ON BERRY-ESSEEN RATES FOR STATISTICAL FUNCTIONS, WITH APPLICATION TO L-ESTIMATES
by Dennis D. Boos and R. J. Serfling

FSU Department of Statistics Report M499
FSU Technical Report 136

Research supported by the Army, Navy, and Air Force under Office of Naval Research Contract No. N00014-76-C-0608
Reproduction in whole or in part is permitted for any purpose of the United States Government.

Department of Statistics, North Carolina State University

400 2 77
A parameter expressed as a functional $T(F)$ of a distribution function (d.f.) $F$ may be estimated by the "statistical function" $T(F_n)$ based on the sample d.f. $F_n$. Typically, $T(F_n)$ is asymptotically normal. We investigate the rate of this convergence by utilizing the von Mises (1947) representation to express $T(F_n) - T(F)$ as an approximate U-statistic plus $R_n$, and applying the Berry-Esséen rate $O(n^{-1/2})$ established for U-statistics by Callaert and Janssen (1978). This essentially reduces the problem to a handling of $R_n$. We carry out this method for linear functions of order statistics ("L-estimates") and obtain results competitive with Bjerve (1977) and Helmers (1977). Also, we briefly indicate the application of the method to M-estimates.
0. Summary. Let $T(\cdot)$ be a real-valued functional defined on distribution functions (d.f.'s). For a sample $X_1, \ldots, X_n$ from a d.f. $F$, consider estimation of the "parameter" $T(F)$ by the sample analogue estimator $T(F_n)$ based on the usual sample d.f. $F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i < x)$, $-\infty < x < \infty$. In many cases, for a suitable constant $\sigma^2(T,F)$, the d.f. of $n^{1/2}(T(F_n) - T(F))/\sigma(T,F)$ converges weakly to the standard normal d.f. $\phi$. In Section 1 we formulate a general approach toward investigation of the rate of this convergence. Taking two terms of the von Mises (1947) Taylor expansion for statistical functions, we represent $T(F_n) - T(F)$ as an approximate U-statistic plus a remainder term $R_n$. Under appropriate conditions, we may dispense with $R_n$, switch to an exact U-statistic, and exploit the Berry-Esseen rate $O(n^{-1/2})$ established for U-statistics by Callaert and Janssen (1978). (On the basis of only one term from von Mises' expansion, the method typically yields a somewhat weaker rate.) In Section 2 the method is applied to linear functions of order statistics ("L-estimates"), as represented by functionals of the form $T(F) = \int_0^1 F^{-1}(u)J(u)\,du$. The rate $O(n^{-1/2})$ is obtained under conditions on $J$ and $F$ competitive with Bjerve (1977) and Helmers (1977). In Section 3 the application of the approach to M-estimates is sketched.

1. A general approach. Utilizing the notion of Taylor expansion for statistical functions introduced by von Mises (1947), and taking two terms, we represent $T(F_n) - T(F)$ as $\nu_n^* + \text{remainder}$, where

$$\nu_n^* = \frac{d}{d\lambda} T((1-\lambda)F+\lambda F_n)\big|_{\lambda=0^+} + \lambda \frac{d^2}{d\lambda^2} T((1-\lambda)F+\lambda F_n)\big|_{\lambda=0^+}.$$
Typically, the terms in $V^*_n$ may be represented in the forms $\int a(x)\,dF_n(x)$ and $\int b(x,y)\,dF_n(x)\,dF_n(y)$, respectively. In this case $V^*_n$ has the form (1.1) below, with $h(x,y) = \frac{1}{2}[a(x) + a(y) + b(x,y)]$. The following theorem, stated for this setting of a statistical function $T(F_n)$ estimating $T(F)$, is in fact valid in connection with any statistic $T_n = T_n(X_1, \ldots, X_n)$ estimating a parameter $T_0$, but depends upon finding a suitable function $h(x,y)$.

**THEOREM 1.1.** Suppose that $T(F_n) - T(F)$ may be represented as

$$V_n + R_n,$$

where $V_n$ is given by

(1.1) \[ V_n = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j), \]

with $h(x,y)$ symmetric in its arguments and satisfying

(1.2a) \[ E_F[h(X_1, X_2)] = 0, \]

(1.2b) \[ E_F|h(X_1, X_2)|^3 < \infty, \]

(1.2c) \[ E_F|h(X_1, X_1)|^{3/2} < \infty, \]

and where, for some $A > 0$,

(1.3) \[ P(|R_n| > An^{-1}) = O(n^{-b}), \ n \to \infty. \]

Put $\sigma^2(T, F) = 4\text{Var}_F(g(X_1))$, where $g(x) = E_F(h(x, X_1))$. Then

(1.4) \[ \sup_{t} |P(n^{b/2}[T(F_n) - T(F)]/\sigma(T,F) \leq t) - \Phi(t)| = O(n^{-b}), \ n \to \infty. \]

We obtain Theorem 1.1 by successive application of the following well-known and easily proved device.
**Lemma 1.1** Let the sequence of rv's \( \{ \xi_n \} \) satisfy

\[ \sup_t |P(\xi_n < t) - \Phi(t)| = O(n^{-1/2}) . \tag{1} \]

Then, for any sequence of rv's \( \{ \Delta_n \} \) and constant \( B \),

\[ \sup_t |P(\xi_n + \Delta_n < t) - \Phi(t)| = O(n^{-1/2}) + P(|\Delta_n| > Bn^{-1/2}) . \tag{2} \]

For the first application, put \( \xi_n = n^{1/2}v_n / \sigma(T,F) \) and \( \Delta_n = n^{1/2}r_n / \sigma(T,F) \). Then, by (1.3), it suffices for (1.4) to establish (1). Now, for the

*U-statistic* of Hoeffding (1948) associated with the kernel \( h \), namely

\[ U_n = \frac{1}{n} \sum_{1 \leq i < j \leq n} h(X_i, X_j), \]

Corllet and Janssen (1978) establish (1) for \( \xi_n = n^{1/2}u_n / \sigma(T,F) \), under the conditions (1.2a,b). Therefore,

To complete the proof of Theorem 1.1, it suffices to show that

\[ P(n^{1/2}|U_n - V_n| > Bn^{-1/2}) = O(n^{-1/2}) \]

for some constant \( B \). This is established in the following result.

**Lemma 1.2.** Suppose that \( h(x,y) \) is symmetric in its arguments

and satisfies \( Eh^2(X_1, X_2) = 1 \) and \( Eh(X_1, X_1)|^{3/2} = 1 \). Then, for \( B > 2|E(h(X_1, X_2) - h(X_1, X_1))| \),

\[ P(|U_n - V_n| > Bn^{-1}) = o(n^{-1/2}) , n \to \infty . \]

**Proof.** Check that \( U_n - V_n = n^{-1}(U_n - W_n) \), where \( W_n = n^{-1} \sum_{i=1}^{n} h(X_i, X_i) \). Then, for \( B > 2|E(h(X_1, X_2) - h(X_1, X_1))| \), we have

\[ P(|U_n - V_n| > Bn^{-1}) = P(|U_n - W_n| > B) \]

\[ \leq P(|U_n - W_n - E(U_n) + E(W_n)| > B) \]

\[ \leq P(|U_n - E(U_n)| > B/4) + P(|W_n - E(W_n)| > B/4) . \]
The first term on the right is $O(n^{-1})$ by Chebyshev's inequality and the well-known rate $\text{Var}(U_n) = O(n^{-1})$. For the second term, we use Theorem 4 of Baum and Katz (1965), which implies: for $\{Y_i\}$ i.i.d. with $E(Y_1) = 0$ and $E|Y_1|^r < \infty$, where $r \geq 1$, $P(|Y| > \varepsilon) = o(n^{1-r})$, for all $\varepsilon > 0$.

We thus apply this result with $r = 3/2$. 

**Remark.** Note (using (**) that even if the optimum rate $O(n^{-\frac{3}{2}})$ is not known for $R_n$ in (1.3), we still may obtain a (weaker) Berry-Esséen rate by replacing $O(n^{-\frac{3}{2}})$ in (1.4) by $O(P(|R_n| > An^{-1})).$

2. *L*-estimates. Consider the functional $T(F) = \int_0^1 F^{-1}(u)J(u)du$ and the corresponding $L$-estimate $T(F_n)$. Implementing the approach of Section 1, we find (under regularity conditions on $J$)

$$V_n^* = \int_{-\infty}^{\infty} [F_n(x) - F(x)]J\circ F(x)dx - \frac{1}{2} \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 J'\circ F(x)dx$$

and thus $V_n^* = V_n$ of form (1.1) with $h(x, y) = \frac{1}{2} [\alpha(x) + \alpha(y) + \beta(x, y)]$, where

(2.1a) $\alpha(x) = - \int_{-\infty}^{\infty} [I(x < t) - F(t)]J\circ F(t)dt$

and

(2.1b) $\beta(x, y) = - \int_{-\infty}^{\infty} [I(x < t) - F(t)][I(y < t) - F(t)]J'\circ F(t)dt$ .

Writing $K(u) = \int_0^u J(u)du$, we obtain by integration by parts that

$T(G) - T(F) = - \int_{-\infty}^{\infty} [K\circ G(x) - K\circ F(x)]dx$. Therefore, for $R_n = T(F_n) - T(F) - V_n$ we obtain
(2.2) \[ R_n = - \int_{-\infty}^{\infty} W_{G,F}(x)dx , \]

where

\[ W_{G,F} = K \circ G - K \circ F - (J \circ F)(G-F) - \frac{1}{2}(J \circ (G-F))^2 . \]

Having found \( h \) and \( R_n \) heuristically, we now rigorously establish two Berry-Esséen theorems for L-estimates by introducing sets of conditions on \( J \) and \( F \) under which \( h \) and \( R_n \) are well-defined and satisfy the requirements of Theorem 1.1. The relevant asymptotic variance parameter is

\[ \sigma^2(J,F) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J(x)J(y)[F(\min(x,y)) - F(x)F(y)]dx dy . \]

**THEOREM 2.1.** Suppose that \( J \) vanishes outside an interval \([a,b]\), \( 0 < a < b < 1 \), and that \( J' \) exists and satisfies a Lipschitz condition of order \( \delta > 0 \) on an open interval containing \([a,b]\). Assume \( \sigma^2(J,F) > 0 \). Then

\[ \sup_t |P(n^{1/2}[T_n - T]} \sigma(J,F) < t) - \theta(t)| = O(n^{-1/2}), \quad n \to \infty . \]

**THEOREM 2.2.** Suppose that \( J' \) exists and satisfies a Lipschitz condition of order \( \delta > 1/3 \) on \((0,1)\). Assume that \( E_F[X_1]^3 < \infty \) and that \( 0 < \sigma^2(J,F) < \infty \). Then (2.3) holds.

**REMARKS.** (1) Theorem 2.1 requires that \( J \) be trimmed away from 0 and 1, but imposes no conditions on \( F \) other than \( \sigma^2(J,F) > 0 \).

Bjerve (1977), also for the case of \( J \) trimmed, obtains (2.3) requiring only that \( J \) be Lipschitz of order 1 on an open interval containing \([a,b]\), but furthermore requiring that \( F^{-1} \) possess a second derivative also satisfying such a Lipschitz condition.
(ii) Theorem 2.2 allows $J$ to have support $(0,1)$ but imposes a standard moment condition on $F$. Helmers (1977) proves a similar theorem. He allows $J'$ not to exist at finitely many points but requires $J'$ to be Lipschitz of order $> \frac{1}{2}$ on the open intervals where it exists and $F^{-1}$ to be Lipschitz of order $> \frac{1}{2}$ on neighborhoods of the points where $J'$ does not exist. (The additional restriction $\int |J'|dF^{-1} < \infty$ required in Helmers (1977a) is eliminated in Helmers (1977b).)

(iii) Theorems 2.1 and 2.2 remain valid with $T(F_n)$ replaced in (2.3) by the closely related statistic $T_n = n^{-1}\sum_{i=1}^n J(i/(n+1))X_{ni}$, where $X_{n1} \leq \ldots \leq X_{nn}$ denote the ordered sample values. To see this, write $T(F_n) = \sum_{i=1}^n [K(i/n) - K((i-1)/n)]X_{ni}$ and apply standard arguments. For example, in the case of Theorem 2.2, we use boundedness of $J'$ to write $|T(F_n) - T_n| \leq n^{-2}\sum_{i=1}^n |X_i|$ for a constant $M$. Put $\Delta_n = n^k[T(F_n) - T_n]$ and obtain $P(|\Delta_n| > 2ME|X_1|n^{-k}) \leq P(n^{-1}\sum_{i=1}^n |X_i| > 2E|X_1|) \leq P(n^{-1}\sum_{i=1}^n |X_i| - E|X_1| > E|X_1|) = O(n^{-1})$ by Chebyshev's inequality, since $E|X_1|^2 < \infty$. Then apply Lemma 1.1. □

PROOF OF THEOREM 2.1. First we utilize the assumption that $J$ is trimmed. Let $0 < \epsilon < \min(a,l-b)$. Then there exist $A,B$ such that $-\infty < A < F^{-1}(a-\epsilon) < F^{-1}(b+\epsilon) < B < \infty$. Hence, defining $||q||_\infty = \sup_x |q(x)|$, we have that $||G - F||_\infty < \epsilon$ implies $W_{G,F}(x) = 0$ for $x \notin [A,B]$. Therefore, defining

$$R_n(A,B) = -\int_A^B W_{F_n,F}(x)dx,$$

we have

$$P(|R_n| > Cn^{-1}) \leq P(|R_n(A,B)| > Cn^{-1}) + P(||F_n - F||_\infty \geq \epsilon).$$
Now apply the Lipschitz condition on \( J' \), whereby

\[
|J'(u_1) - J'(u_2)| \leq D|u_1 - u_2|^{\delta},
\]

to obtain

\[
|R_n(A,B)| \leq \frac{1}{2}(B-A)D \left\| F_n - F \right\|_{\infty}^{2+\delta},
\]

where \( \delta > 0 \). Thus, with \( C_1 = 2C/(B-A)D \),

\[
(P(|R_n| > C_n^{-1}) \leq P(n\left\| F_n - F \right\|_{\infty}^{2+\delta} > C_1) + P(\left\| F_n - F \right\|_{\infty} \geq \varepsilon).
\]

We now apply the inequality \( P(n^{\frac{k}{2}}\left\| F_n - F \right\|_{\infty} > d) \leq D_0 \exp(-2d^2) \), \( d > 0 \),

where \( D_0 \) is a constant, due to Dvoretzky, Kiefer and Wolfowitz (1956).

It is readily seen that the terms on the right in (2.4) are each \( O(n^{-\frac{k}{2}}) \),

so that \( R_n \) satisfies (1.3).

The required properties of \( h \) are obtained easily. Restricting

the range of integration in (2.1) to the interval \([A,B]\], finiteness of

moments follows. Further, interchanging expectation and integration by

Fubini's theorem, we find \( E(\alpha(X_1)) = 0, E(\beta(x,X_1)) = 0 \) and thus also

\( E(\beta(X_1,X_2)) = 0 \) and \( g(x) = E(h(x,X_1)) = \frac{1}{2}\alpha(x) \). Thus \( \sigma^2(T,F) \) of

Theorem 1.1 is given by \( \sigma^2(X_1) = \sigma^2(J,F) \). \( \square \)

In proving Theorem 2.2, we will need a property of \( \left\| F_n - F \right\|_{L^2} \),

where \( \left\| \cdot \right\|_{L^2} \) denotes the \( L^2 \)-norm, \( \left\| q \right\|_{L^2} = [\int q^2(x)dx]^{\frac{1}{2}} \).

**Lemma 2.1.** Let \( k \) be a positive integer and assume \( E_F |X_1|^k < \infty \).

Then

(2.5) \[
E(\left\| F_n - F \right\|_{L^2}^{2k}) = O(n^{-k}), \quad n \to \infty.
\]
PROOF. Put $Y_i(t) = I(X_i \leq t) - F(t)$, $1 \leq i \leq n$. Then

$$E\|F_n - F\|_{L_2}^{2k} = n^{-2k} \sum_{1=1}^{n} \sum_{j_1=1}^{n} \sum_{i_1=1}^{n} \cdots \sum_{j_k=1}^{n} \sum_{i_k=1}^{n} \sum_{j_k=1}^{n} E\{ Y_i(t) Y_j(t) \} dt_1 \cdots dt_k.$$  

Since $E|X_1| < \infty$, integration by parts yields the inequality

$$\int |I(X_1 \leq t) - F(t)| dt \leq |X_1| + E|X_1|.$$  

Consequently, the expectation term on the right in (2.6) is finite since

$$E|X_1|^k < \infty$$  

and, by Fubini's theorem,

$$E\{ \prod_{i=1}^{k} Y_{i_1}(t) Y_{i_2}(t) \} = \cdots \int E\{ \prod_{i=1}^{k} Y_{i_1}(t) Y_{i_2}(t) \} dt_1 \cdots dt_k.$$  

By independence of $Y_1(s_1), \ldots, Y_n(s_n)$ for any $s_1, \ldots, s_n$, we have

$$E\{ Y_{i_1}(t_1) Y_{j_1}(t_1) \cdots Y_{i_k}(t_k) Y_{j_k}(t_k) \} = 0$$  

except possibly in the case that each index in the list $i_1, j_1, \ldots, i_k, j_k$ appears at least twice. In this case the number of distinct elements in the set $\{i_1, j_1, \ldots, i_k, j_k\}$ is $\leq k$. It follows that the number of ways to choose $i_1, j_1, \ldots, i_k, j_k$ such that the expectation in (2.8) is nonzero is $O(n^k)$. Thus the number of nonzero terms in the summation in (2.6) is $O(n^k)$.

PROOF OF THEOREM 2.2. Applying the Lip condition on $J^-$, we obtain

$$|F_n| \leq kD \int |F_n(x) - F(x)|^{2+\delta} dx \leq kD |F_n - F|_\infty^\delta \cdot |F_n - F|_{L_2}^2.$$
Thus, for any $A > 0$,

$$P(\left| R_n \right| > An^{-1}) \leq P(n \left| \Delta F_n - F \right|_L^2 > 2A/D) \leq P(n^{1/6} \left| \Delta F_n - F \right|_L > 2A/D) + P(n^{5/6} \left| \Delta F_n - F \right|_L^2 > 1).$$

For $\delta > 1/3$, the first term above is $O(n^{-3/2})$ by the DKW inequality used in the proof of Theorem 2.1. The second term above is $O(n^{-3/2})$ by Lemma 2.1, applied for $k = 3$. Therefore, $R_n$ satisfies (1.3).

The required properties of $h$ are established in similar fashion as in the proof of Theorem 2.1. In this connection, the inequality (2.7) is useful in proving finiteness of moments.

REMARK. It is evident from the preceding proof that under a higher moment assumption $E|X_1|^\nu < \infty$, where $\nu$ is an integer $> 3$, the Lip condition on $J^*$ may be relaxed to order $\delta > 1/\nu$.

3. M-estimates. Let $\psi$ be a function such that the parameter of interest $T(F)$ may be defined as a solution of the equation $\lambda_F(t) = 0$, where $\lambda_F(t) = \int \psi(x-t)dF(x)$. Thus $T(\cdot)$ represents an "M-functional with respect to $\psi"$ if $\lambda_F(T(F)) = 0$, all $F$. In the following let us put $T_n$ for $T(F_n)$ and $t_\circ$ for $T(F)$.

The approach of Section 1 leads to

$$\alpha(x) = -\frac{\psi(x-t_\circ)}{\lambda_F'(t_\circ)},$$

and

$$\beta(x,y) = \frac{2\alpha(x)}{\lambda_F'(t_\circ)} \left[ \psi'(y-t_\circ) + \lambda_F'(t_\circ) - \frac{1}{2} \lambda_F''(t_\circ) \alpha(y) \right].$$
Defining $h(t) = [\lambda_F(t) - \lambda_F(t_0)]/(t-t_0)$ or $\lambda_F''(t_0)$, according as $t \neq t_0$ or $t = t_0$, we obtain for $R_n = T_n - t_0 - V_n$

$$R_n = -\frac{\int \psi(x-T_n) - \psi(x-t_0) d[F_n(x)-F(x)]}{h(T_n)} + \left[\frac{1}{\lambda_F''(t_0)} - \frac{1}{h(T_n)}\right] \int \psi(x-t_0) d[F_n(x)-F(x)]$$

$$+ \frac{\int \psi(x-t_0) d[F_n(x)-F(x)] \int \psi'(x-t_0) d[F_n(x)-F(x)]}{[\lambda_F''(t_0)]^2}$$

$$+ \frac{\lambda_F''''(t_0) \int \psi(x-t_0) d[F_n(x)-F(x)]^2}{2[\lambda_F''(t_0)]^3}.$$

A brute force treatment of $R_n$ leads to

$$|R_n| \leq C_1 \|F_n - F\|_\infty^3$$

under moderate restrictions on $\psi, \psi', \psi''$, and on $\lambda_F''(t), \lambda_F''''(t)$ and $\lambda_F''''(t)$ for $t$ in a neighborhood of $t_0$. An application of the DKW inequality cited in Section 2 then yields (1.3) for $R_n$. Condition (1.2) is easily verified. For other discussion of the Berry-Esseen rate for $M$-estimates, see Bickel (1974).
REFERENCES


A parameter expressed as a functional $T(F)$ of a distribution function (d.f.) $F$ may be estimated by the "statistical function" $T(F_n)$ based on the sample d.f. $F_n$. Typically, $T(F_n)$ is asymptotically normal. We investigate the rate of this convergence by utilizing the von Mises (1947) representation to express $T(F_n) - T(F)$ as an approximate U-statistic plus $R_n$, and applying the Berry-Essén rate $O(n^{-1})$ established for U-statistics by Callaert and Janssen (1978). This essentially reduces the problem to a handling of $R_n$. We carry out this method for linear functions of order statistics ("L-estimates") and obtain results competitive with Bjerve (1977) and Helmers (1977). Also, we briefly indicate the application of the method to M-estimates.