The variance function

Suppose $X$ is symmetric about $\theta$. If the reasonable criterion of minimum asymptotic variance is accepted then the optimum estimator obtained by the above method is $\tilde{\theta}_n(t_0)$, where $t_0$ minimises $\sigma^2(t)$. Theorem 2 of Chambers and Heathcote [1] shows that in the class $\mathcal{D}$ of distributions which are normal, or lack a variance, or are leptokurtic, the normal is the only one for which $t_0 = 0$; that is the sample mean $\overline{X}_n$ is optimal. For

https://doi.org/10.1017/S0004972700005402 Published online by Cambridge University Press

other members of $\mathcal{D}$ the value of $t_0$ is strictly non-zero.

This characterisation of the normal distribution in terms of the variance function also leads to a method for detecting outliers with respect to the normal, similar to that described for regression models by Chambers and Heathcote [1]. The procedure is the following. Assume $X$ has a distribution in the above class $\mathcal{D}$ of possibly long-tailed distributions, and suppose $\sigma_n^2(t)$ is an estimate of $\sigma(t)$. Plot $\sigma_n^2(t)$ and determine $t_n$ at which it achieves a minimum. If $t_n = 0$ we infer that $X$ is normally distributed; otherwise delete a suspected outlier and estimate the minimising $t_{n-1}$ for the reduced sample of $n - 1$. Continue the process until the origin is estimated to be the minimising value of $\sigma^2(t)$. The deleted sample members are then classified as outliers with respect to the remaining ones, inferred to be generated by an underlying normal distribution.

In the case of symmetry, the variance function of (2.3) reduces to

$$\sigma^2(t) = \{1-u_\theta(2t)\}/\left\{2t^2u_\theta^2(t)\right\}$$

with $u_\theta(t) = E\cos\{t(X-\theta)\}$. In practice $\sigma^2(t)$ is unknown, and can be estimated by

$$\sigma_n^2(t) = (2t^2)^{-1}g\{U_n(t),\ V_n(t),\ U_n(2t),\ V_n(2t)\}\ .$$

It follows from the almost sure uniform consistency of the empirical characteristic function $C_n(t)$ that, as $n \to \infty$,

$$\sup_{t\in S}\left|\sigma_n^2(t)-\sigma^2(t)\right| \to 0 \quad \text{almost surely.} \tag{3.1}$$

Thus if $t_n$ minimises $\sigma_n^2(t)$ we take $\tilde{\theta}_n(t_n)$ as estimator of $\theta$. Note that $\sigma^2(t)$, $\sigma_n^2(t)$ are symmetric about the origin so that we need consider only $t \geq 0$.

To formalise this procedure suppose $t_0 > 0$ is given by

https://doi.org/10.1017/S0004972700005402 Published online by Cambridge University Press

$$t_0 = \inf\left\{s : \sigma^2(s) = \inf_{0\le t} \sigma^2(t)\right\},$$

and let $S_0 = [a_0, b_0]$ be an interval such that $0 < a_0 < t_0 < b_0$ and $\sigma^2(t_0) < \sigma^2(t)$ for any other $t$ in $S_0$. Let

$$t_n = \inf\left\{s : \sigma_n^2(s) = \inf_{t\in S_0} \sigma_n^2(t)\right\}.$$

The argument of Lemma 5 of Csörgő [4] shows that, as $n \to \infty$,

$$t_n \to t_0 \quad \text{almost surely.}$$

Then from (1.4), $\tilde{\theta}_n(t_n) \to \theta$ almost surely, and imitating the proof of Theorem 4 of [4] we obtain the following:

THEOREM 4. *If condition* (2.1) *is satisfied then*

$$\lim_{n\to\infty} \Pr\{T_n(t_n) \le x\} = \Phi\{x/\sigma(t_0)\}$$

*for all real* $x$, *where* $\Phi(\cdot)$ *denotes the standard normal distribution function.*

Observe that if the underlying distribution has finite variance then $S$ and $S_0$ may be extended to include the origin.

## References

[1] R.L. Chambers and C.R. Heathcote, "On the estimation of slope and the identification of outliers in linear regression", *Biometrika* 68 (1981), 21-33.

[2] Sándor Csörgő, "Limit behaviour of the empirical characteristic function", *Ann. Probab.* 9 (1981), 130-144.

[3] Sándor Csörgő, "Multivariate empirical characteristic functions", *Z. Wahrsch. Verw. Gebiete* 55 (1981), 203-229.

[4] Sándor Csörgő, "The theory of functional least squares", *J. Austral. Math. Soc. Ser. A* (to appear).

[5] Andrey Feuerverger and Roman A. Mureika, "The empirical characteristic function and its applications", *Ann. Statist.* 5 (1977), 88-97.

[6] C.R. Heathcote, "Linear regression by functional least squares", *Essays in Statistical Science*. *J. Appl. Probab.* (to appear).

[7] Michael B. Marcus, "Weak convergence of the empirical characteristic function", *Ann. Probab.* 9 (1981), 194-201.

József Attila Tudományegyetem,
Bolyai Intézet,
6720 Szeged,
Aradi vértamik tere 1,
Hungary;

Department of Statistics,
The Faculties,
Australian National University,
PO Box 4,
Canberra,
ACT 2600,
Australia.