The Annals of Mathematical Statistics
1970, Vol. 41, No. 4, 1223–1226

A NOTE ON SYMMETRIC BERNOULLI RANDOM VARIABLES[1]

BY MORRIS L. EATON

University of Chicago

Consider independent random variables $X_1, X_2, \cdots$ such that X_i takes the values ± 1 each with probability $\frac{1}{2}$. If $\theta = (\theta_1, \cdots, \theta_n)$ satisfies $\sum_{i=1}^n \theta_i^2 = 1$, let $S_n(\theta) = \sum_1^n \theta_i X_i$ and $S_n = n^{-\frac{1}{2}} \sum_1^n X_i$. Recently, Efron (1969) has shown that $E(S_n(\theta))^{2k} \leqq ES_n^{2k}$ for $k = 1, 2, \cdots$ and for all n. In the present note, sufficient conditions on a continuously differentiable function f are given so that $Ef(S_n(\theta)) \leqq Ef(S_n)$ for all n. This result is then used to derive probability bounds related to results of Hoeffding (1963).

DEFINITION 1. Let $a = (a_1, \cdots, a_m)$ and $b = (b_1, \cdots, b_m)$ be real vectors. Reorder the components of a and b such that $a_{i_1} \geqq a_{i_2} \geqq \cdots \geqq a_{i_m}$ and $b_{j_1} \geqq b_{j_2} \geqq \cdots \geqq b_{j_m}$. Then, *a majorizes b* if and only if $\sum_{\alpha=1}^k a_{i_\alpha} \geqq \sum_{\alpha=1}^k b_{j_\alpha}$ for $k = 1, \cdots, m-1$ and $\sum_{\alpha=1}^m a_{i_\alpha} = \sum_{\alpha=1}^m b_{j_\alpha}$.

DEFINITION 2. A real-valued function φ defined on an open subset of R^n which has continuous first partial derivatives is called a Schur function if

$$\frac{\partial \varphi}{\partial x_i} - \frac{\partial \varphi}{\partial x_j} \geqq 0 \tag{1}$$

when $x_i > x_j$ for $i, j = 1, \cdots, n$ and x in the domain of φ.

A result which relates Schur functions and one vector majorizing another is

THEOREM. *Let C be an open symmetric convex set in R^n and suppose φ is a Schur function on C which is a symmetric function of its arguments. If $a \in C$ majorizes $b \in C$, then $\varphi(a) \geqq \varphi(b)$.*

For a proof of this theorem, see Schur (1923) and Ostrowski (1952).

Now, let F be the set of all functions f on R to R which are continuously differentiable and satisfy

$$t^{-1}[f'(t+\Delta) - f'(-t+\Delta) + f'(t-\Delta) - f'(-t-\Delta)] \tag{2}$$

is non-decreasing in t for $t > 0$ and $\Delta \geqq 0$. Note that $f \in F$ is equivalent to

$$t^{-1} E_W[f'(t+W) - f'(-t+W)] \tag{3}$$

Received February 13, 1969.

[1] This research was supported in part by Research Grant No. NSF 8026 from the Division of Mathematical, Physical, and Engineering Sciences of the National Science Foundation, and in part by the Department of Statistics, University of Chicago, under partial sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

is non-decreasing for $t > 0$ for all bounded symmetric random variables W. Also, a useful sufficient condition that $f \in F$ is that

$$t^{-1}[f'(t+\Delta)-f'(-t+\Delta)] \tag{4}$$

be non-decreasing in t for $t > 0$ and for all real Δ.

THEOREM 1. *If* $f \in F$ *and if* $(\theta_1{}^2, \cdots, \theta_n{}^2)$ *majorizes* $(\xi_1{}^2, \cdots, \xi_n{}^2)$, *then*

$$Ef(S_n(\theta)) \leqq Ef(S_n(\xi)). \tag{5}$$

PROOF. Define g on $\{\eta \mid \eta = (\eta_1, \cdots, \eta_n); \eta_i > 0\}$ by

$$g(\eta) = Ef(\textstyle\sum_1^n \eta_i{}^{\frac{1}{2}} X_i). \tag{6}$$

It will be shown that $-g$ is a Schur function. Since g is a symmetric function of its arguments, it is sufficient to verify (1) for $i = 2$ and $j = 1$. However,

$$\frac{\partial g}{\partial \eta_1} - \frac{\partial g}{\partial \eta_2} = (2(\eta_1 \eta_2)^{\frac{1}{2}})^{-1} E(X_1 \eta_2{}^{\frac{1}{2}} - X_2 \eta_1{}^{\frac{1}{2}}) f'(\eta_1{}^{\frac{1}{2}} X_1 + \eta_2{}^{\frac{1}{2}} X_2 + W) \tag{7}$$

where $W = \sum_{i=3}^n \eta_i{}^{\frac{1}{2}} X_i$ is a bounded symmetric random variable. Computing the expectation on (X_1, X_2), we find after some manipulation that

$$\frac{\partial g}{\partial \eta_1} - \frac{\partial g}{\partial \eta_2} = \frac{(\eta_1{}^{\frac{1}{2}} + \eta_2{}^{\frac{1}{2}})(\eta_2{}^{\frac{1}{2}} - \eta_1{}^{\frac{1}{2}})}{8(\eta_1 \eta_2)^{\frac{1}{2}}} E\left\{\frac{f'(\eta_1{}^{\frac{1}{2}} + \eta_2{}^{\frac{1}{2}} + W) - f'(-\eta_1{}^{\frac{1}{2}} - \eta_2{}^{\frac{1}{2}} + W)}{\eta_1{}^{\frac{1}{2}} + \eta_2{}^{\frac{1}{2}}} - \frac{f'(\eta_2{}^{\frac{1}{2}} - \eta_1{}^{\frac{1}{2}} + W) - f'(-\eta_2{}^{\frac{1}{2}} + \eta_1{}^{\frac{1}{2}} + W)}{\eta_2{}^{\frac{1}{2}} - \eta_1{}^{\frac{1}{2}}}\right\} \tag{8}$$

which is non-negative for $\eta_2 > \eta_1$ since $f \in F$. Hence $-g$ is a Schur function. Since the distribution of $\sum_{i=1}^n \theta_i X_i$ is the same as the distribution of $\sum_{i=1}^n |\theta_i| X_i$, the theorem holds for any vectors $\theta = (\theta_1, \cdots, \theta_n)$ and $\xi = (\xi_1, \cdots, \xi_n)$ which have no components zero when $(\theta_1{}^2, \cdots, \theta_n{}^2)$ majorizes $(\xi_1{}^2, \cdots, \xi_n{}^2)$. However, when some of the θ_i or ξ_i are zero, the conclusion follows from the continuity of f when $(\theta_1{}^2, \cdots, \theta_n{}^2)$ majorizes $(\xi_1{}^2, \cdots, \xi_n{}^2)$. This completes the proof.

COROLLARY 1. *If* $f \in F$, *then*

$$Ef(S_n(\theta)) \leqq Ef(S_n) \tag{9}$$ *and*

$$Ef(S_{n-1}) \leqq Ef(S_n). \tag{10}$$

PROOF. If $\eta_i \geqq 0$, $i = 1, \cdots, n$ are such that $\sum_1^n \eta_i = 1$, then the vector $(\eta_1, \cdots, \eta_n)$ majorizes $(1/n, \cdots, 1/n)$ and (9) follows. Choosing $\theta = ((n-1)^{-\frac{1}{2}}, \cdots, (n-1)^{-\frac{1}{2}}, 0)$ in (9) shows (10) holds.

EXAMPLE 1. To obtain Efron's (1969) result, let $f(x) = x^{2k}$ for k a positive integer. That $f \in F$ is immediate from (4) and Theorem 1 holds.

EXAMPLE 2. Consider $f(x) = e^{ax}$ for $a \neq 0$. Again, verification of (4) is immediate so $f \in F$. From this example, it is clear that the function $e^{bx} + e^{-bx}, b \neq 0$ is also in F.

THEOREM 2. *If $f \in F$ and if there exists a $\delta > 0$ and a constant M such that $E|f(S_n)|^{1+\delta} \leqq M$ for all n, then*

$$Ef(S_n(\theta)) \leqq Ef(Z) \tag{11}$$

where Z has a unit normal distribution, and we assume that $E(|f(Z)|) < +\infty$.

PROOF. The continuity of f implies that $f(S_n)$ converges in distribution to $f(Z)$. Since $E|f(S_n)|^{1+\delta} \leqq M$ for all n, $E|f(Z)| < +\infty$, and $f(S_n)$ converges in distribution to $f(Z)$, we conclude that $Ef(S_n)$ converges to $Ef(Z)$. From Corollary 1, $Ef(S_n)$ is a non-decreasing sequence and the conclusion follows.

The next result extends Theorem 2 to symmetric random variables taking values in $[-1, 1]$. Let $Y_1, \cdots, Y_n$ be independent symmetric random variables such that $|Y_i| \leqq 1$. It is clear that there exist independent random variables $V_1, \cdots, V_n$ and $X_1, \cdots, X_n$ such that $0 \leqq V_i \leqq 1$, X_i takes the values ± 1 each with probability $\frac{1}{2}$, and the distribution of Y_i is that of $V_i X_i$. If $\theta = (\theta_1, \cdots, \theta_n)$ satisfies $\sum \theta_i^2 = 1$, let $T_n(\theta) = \sum_1^n \theta_i Y_i$. Further, let

$$U_i = \frac{\theta_i V_i}{(\sum \theta_i^2 V_i^2)^{\frac{1}{2}}} \quad \text{if} \quad \sum \theta_i^2 V_i^2 > 0,$$
$$= 0 \qquad \text{otherwise;}$$

and note that

$$T_n(\theta) = (\textstyle\sum \theta_i^2 V_i^2)^{\frac{1}{2}} \sum_{i=1}^n U_i X_i. \tag{12}$$

For $c \geqq 0$ and $f \in F$, let f_c be defined by $f_c(x) \equiv f(cx)$. Obviously, $f_c \in F$ for $c > 0$ when $f \in F$.

THEOREM 3. *Let $f \in F$ and assume*

(i) *for each $c \in (0, 1]$, f_c satisfies the assumptions of Theorems 1 and 2*

(ii) *for each $c \in [0, 1]$, $Ef_c(Z) \leqq Ef(Z)$. Then*

$$Ef(T_n(\theta)) \leqq Ef(Z) \tag{13}$$

where Z has a unit normal distribution.

PROOF. From (12)

$$\begin{aligned} Ef(T_n(\theta)) &= Ef((\textstyle\sum \theta_i^2 V_i^2)^{\frac{1}{2}} \sum U_i X_i) \\ &= E[E(f((\textstyle\sum \theta_i^2 V_i^2)^{\frac{1}{2}} \sum U_i X_i) \mid V_1, \cdots, V_n)] \\ &\leqq E[E(f((\textstyle\sum \theta_i^2 V_i^2)^{\frac{1}{2}} Z) \mid V_1, \cdots, V_n)] \leqq Ef(Z). \end{aligned} \tag{14}$$

The first inequality follows from the application of Theorem 2 to f_c with $c = (\sum \theta_i^2 V_i^2)^{\frac{1}{2}} \leqq 1$ and the second inequality follows from assumption (ii). This completes the proof.

COROLLARY 2. *Let $\alpha > 0$ and assume $f \in F$ satisfies the assumptions of Theorem* 3. *If $f \geqq 0$ is symmetric and if $f(x) \geqq 1$ for $|x| \geqq \alpha$, then*

$$P\{T_n(\theta) \geqq \alpha\} \leqq \tfrac{1}{2}Ef(Z). \tag{15}$$

PROOF. From the assumptions on f,

$$P\{T_n(\theta) \geqq \alpha\} = \tfrac{1}{2}P\{|T_n(\theta)| \geqq \alpha\} \leqq \tfrac{1}{2}Ef(T_n(\theta)) \leqq \tfrac{1}{2}Ef(Z),$$

the last inequality following from Theorem 3.

EXAMPLE 3. Let $f(x) = (e^{\alpha x} + e^{-\alpha x})/(e^{\alpha^2} + e^{-\alpha^2})$. Application of (15) yields

$$P\{T_n(\theta) \geqq \alpha\} \leqq \frac{e^{-\frac{1}{2}\alpha}}{1 + e^{-2\alpha^2}} \tag{16}$$

for $\alpha > 0$. (16) is useless for small values of α as the bound is greater than $\frac{1}{2}$. Let α_0 be the unique positive solution to $2 = e^{\frac{1}{2}\alpha^2} + e^{-3\alpha^2/2}$. Then for $\alpha > \alpha_0$, (16) is less than $\frac{1}{2}$.

EXAMPLE 4. A somewhat more complicated bound can be given by choosing $f(x) = (e^{h\alpha x} + e^{-h\alpha x} - 2)/(e^{h\alpha^2} + e^{-h\alpha^2} - 2)$, $h \neq 0$. (14) then yields

$$P\{T_n(\theta) \geqq \alpha\} \leqq \min_h \frac{e^{\frac{1}{2}h^2\alpha^2} - 1}{e^{h\alpha^2} + e^{-h\alpha^2} - 2}. \tag{17}$$

By setting $h = 1$, the bound in (17) is a uniform improvement on (16) for $\alpha > \alpha_0$.

The inequalities (16) and (17) are related to results by Okamoto (1958) and Hoeffding (1963). It seems likely that uniform improvements on (17) are possible by a more clever choice of f in (15). However the author has been unable to do so.

Acknowledgment. I wish to thank the referee for many helpful suggestions which greatly improved this paper. In particular, the referee's comments led to Theorem 3 and to the inequality (17).

REFERENCES

[1] EFRON, B. (1969). Student's *t*-test under non-normal conditions. Harvard Univ. Technical Report No. 21.

[2] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[3] OKAMOTO, M. (1958). Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.* **10** 29–35.

[4] OSTROWSKI, A. (1952). Sur quelques applications des fonctions convexes et concave au sens de I. Schur. *J. Math. Pures Appl.* **31** 253–292.

[5] SCHUR, I. (1923). Uker eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. *Sitzber. Berl. Math. Ges.* **22** 9–20.