A ROBUST TEST FOR OMNIBUS ALTERNATIVES

GUTTI JOGESH BABU\textsuperscript{1,*} and A. R. PADMANABHAN\textsuperscript{2}

\textsuperscript{1}Department of Statistics, Pennsylvania State University, University Park, PA 16802, U.S.A.
\textsuperscript{2}Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia

ABSTRACT

For simultaneous testing for differences in location, scale, symmetry (or skewness) and tailweight between two unknown continuous distribution functions, a statistic $V$ is proposed. Its quantiles are estimated by the bias-corrected bootstrap. Monte Carlo studies show that the $V$-test tends to be more robust than its competitor. Asymptotics to justify the use of the bootstrap are presented.

The methodology is illustrated by testing simultaneously for differences in location, scale, symmetry (or skewness) and tailweight between the amounts of rainfall in February and August during 1950–1979 in New York’s Central Park.

1. INTRODUCTION

Consider the problem of testing for the equality of two continuous distribution functions $F_1$ and $F_2$. Suppose that an omnibus test such as the Kolmogorov–Smirnov rejects the equality of $F_1$ and $F_2$, thereby suggesting the difference is statistically significant. It is hard to interpret this significance (cf. (Boos, 1986, p. 1018)).

Therefore it is worthwhile to devise a simultaneous test of differences in four important characteristics, viz.: location, scale, symmetry (or skewness) and tailweight. Such a test is studied in (Boos, 1986, pp. 1018–1025). However, its practical applications are typically confined to symmetric populations and even here problems arise with his skewness and kurtosis statistics. This paper proposes a statistic $V$ for testing simultaneously such omnibus alternatives, which is free of the above limitations.

* Supported by NSF Grant DMS-9208066.
Section 2 describes the test. Section 3 contains the Monte Carlo studies and also explains how our procedure is more robust than its competitor. Section 4 applies the methodology to test simultaneously for differences in location, scale, symmetry (or skewness) and tailweight in the amounts of rainfall in February and August during 1950–1979 in New York’s Central Park. Section 5 details the asymptotics to justify the use of the bootstrap.

2. THE TEST

Let $Z_1 = (Z_{11}, Z_{12}, \ldots, Z_{1n_1})$ and $Z_2 = (Z_{21}, Z_{22}, \ldots, Z_{2n_2})$ be independent samples from continuous distribution functions $F_1$ and $F_2$. We first describe a series of statistics to bring out various differences between the two populations. For $i = 1, 2$, let $U_{\gamma,i}$ and $L_{\gamma,i}$ denote the averages of the top $[n_i\gamma]$ and bottom $[n_i(1-\gamma)]$ observations in the sample. The tailweight of $F_i$ (cf. (Hogg et al., 1975)) can be estimated by

$$Q_{2,i} = \frac{U_{0.05,i} - L_{0.05,i}}{U_{0.5,i} - L_{0.5,i}}.$$  

Let $Z_{i,\text{med}}$ denote the median of the $Z_i$-sample and $Y_i = Z_i - Z_{i,\text{med}}$ denote the $Z_i$-sample values centered at the sample median. Write $G_{n,i}$ for the empirical distribution function of $Y_i$. Then $M_i$ denotes the modified Butler statistic

$$M_i = \sup_y |G_{n,i}(y) + G_{n,i}(-y) - 1|.$$  

(Butler (1969) works with the sample values centered at the population median, assuming the latter to be known.) Finally, let $Z_{i,0.10}$ and $s_{i,0.10}^2$ denote respectively the ten percent trimmed mean and ten percent trimmed variance corresponding to the $Z_i$-sample.

Write

$$V_1 = \frac{|Z_{1,0.10} - Z_{2,0.10}|}{1 + |Z_{1,0.10} - Z_{2,0.10}|},$$

$$V_2 = \frac{|s_{1,0.10}^2 - s_{2,0.10}^2|}{1 + |s_{1,0.10}^2 - s_{2,0.10}^2|},$$

$$V_3 = \frac{|M_1 - M_2|}{1 + |M_1 - M_2|}$$

and

$$V_4 = \frac{|Q_{2,1} - Q_{2,2}|}{1 + |Q_{2,1} - Q_{2,2}|}.$$
For \( i = 1, 2 \), let \( F_{i,0.10} \) denote the ten percent truncated distribution corresponding to \( F_i \). Let \( m_i \) and \( \sigma_i^2 \) be respectively the mean and variance of \( F_i \), while \( m_{i,0.10} \) and \( \sigma_{i,0.10}^2 \) have similar meanings in relation to \( F_{i,0.10} \). The statistics \( V_1 \) and \( V_2 \) are measures of differences between \( F_1 \) and \( F_2 \) regarding \( m_{1,0.10} \) and \( m_{2,0.10} \) and \( \sigma_{1,0.10}^2 \) and \( \sigma_{2,0.10}^2 \) respectively. For their greater robustness, \( m_{i,0.10} \) and \( \sigma_{i,0.10}^2 \) are chosen in preference to \( m_i \) and \( \sigma_i^2 \) (cf. (Bickel and Lehmann, 1975, 1976)). \( V_3 \) and \( V_4 \) are measures of differences between \( F_1 \) and \( F_2 \) regarding symmetry (or skewness) and tailweight respectively. Thus our test statistic

\[
V = \max (V_1, V_2, V_3, V_4)
\]

brings out the differences in location, scale, skewness and the tailweight. The exact null distribution of \( V \) is intractable. On the other hand, the bootstrap method is known to yield a better approximation than the one based on the normal approximation theory. Besides, bootstrap is known to correct for skewness of the sampling distribution (cf. (Babu and Singh, 1984a)). In this paper, we shall use the bias-corrected refinement (cf. (Efron and Tibshirani, 1986, p. 68)) of the percentile bootstrap method.

Let \( \hat{F}_{1,n_1} \) and \( \hat{F}_{2,n_2} \) be the empirical distribution functions corresponding to \( F_1 \) and \( F_2 \) respectively. Bootstrap samples are drawn separately from \( \hat{F}_{1,n_1} \) and \( \hat{F}_{2,n_2} \). Let \( \hat{G} \) be the resulting bootstrap distribution of \( V \). For \( 0 < q < 1 \), let \( v_q \) denote the \( q \)th quantile of the null distribution of \( V \). The bias-corrected bootstrap estimator of \( v_q \) is given by

\[
b_q = \hat{G}^{-1}(\Phi(2z + n_q)),
\]

where \( n_q \) denotes the \( q \)th quantile of the cumulative standard normal distribution function \( \Phi \) and \( z = \Phi^{-1}(\hat{G}(V)) \).

Remark 1. Note that the \( q \)th quantile \( p_q \) of \( \hat{G} \) is a consistent estimator of \( v_q \). Although for large samples, \( p_q \) and \( b_q \) give nearly the same result, \( b_q \) gives a somewhat better result for practical sample sizes. Hence, in all our Monte Carlo studies, only \( b_q \) is used.

3. THE MONTE CARLO STUDIES

Five thousand random samples of sizes \( n_1 = n_2 = n \) \((n = 20, 30)\) were drawn using IMSL subroutines on a VAX 23 computer from a number of distributions to be specified shortly. Following the recommendations of Efron and Tibshirani (1986, p. 72), one thousand bootstrap samples were drawn from each sample. The nominal level was always kept at five percent.

We use the notation \( B(m, n) \), \( N \), \( 4N \), \( N(2, 4^2) \), \( E \) and \( L \) to denote respectively, the Beta distribution with parameters \( m \) and \( n \), the standard normal, the normal distribution with mean zero and standard deviation four, the normal distribution with mean two and standard deviation four, the standard exponential distribution and the standard lognormal distribution. For \( \lambda = 5, 10 \) and 15, \( \lambda\%4N \) and \( \lambda\%N(2, 4^2) \) will stand for the mixtures \((1 - \lambda/100)N + (\lambda/100)4N \) and
Table 1.
Empirical levels (in percent) of the $V$-test $F_1 = F_2 = F$

<table>
<thead>
<tr>
<th>$F$</th>
<th>$n_1 = n_2 = 20$</th>
<th>$n_1 = n_2 = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B(3, 3)$</td>
<td>4.1</td>
<td>5.2</td>
</tr>
<tr>
<td>$N$</td>
<td>4.35</td>
<td>5.08</td>
</tr>
<tr>
<td>5%4$N$</td>
<td>4.36</td>
<td>4.62</td>
</tr>
<tr>
<td>10%4$N$</td>
<td>4.1</td>
<td>4.4</td>
</tr>
<tr>
<td>15%4$N$</td>
<td>3.9</td>
<td>4.2</td>
</tr>
<tr>
<td>$E$</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td>$L$</td>
<td>5.5</td>
<td>5.25</td>
</tr>
</tbody>
</table>

Table 2.
Empirical powers (in percent) $n_1 = n_2 = 20$

<table>
<thead>
<tr>
<th>Case</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>Location</th>
<th>Scale</th>
<th>Symmetry or Skewness</th>
<th>Tail-weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$N$</td>
<td>5%4$N$</td>
<td>2.4</td>
<td>5.2</td>
<td>2.7</td>
<td>8.8</td>
</tr>
<tr>
<td>2</td>
<td>$N$</td>
<td>5%$N(2.4^2)$</td>
<td>5.5</td>
<td>5.5</td>
<td>18.75</td>
<td>9.7</td>
</tr>
<tr>
<td>3</td>
<td>$N$</td>
<td>15%4$N$</td>
<td>3.1</td>
<td>14.6</td>
<td>3.15</td>
<td>14.1</td>
</tr>
<tr>
<td>4</td>
<td>$N$</td>
<td>15%$N(2.4^2)$</td>
<td>9.75</td>
<td>15.5</td>
<td>27.5</td>
<td>15.0</td>
</tr>
<tr>
<td>5</td>
<td>$N$</td>
<td>$E$</td>
<td>35.2</td>
<td>3.9</td>
<td>50.0</td>
<td>10.5</td>
</tr>
<tr>
<td>6</td>
<td>$N$</td>
<td>$B(3, 3)$</td>
<td>40.2</td>
<td>39.8</td>
<td>2.5</td>
<td>24.6</td>
</tr>
<tr>
<td>7</td>
<td>$N$</td>
<td>$B(6, 3)$</td>
<td>52.5</td>
<td>30.0</td>
<td>31.0</td>
<td>20.0</td>
</tr>
<tr>
<td>8</td>
<td>$E$</td>
<td>$L$</td>
<td>34.7</td>
<td>7.8</td>
<td>50.0</td>
<td>9.7</td>
</tr>
<tr>
<td>9</td>
<td>$E$</td>
<td>Chi(4)</td>
<td>85.0</td>
<td>79.1</td>
<td>30.0</td>
<td>8.4</td>
</tr>
</tbody>
</table>

$(1 - \lambda/100)4N + (\lambda/100)N(2, 4^2)$. Chi(4) will denote the chi-square distribution with four degrees of freedom.

Table 1 gives the empirical levels of the $V$-test for $n_1 = n_2 = 20$ and $n_1 = n_2 = 30$ respectively.

Table 1 shows the robustness of the $V$-test for practical sample size.

Table 2 shows that the $V$-test effectively picks up any significant difference between $F_1$ and $F_2$ regarding location, scale, symmetry (or skewness) and tailweight.

Let WILCOXON, MOOD, SKEW and KURT denote respectively the Wilcoxon and Mood tests, and tests for skewness and kurtosis based on appropriate linear rank statistics. Then for the (above) two-sample problem, Boos (1986, p. 1019) initially proposes

$$\text{GLOBE} = \text{WILCOXON} + \text{MOOD} + \text{SKEW} + \text{KURT}.$$ 

However, location differences would drastically affect the performance of MOOD, SKEW and KURT. Therefore, it is good to base these statistics on the samples
centered at the respective sample medians. Nevertheless, this centering upsets the
distribution-free property of MOOD, SKEW and KURT, although an asymptotic
distribution-free property holds for MOOD and KURT in the case of symmetric
populations. This effectively restricts the applicability of GLOBE to symmetric
populations. If skewness is suspect, Boos (1986, p. 1024, II column) recommends
transformation of the data.

In view of the drawbacks of GLOBE, Boos (1986) proposes

\[ \text{GLOBE}^* = \text{WILCOXON} + \text{MOOD}^* + \text{KURT}^* + \text{SKEW}, \]

where MOOD* and KURT* are respectively MOOD and KURT, based on the
samples centered at the respective sample medians. However, for reasons given
in the preceding paragraph, GLOBE* also applies only to symmetric populations
and even here problems arise. The kurtosis test will have good power only when
both location and scale differences (between the samples) are eliminated. This re-
quires basing it on the sample values aligned not only for location (by subtracting
a location estimate) but also for scale (by dividing the sample values by a scale
estimate). Such double-alignment, however, affects the levels of the kurtosis test.
Boos (1986, p. 1025) says: “... In general, one should be cautious about inter-
preting the \( p \)-value of the Kurtosis test after aligning for both location and scale
...”.

Remark 2. Even when \( F_1 \) and \( F_2 \) have the same shape and equal locations, the
level of the Wilcoxon test can be seriously affected by unequal scales (cf. (Pratt,
1964)). Besides, unequal scales may also damage the effectiveness of the Wilcoxon
test in detecting location differences. Therefore, even in GLOBE*, WILCOXON
should be replaced by WILCOXON*, which is WILCOXON based on the scale-
aligned samples.

The question arises as to whether the performance of GLOBE or GLOBE* can
be improved by bootstrapping. The answer is “No”. For, as explained in the next
paragraph, due to the conflicting requirements of the test statistics, it is impossible
to find samples, bootstrapping which will achieve efficiency (of power) for all the
statistics.

The Wilcoxon test is effective only when based on the scale-aligned samples
(cf. Remark 2). However, this scale alignment will destroy the effectiveness of the
Mood test (in detecting scale differences). Next, the Mood test is effective only
when based on the location-aligned samples. Nevertheless, this location alignment
will destroy the effectiveness of the Wilcoxon test. Finally, the kurtosis test is
effective only when based on the samples aligned for both location and scale.
However, this double alignment will destroy the effectiveness of both the Wilcoxon
and the Mood tests. On the other hand, our \( V \)-statistic requires no location or scale-
alignment and applies to symmetric as well as skewed distributions (cf. Table 1).
Table 3.
Data on rainfall (in inches) in New York’s Central Park from 1950 to 1979

<table>
<thead>
<tr>
<th>Month</th>
<th>February</th>
<th>August</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.15</td>
<td>5.29</td>
</tr>
<tr>
<td></td>
<td>2.46</td>
<td>5.87</td>
</tr>
<tr>
<td></td>
<td>2.33</td>
<td>3.15</td>
</tr>
<tr>
<td></td>
<td>1.63</td>
<td>6.58</td>
</tr>
<tr>
<td></td>
<td>2.48</td>
<td>13.82</td>
</tr>
<tr>
<td></td>
<td>4.65</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>5.09</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>1.79</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>4.42</td>
<td>5.94</td>
</tr>
<tr>
<td></td>
<td>3.98</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td>2.85</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>3.44</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>2.91</td>
<td>6.26</td>
</tr>
<tr>
<td></td>
<td>6.58</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td>1.13</td>
<td>3.05</td>
</tr>
<tr>
<td></td>
<td>4.52</td>
<td>5.33</td>
</tr>
<tr>
<td></td>
<td>3.05</td>
<td>5.90</td>
</tr>
<tr>
<td></td>
<td>5.40</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>2.88</td>
<td>2.53</td>
</tr>
<tr>
<td></td>
<td>2.47</td>
<td>9.37</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
<td>3.08</td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td>3.13</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>1.59</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>5.99</td>
<td>3.05</td>
</tr>
<tr>
<td></td>
<td>6.52</td>
<td>4.57</td>
</tr>
<tr>
<td></td>
<td>4.50</td>
<td>5.50</td>
</tr>
<tr>
<td></td>
<td>4.27</td>
<td></td>
</tr>
</tbody>
</table>

4. AN ILLUSTRATION

Table 3 contains the rainfall data at New York’s Central Park for the months of February and August, using data from 1950–1979 (cf. (Barnett and Eisen, 1982)). The estimated 95% quantile of the V-statistic was 0.367. The statistics for location, scale, symmetry (or skewness) and tailweight were respectively 0.43, 0.52, 0.22 and 0.136. Thus only the location and scale differences were significant. This result is consistent with the findings of Barnett and Eisen (1982) and Boos (1986).

5. ASYMPTOTICS

Let $X_1 < \cdots < X_n$ be order statistics of an i.i.d. sample $X_1, \ldots, X_n$ from a continuous distribution $F$. Let $X_1^* \leq \cdots \leq X_n^*$ be the order statistics of a smooth bootstrap sample $X_1^*, \ldots, X_n^*$ from $X_1, \ldots, X_n$. Let $U_\gamma$ and $L_\gamma$ denote the averages of the top $n\gamma$ and bottom $n\gamma$ observations in the sample and let

$$Q_2 = (U_{0.05} - L_{0.05})/(U_{0.5} - L_{0.5}).$$

For $0 < u \leq 1$, let

$$F^{-1}(u) = \inf\{x: F(x) \geq u\} \quad \text{and} \quad F^{-1}(0) = \lim_{\varepsilon \downarrow 0} F^{-1}(\varepsilon).$$

**Theorem 3.** Let $0 \leq \alpha < \beta \leq 1$. Suppose $F^{-1}$ is continuous at $\alpha$, if $\alpha > 0$ and continuous at $\beta$ if $\beta < 1$. Then

$$\frac{1}{n} \sum_{\alpha n < i \leq \beta n} X_{(i)} = \int_{\alpha}^{\beta} F^{-1}(u) \, du + \frac{1}{n} \sum_{i=1}^{n} Z(X_i, \alpha, \beta) + o_p(n^{-1/2}),$$

(a)
where

\[ Z(X, \alpha, \beta) = \int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} (F(x) - I(X \leq x)) \, dx. \]

We also have, for almost all samples,

\[ \frac{1}{n} \sum_{\alpha < i \leq \beta} (X_{i}^{c} - X_{i}) = \frac{1}{n} \sum_{i=1}^{n} (Z(X_{i}, \alpha, \beta) - Z(X_{i}, \alpha, \beta)) + o_p(n^{-1/2}). \]

**Proof.** Part (a) follows from (P 9) of Babu and Singh (1984b). Part (b) follows from a proof similar to Theorem 3 of Babu and Singh (1984b). Even though Theorem 3 there, was proved for the standard bootstrap, a minor modification yields part (b).

**Remark 4.** If a Lipschitz condition of order \( \eta \in (0, 1/2) \) is assumed for \( F^{-1} \) in small neighborhoods of \( \alpha \) and \( \beta \) (i.e., for some \( C > 0, \varepsilon > 0, \)

\[ |F^{-1}(\mu) - F^{-1}(\alpha)| \leq C|\mu - \alpha|^\eta \quad \text{and} \quad |F^{-1}(\nu) - F^{-1}(\beta)| \leq C|\nu - \beta|^\eta \]

for \( \mu \in (\alpha - \varepsilon, \alpha + \varepsilon) \) and \( \nu \in (\beta - \varepsilon, \beta + \varepsilon), \) \( 0 < \alpha - \varepsilon < \beta + \varepsilon < 1) \) then the error terms can be improved.

**Remark 5.** Note that if \( F^{-1} \) is continuous at 0.05, 0.5 and 0.95, then

\[ U_{0.05} = \frac{1}{0.05} \left( a_{1} + \frac{1}{n} \sum_{i=1}^{n} Z(X_{i}, 0.95, 1) \right) + o_p(n^{-1/2}), \]

\[ L_{0.05} = \frac{1}{0.05} \left( a_{2} + \frac{1}{n} \sum_{i=1}^{n} Z(X_{i}, 0, 0.05) \right) + o_p(n^{-1/2}), \]

\[ U_{0.5} = 2 \left( a_{3} + \frac{1}{n} \sum_{i=1}^{n} Z(X_{i}, 0.5, 1) \right) + o_p(n^{-1/2}), \]

\[ L_{0.5} = 2 \left( a_{4} + \frac{1}{n} \sum_{i=1}^{n} Z(X_{i}, 0, 0.5) \right) + o_p(n^{-1/2}), \]

and

\[ a_{1} = \int_{0.95}^{1} F^{-1}(u) \, du, \quad a_{2} = \int_{0}^{0.05} F^{-1}(u) \, du, \]
\[ a_3 = \int_{0.5}^{1} F^{-1}(u) \, du, \quad \text{and} \quad a_4 = \int_{0}^{0.5} F^{-1}(u) \, du. \]

Consequently,

\[ Q_2 = \frac{a_1 - a_2 + \frac{1}{n} \sum_{i=1}^{n} [Z(X_i, 0.95, 1) - Z(X_i, 0, 0.05)]}{a_3 - a_4 + \frac{1}{n} \sum_{i=1}^{n} [Z(X_i, 0.5, 1) - Z(X_i, 0, 0.5)]} + o_p(n^{-1/2}). \]

A similar proof using Theorem 1 of Babu and Singh (1984b) leads to similar representations for trimmed means and trimmed variances.

Remark 6. Let \( F_n \) denote the empirical distribution of \( X_i - \bar{X} \) and \( F_n^* \) denote the empirical distribution of the bootstrap sample \( X_i^* - \bar{X}_n^* \). Note that if \( F^{-1} \) is continuous at 0.5, then \( \bar{X}_n \) converges to the population median \( m \).

Let \( Y_n,F(x) = F_n(x) + F_n(-x) - 1 - (F(x + m) + F(-x + m) - 1) \)

and

\[ Y_n,F^*(x) = F_n^*(x) + F_n^*(-x) - 1 - (F_n(x + \bar{X}) + F_n(-x + \bar{X}) - 1). \]

By standard results on weak convergence of empirical processes (see (Shorack and Wellner, 1986, Chapter 3)), and by Theorem 1 and its corollary on p. 764 of (Shorack and Wellner, 1986) for bootstrapped version, it follows that the processes \( \sqrt{n} Y_n,F \) and (for almost all sample sequences) \( \sqrt{n} Y_n,F^* \) converge weakly to the same Gaussian process \( Y_F \). The covariance function of \( Y_F \) is given by

\[ \text{Cov}(Y_F(x), Y_F(y)) = 2F(x + m) + (F(y + m) + F(-y + m))(1 - F(x + m) - F(-x + m)), \]

for all \( x \leq y < 0 \). If \( F \) is continuous and symmetric around its median, it follows that for almost all sample sequences \( X_1, \ldots, X_n \), both

\[ S_n = \sqrt{n} \sup_{x \leq 0} |F_n(x) + F_n(-x) - 1| \quad \text{and} \quad S_n^* = \sqrt{n} \sup_{x \leq 0} |Y_n,F^*(x)| \]

have the same limiting distribution as that of \( \sup_{0 \leq t \leq 1} |Y(t)| \), where \( Y \) is the standard Brownian motion on \( 0 \leq t \leq 1 \).

Suppose two samples of sizes \( n_1 \) and \( n_2 \) are drawn from the continuous populations \( F_1 \) and \( F_2 \) respectively. Let \( F_1^{-1} \) and \( F_2^{-1} \) be continuous at 0.05, 0.5, 0.95. For some \( \delta > 0 \), let \( \delta < n_1/n_2 < 1 - \delta \). By the theorem, Remarks 2 and 3, it follows that \( (V_1, \ldots, V_4) \) and its bootstrap version \( (V_1^*, \ldots, V_4^*) \) both have
the same asymptotic distribution, for almost all samples. Hence by Lemma 2.1 of Babu and Bose (1988) for any smooth real valued function $H$ on four-dimensional Euclidean space,

$$P(H(V_1, \ldots, V_4) \leq t^*_{H,\gamma}) - \gamma \to 0,$$

where $t^*_{H,\gamma}$ is the $\gamma$th quantile of the distribution of $H(V^*_1, \ldots, V^*_4)$.

Acknowledgement

The authors are grateful to Professor Boos for some useful correspondence on this problem.

REFERENCES