The Annals of Statistics
1980, Vol. 8, No. 3, 618–624

A NOTE ON DIFFERENTIALS AND THE CLT AND LIL FOR STATISTICAL FUNCTIONS, WITH APPLICATION TO M-ESTIMATES[1]

BY DENNIS D. BOOS AND R. J. SERFLING

North Carolina State University and Florida State University

A parameter expressed as a functional $T(F)$ of a distribution function (df) F may be estimated by the "statistical function" $T(F_n)$ based on the sample df F_n. For analysis of the estimation error $T(F_n) - T(F)$, we adapt the differential approach of von Mises (1947) to exploit stochastic properties of the Kolmogorov-Smirnov distance $\sup_x|F_n(x) - F(x)|$. This leads directly to the central limit theorem (CLT) and law of the iterated logarithm (LIL) for $T(F_n) - T(F)$. The adaptation also incorporates innovations designed to broaden the scope of statistical application of the concept of differential. Application to a wide class of robust-type M-estimates is carried out.

0. Introduction. Parameters of interest in statistics can often be expressed as functionals $T(F)$ of the underlying population distribution function (df), in which case a natural sample analogue estimator is provided by the "statistical function" $T(F_n)$ based on the sample df F_n. The functional representation of statistical parameters was first studied in detail by von Mises (1947), who developed a theory of differentiation of statistical functions $T(F_n)$ and employed related Taylor expansions as a tool. This work was extended in the framework of stochastic process theory by Filippova (1962). In the present paper, attention is focused upon the *differential* of a functional T as the key concept and tool. This approach bypasses a higher-order remainder term in the Taylor expansion but introduces the difficulty of handling a norm. However, for the *sup-norm* $\|\cdot\|_\infty$, this enables us to take advantage of known stochastic properties of the Kolmogorov-Smirnov distance $\|F_n - F\|_\infty$. The approach then yields in a very direct way the central limit theorem (CLT) and the law of the iterated logarithm (LIL) for $T(F_n) - T(F)$. This development is presented in Section 1. Although essentially in the spirit of Frechet differentiation, the treatment incorporates certain modifications giving the method greater scope and flexibility in statistical applications.

Section 2 provides application to M-estimates, obtaining the CLT, the LIL and a differential for "M-functionals" $T(\cdot)$ of the type of interest in *robust* estimation. A similar application to L-estimates is provided in Boos (1979).

Section 3 provides brief complements. Also see Boos (1977).

Received November 1977; revised March 1979.

[1]Research supported by the Army, Navy, and Air Force under Office of Naval Research Contract No. N00014-76-C-0608.

AMS 1970 *subject classifications.* Primary 62E20; secondary 62G35.

Key words and phrases. Differentials, functionals, statistical functions, asymptotic normality, law of the iterated logarithm, M-estimates.

1. The differential approach for statistical functions. Consider a functional $T(\cdot)$ defined on df's G, for example, the variance functional $T(G) = \int[x - \int x dG(x)]^2 dG(x)$. Corresponding to a sample $X_1, \cdots, X_n$ from a df F, let F_n denote the usual sample df $n^{-1}\Sigma_1^n \delta_{X_i}$, where δ_x denotes the df with point mass 1 at x. The "statistical function" $T(F_n)$ is the natural sample analogue estimator of the "parameter" $T(F)$.

Let $\mathcal{F}$ be the set of df's and $\mathfrak{D}$ the linear space generated by differences $G - H$ of members of $\mathcal{F}$. Let $\mathfrak{D}$ be equipped with a norm $\|\cdot\|$. The functional T defined on $\mathcal{F}$ is said to have a *differential* at the point $F \in \mathcal{F}$ with respect to the norm $\|\cdot\|$ if there exists a functional $T(F; \Delta)$, defined on $\Delta \in \mathfrak{D}$ and linear in the argument Δ, such that

$$(D) \qquad T(G) - T(F) - T(F; G - F) = o(\|G - F\|)$$

as $\|G - F\| \to 0$ ($T(F; \Delta)$ is called the "differential").

If T is differentiable with respect to the *sup-norm*, $\|h\|_\infty = \sup_x|h(x)|$, then direct application of (D) in conjunction with known stochastic properties of the Kolmogorov-Smirnov distance $\|F_n - F\|_\infty$ yields the CLT and the LIL for the estimation error $T(F_n) - T(F)$. For, utilizing the fact that $n^{\frac{1}{2}}\|F_n - F\|_\infty = O_p(1)$, which follows from a well-known inequality of Dvoretzky, Kiefer and Wolfowitz (1956), we immediately obtain from (D) that

$$n^{\frac{1}{2}}\left[T(F_n) - T(F) - T(F; F_n - F)\right] = o_p(1),$$

so that the limit distribution of $n^{\frac{1}{2}}[T(F_n) - T(F)]$ is given by that of $n^{\frac{1}{2}}T(F; F_n - F)$. But, by the linearity of the differential, $T(F; F_n - F)$ is an average of i.i.d. random variables,

$$T(F; F_n - F) = n^{-1}\Sigma_{i=1}^n T(F; \delta_{X_i} - F).$$

Put $\mu(T, F) = E\{T(F; \delta_X - F)\}$ and $\sigma^2(T, F) = \text{Var}\{T(F; \delta_X - F)\}$. In the case $\mu(T, F) = 0 < \sigma^2(T, F) < \infty$, the classical CLT yields

$$(\text{CLT}) \qquad n^{\frac{1}{2}}\left[T(F_n) - T(F)\right] \to_d N(0, \sigma^2(T, F)).$$

Further, by a similar argument using the fact that $n^{\frac{1}{2}}\|F_n - F\|_\infty = O(\log\log n)^{\frac{1}{2}})$ with probability 1(w.p. 1), which follows from the LIL for $\|F_n - F\|_\infty$ due to Chung (1949) for continuous F and extended by Richter (1974) to arbitrary F, the classical LIL of Hartman and Wintner (1941) yields

$$(\text{LIL}) \qquad \limsup_{n\to\infty} \frac{n^{\frac{1}{2}}\left[T(F_n) - T(F)\right]}{(2\sigma^2(T, F)\log\log n)^{\frac{1}{2}}} = 1 \text{ w.p. } 1.$$

Differentiability of T has additional applications. In the problem of "efficient" estimation, Kallianpur and Rao (1955) utilize Frechet differentiation w.r.t. $\|\cdot\|_\infty$ to characterize lower bounds to the asymptotic variance of $T(F_n)$. In the problem of "robust" estimation, Hampel (1968, 1974) exploits the approximation of the estimation error by $T(F; F_n - F)$ as a basis for suggesting $T(F; \delta_{X_i} - F)$ as a measure of the "influence" of the observation X_i toward this error. Thus $\Omega_{T,F}(x) = T(F; \delta_x - F)$, $-\infty < x < \infty$, is called the "influence curve" of the estimator $T(F_n)$ for $T(F)$. In a treatment of efficient *and* robust estimation, Beran (1977a, b) utilizes Hellinger-metric differentiation of functionals of densities.

The "candidate" differential $T(F; G - F)$ to be employed in verifying (D) is found by routine calculus methods:

$$(C) \qquad T(F; G - F) = \lim_{t \to 0+} \frac{T(F + t(G - F)) - T(F)}{t}.$$

The norm $\|\cdot\|$ is not involved in this computation.

The problem of actually verifying (D) may be approached in various ways. A device to be applied in Section 2 introduces an auxiliary functional $T_F(\cdot)$, chosen for convenience, such that

$$(D1) \qquad \lim_{\|G-F\| \to 0} T_F(G) = 1.$$

Then (D) may be obtained by establishing

$$(D2) \qquad T(G) - T(F) - T_F(G)T(F; G - F) = o(\|G - F\|)$$

and

$$(D3) \qquad [T_F(G) - 1]T(F; G - F) = o(\|G - F\|).$$

This device has a significant further application. If our purpose is merely to obtain the CLT and LIL for $T(F_n) - T(F)$, then it clearly suffices to reduce to the random variable $T_F(F_n) \cdot T(F; F_n - F)$ instead of to $T(F; F_n - F)$. Thus the previous argument leading to (CLT) and (LIL) carries through with (D) replaced by the *weaker* conditions $\{(D1), (D2)\}$. Moreover, it suffices merely to establish $\{(D1), (D2)\}$ *stochastically*, with G replaced by F_n, as $n \to \infty$. Also, to find $T(F; F_n - F)$ and the quantities $\mu(T, F)$ and $\sigma^2(T, F)$, we need only to compute $T(F; \delta_x - F)$ via (C). This general scheme is summarized as follows.

LEMMA 1.1. *Let $T(\cdot)$ be a functional on $\mathfrak{F}$ and F an element of $\mathfrak{F}$. For $T(F; \delta_x - F)$ defined by (C), suppose that $\mu(T, F) = 0 < \sigma^2(T, F) < \infty$. Let $T_F(\cdot)$ be an auxiliary functional. If $T_F(F_n) \to_p 1$ and $T(F_n) - T(F) - T_F(F_n)T(F; F_n - F) = o_p(\|F_n - F\|_\infty)$, then* (CLT) *holds for $T(F_n) - T(F)$. If these convergences hold* w.p. 1, *then* (LIL) *holds also.*

2. Application to M-estimation.

The setting for "M-estimation" of a parameter is based on a sample $X_1, \cdots, X_n$ from a df F and a function $\psi(x, t)$ such that the parameter of interest may be defined as the solution T of the equation

$$\int \psi(x, T)\, dF(x) = 0.$$

Often there are a variety of possible ψ for the same parameter. Here we establish properties of the functional T and the estimation error $T(F_n) - T(F)$ for a class of

ψ functions containing the typical examples arising in robust estimation. Specifically, we show that if ψ is continuous and bounded, then the functional T may be defined so as to be continuous at F with respect to $\|\cdot\|_\infty$ and thus to satisfy $T(F_n) \to T(F)$ w.p. 1. If ψ is merely continuous and satisfies (2.4) and $T(F_n)$ is consistent for $T(F)$, then the CLT and LIL are shown to hold for $T(F_n) - T(F)$ (ψ need not be bounded for consistency, e.g., Huber (1964) with ψ monotone). Finally under the further condition that ψ is of bounded variation, it is shown that T is differentiable w.r.t. $\|\cdot\|_\infty$.

Defining

$$\lambda_G(t) = \int \psi(x, t)\, dG(x), \qquad -\infty < t < \infty, G \in \mathfrak{F},$$

we call T an "M-functional w.r.t. ψ" if $\lambda_G(T(G)) = 0$, $G \in \mathfrak{F}$.

THEOREM 2.1. *Let $\psi(x, t)$ be continuous in each argument and bounded. Let F be given. Suppose that $\lambda_F(t_0) = 0$ and that at t_0 $\lambda_F(\cdot)$ changes sign uniquely in a neighborhood of t_0. Let $G_n \Rightarrow F$. Then, for every $\varepsilon > 0$, the interval $t_0 \pm \varepsilon$ contains a solution of $\lambda_{G_n}(t) = 0$ for all n sufficiently large.*

PROOF. Let $\varepsilon > 0$ be given. Without loss of generality, let $t_0 \pm \varepsilon$ lie within the hypothesized neighborhood of t_0. Now, since $\psi(x, t)$ is continuous in t and is bounded, by the dominated convergence theorem the functions $\lambda_{G_n}(\cdot)$, each n, and $\lambda_F(\cdot)$ are continuous. By continuity of $\lambda_F(\cdot)$ and the assumed uniqueness condition for the sign change at t_0, $\lambda_F(t_0 - \varepsilon)$ and $\lambda_F(t_0 + \varepsilon)$ must be opposite in sign. By the Helly-Bray theorem, since $\psi(x, t)$ is continuous in x and is bounded, $\lambda_{G_n}(t_0 - \varepsilon) \to \lambda_F(t_0 - \varepsilon)$ and $\lambda_{G_n}(t_0 + \varepsilon) \to \lambda_F(t_0 + \varepsilon)$, $n \to \infty$. Hence, for all n sufficiently large, $\lambda_{G_n}(t_0 - \varepsilon)$ and $\lambda_{G_n}(t_0 + \varepsilon)$ have opposite signs and thus, by continuity of each $\lambda_{G_n}(\cdot)$, there exists a solution of $\lambda_{G_n}(t) = 0$ in the interval $(t_0 - \varepsilon, t_0 + \varepsilon)$. □

The weak convergence $G_n \Rightarrow F$ is implied, in particular, by the condition $\|G_n - F\|_\infty \to 0$. Thus, under the conditions of the theorem, an M-functional T may be defined so as to be continuous at F w.r.t. $\|\cdot\|_\infty$. By the Glivenko-Cantelli theorem, $\|F_n - F\|_\infty \to 0$ w.p. 1. Thus, under the conditions of Theorem 2.1, with probability 1 every ε-neighborhood of $T(F)$ contains a solution $T(F_n)$ of $\lambda_{F_n}(t) = 0$ for all n sufficiently large. That is, the M-functional T admits a strongly consistent estimation sequence T_n for $T(F)$.

Let us now apply the differential approach of Section 1. By (C), we find

$$T(F; G - F) = -\frac{\int \psi(x, T(F))\, d[G(x) - F(x)]}{\lambda_F'(T(F))} = -\frac{\lambda_G(T(F))}{\lambda_F'(T(F))},$$

assuming $\lambda_F'(T(F)) \neq 0$. The device of $(D1) - (D3)$ is found to be productive with the auxiliary functional $T_F(G) = \lambda_F'(T(F))/h(T(G))$, where we define

$$h(t) = \frac{\lambda_F(t) - \lambda_F(T(F))}{t - T(F)}, \qquad t \neq T(F),$$
$$= \lambda_F'(T(F)), \qquad t = T(F).$$

Thus $(D1)$ is satisfied if $T(G) \to T(F)$ as $\|G - F\| \to 0$. And the left-hand side of $(D2)$ becomes simply $[\lambda_F(T(G)) - \lambda_F(T(F)) + \lambda_G(T(F))]/h(T(G))$. Since $\lambda_G(T(G)) = 0$, we may write

$$\lambda_F(T(G)) - \lambda_F(T(F)) + \lambda_G(T(F)) = \\ -\int[\psi(x, T(G)) - \psi(x, T(F))]\, d[G(x) - F(x)].$$

Thus $(D2)$ is equivalent to

$$\int[\psi(x, T(G)) - \psi(x, T(F))]\, d[G(x) - F(x)] = o(\|G - F\|). \tag{2.1}$$

Also, by $(D1)$, it suffices for $(D3)$ to show

$$\int\psi(x, T(F))\, d[G(x) - F(x)] = O(\|G - F\|). \tag{2.2}$$

Below, in proving (2.1) and (2.2) for $\|\cdot\| = \|\cdot\|_\infty$, we will use

$$|\int H\, dK| \leqslant \|H\|_V \cdot \|K\|_\infty, \tag{2.3}$$

which is easily checked, using integration by parts, for H continuous and of finite variation $\|H\|_V$ and K right-continuous, bounded, and 0 at $\pm\infty$.

THEOREM 2.2. *Consider an M-functional T* w.r.t. ψ. *Let F be given and put* $t_0 = T(F)$. *Suppose that* $\lambda_F'(t_0) \neq 0$ *and that* $\sigma^2(T, F) = \int\psi^2(x, t_0)\, dF(x)/[\lambda_F'(t_0)]^2$ *is finite and positive. Suppose that* $\psi(x, t)$ *is continuous in x and satisfies*

$$\lim_{t\to t_0}\|\psi(\cdot, t) - \psi(\cdot, t_0)\|_V = 0. \tag{2.4}$$

Let $T_n = T(F_n)$ *satisfy*

$$T_n \to_p t_0. \tag{2.5}$$

Then

$$n^{\frac{1}{2}}(T_n - t_0) \to_d N(0, \sigma^2(T, F)). \tag{CLT}$$

If the convergence in (2.5) *is* w.p. 1, *then*

$$\limsup_{n\to\infty} \frac{n^{\frac{1}{2}}(T_n - t_0)}{(2\sigma^2(T, F)\log\log n)^{\frac{1}{2}}} = 1 \text{ w.p. } 1. \tag{LIL}$$

PROOF. We apply Lemma 1.1. Note that

$$\mu(T, F) = -E\{\psi(X, t_0)/\lambda_F'(t_0)\} = -\lambda_F(t_0)/\lambda_F'(t_0) = 0.$$

By the inequality (2.3) with $K = F_n - F$ and $H = \psi(\cdot, T_n) - \psi(\cdot, t_0)$, and by (2.4) and (2.5), we obtain (2.1) with $G = F_n$. Thus the appropriate stochastic versions of $(D1)$ and $(D2)$ hold. □

THEOREM 2.3. *Consider an M-functional T* w.r.t. ψ. *Let F be given and put* $t_0 = T(F)$. *Suppose that* $\lambda_F'(t_0) \neq 0$ *and that* $T(\cdot)$ *is continuous at F* w.r.t. $\|\cdot\|_\infty$. *Suppose that* $\psi(x, t)$ *is continuous in x and satisfies* (2.4) *and* $\|\psi(\cdot, t_0)\|_V < \infty$. *Then* $T(\cdot)$ *is differentiable at F* w.r.t. $\|\cdot\|_\infty$, *with* $T(F; \Delta) = -\int\psi(x, t_0)\, d\Delta(x)/\lambda_F'(t_0)$.

PROOF. Similar to that of Theorem 2.2. □

EXAMPLES. Consider M-estimation of a *location* parameter, in which case the function $\psi(x, t)$ may be replaced by $\psi(x - t)$, where $\psi(\cdot)$ is now a function of one argument. Our Theorem 2.2 requires that this ψ be continuous and satisfy

$$(*) \qquad \lim_{b\to 0}\|\psi(\cdot - b) - \psi(\cdot)\|_V = 0.$$

(In checking (*), a helpful relation is $\|H\|_V = \int|H'(x)|\,dx$, for H absolutely continuous.) These conditions are satisfied by typical ψ considered in robust estimation: "least pth power" estimates $\psi(x) = |x|^{p-1}\mathrm{sgn}(x)$, $1 < p \leqslant 2$; "Hubers" $\psi(x) = \min(k, \max(-k, x))$; Hampel's 3-part "redescenders"; the "AMT" sine curves.

Theorem 2.2 exchanges condition (*) for the condition that ψ have a uniformly continuous derivative in an asymptotic normality result (Lemma 5) of Huber (1964), and provides the LIL as well.

Theorem 2.2 has a similar comparison with work of Collins (1976). He requires F to be governed by the standard normal density on an interval $(t_0 - d, t_0 + d)$ and allows F to be arbitrary elsewhere, and requires that ψ be continuous with continuous derivative and skew-symmetric and vanish outside an interval $[-c, c]$, $c < d$. His estimator T_n is the *Newton method* solution of $\lambda_{F_n}(t) = 0$ starting with the sample median. Collins proves that $T_n \to_p t_0$ and establishes the CLT. Our theorem, under the additional assumption of uniformly continuous ψ', extends to one-step Newton method estimators based on consistent starters.

Theorem 2.2 can also be compared with results of Portnoy (1977). He assumes F symmetric and absolutely continuous with density satisfying certain regularity properties, and that ψ is bounded and has a derivative which is bounded and uniformly continuous a.s. (Lebesgue). His T_n is the solution of $\lambda_{F_n}(t) = 0$ nearest to a given consistent estimator $\tilde{T}_n$ for t_0.

3. Complements.

(i) The treatment of $T(F_n) - T(F)$ by the method of Section 1 extends in straightforward fashion to the case of possibly *dependent* observations $X_1, X_2, \cdots$. The probabilistic part of the analysis rests entirely upon the classical CLT and LIL for sums and the stochastic behavior of $\|F_n - F\|_\infty$. The various extensions of the latter results to dependent variables immediately yield corresponding extensions for $T(F_n) - T(F)$.

(ii) The differential approach of Section 1 reduces $T(F_n) - T(F)$ to a random variable of the form $T_F(F_n) \cdot T(F; F_n - F)$, where $T_F(F_n) \to 1$ in a suitable stochastic sense and $T(F; F_n - F)$ has the usual structure of a differential. This leads to a broader concept of differential. We call $T(F; \Delta)$ a *quasi-differential* w.r.t. $\|\cdot\|$ and $T_F(\cdot)$ if $T_F(\cdot)$ satisfies $(D1)$ and $T(F; \Delta)$ satisfies the definition of differential with (D) replaced by $(D2)$. Thus we may think of $T(F; F_n - F)$ as a "stochastic quasi-differential" w.r.t. $\|\cdot\|_\infty$ and $T_F(\cdot)$.

(iii) A different approach toward the CLT and LIL for M-estimates has been carried out by Carroll (1978), via Bahadur-type (see Bahadur (1966)) almost-sure representations. To establish the desired representation, Carroll requires that ψ be bounded and possess two continuous bounded derivatives piecewise on intervals, and that F be Lipschitz in neighborhoods of the endpoints of these intervals. □

Acknowledgement. The authors are very grateful to the referee for providing clarifying viewpoints which inspired a comprehensive revision of the original manuscript.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BERAN, R. (1977a). Robust location estimates. *Ann. Statist.* **5** 431–444.

BERAN, R. (1977b). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BOOS, D. D. (1977). The differential approach in statistical theory and robust inference. Ph.D. dissertation, Florida State Univ.

BOOS, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7** 955–959.

CARROLL, R. J. (1978). On almost sure expansions for M-estimates. *Ann. Statist.* **6** 314–318.

CHUNG, K. L. (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

COLLINS, J. R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.* **4** 68–85.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

FILIPPOVA, A. A. (1962). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and statistical applications. *Theor. Probability Appl.* **7** 24–57.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. California, Berkeley.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HARTMAN, P. and WINTNER, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.* **63** 169–176.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

KALLIANPUR, G. and RAO, C. R. (1955). On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā* **15** 331–342.

PORTNOY, S. L. (1977). Robust estimation in dependent situations. *Ann. Statist.* **5** 22–43.

RICHTER, H. (1974). Das Gesetz vom iterierten Logarithmus für empirische Verteilungsfunktionen im R^k. *Manuscripta Math.* **11** 291–303.

VON MISES, R. (1947). On the asymptotic distributions of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27607

FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306