Probab. Th. Rel. Fields (1989) 81, 453–471

Probability Theory and Related Fields

The Effects on Convergence of Substituting Parameter Estimates into *U*-Statistics and Other Families of Statistics

H.K. Iverson[1] and R.H. Randles[2]

[1] The University of Iowa, Iowa City, IA 52244, USA
[2] The University of Florida, Department of Statistics, Gainesville, FL 32611, USA

Summary. Substituting an estimator in a statistic will often affect its limiting distribution. Sukhatme (1958), Randles (1982), and Pierce (1982) all consider the changes, if any, in the statistic's limiting normal distribution. This paper gives conditions for a law of the iterated logarithm for *U*-statistics which have a kernel with an estimator substituted into it. It also gives conditions for both strong and weak convergence. Applications of the theory are illustrated by constructing a sequential test for scale differences with power one. The theory also produces convergence results for adaptive *M*-estimators and for cross-validation assessment statistics. In addition, it is shown how to extend LIL results to a broad class of statistics with estimators substituted into them by use of the differential. In particular, a law of the iterated logarithm is described for adaptive *L*-statistics and is illustrated by an example of de Wet and van Wyk (1979).

1. Introduction

In many hypothesis testing and estimation problems, one needs to know the effect on the distribution of substituting an estimator into a statistic. Let $X_1, \ldots, X_n$ denote a random sample and consider a statistic

$$T_n(\hat{\underset{\sim}{\lambda}}_n) = T_n(X_1, \ldots, X_n; \hat{\underset{\sim}{\lambda}}),$$

which is a function of $X_1, \ldots, X_n$ and also the estimator $\hat{\underset{\sim}{\lambda}}_n$. Here $\hat{\underset{\sim}{\lambda}}_n$ is also a function of $X_1, \ldots, X_n$ and consistently estimates the parameter $\underset{\sim}{\lambda}$. Replacing $\hat{\underset{\sim}{\lambda}}_n$ with the variable $\underset{\sim}{\gamma}$, we write

$$T_n(\underset{\sim}{\gamma}) = T_n(X_1, \ldots, X_n; \underset{\sim}{\gamma})$$

and denote its limiting mean by

$$\mu(\underset{\sim}{\gamma}) = \lim_{n\to\infty} E_{\underset{\sim}{\lambda}}[T_n(\underset{\sim}{\gamma})],$$

where the notation indicates the actual parameter value is $\underset{\sim}{\lambda}$. In many settings $T_n(\underset{\sim}{\gamma})$ is a member of a common family of statistics to which well known convergence results apply. We desire to know the effects of $\hat{\lambda}_n$ on the convergence of $T_n(\hat{\underset{\sim}{\lambda}}_n)$. In many common settings $T_n(\underset{\sim}{\gamma})$ is not differentiable in $\underset{\sim}{\gamma}$ at $\underset{\sim}{\gamma}=\underset{\sim}{\lambda}$, so a simple expansion will not be useful. Thus, more general techniques are needed.

Sukhatme (1958) demonstrates sufficient conditions for the asymptotic normality of $T_n(\hat{\underset{\sim}{\gamma}}_n)$ when $T_n(\underset{\sim}{\gamma})$ is a U-statistic with a kernel depending on a univariate location parameter γ. Randles (1982) generalizes this result for U-statistics and shows how the differential may be used to extend the asymptotic normality to other classes of statistics. Pierce (1982) gives conditions for the asymptotic normality of $T_n(\hat{\underset{\sim}{\lambda}}_n)$ when $T_n(\cdot)$ is not restricted to a particular family, but $\hat{\underset{\sim}{\lambda}}_n$ is assumed to be asymptotically efficient.

The present paper develops convergence results for $T_n(\hat{\underset{\sim}{\lambda}}_n)-\mu(\underset{\sim}{\lambda})$. It creates a law of the iterated logarithm for settings in which $T_n(\underset{\sim}{\gamma})$ is not necessarily differentiable at $\underset{\sim}{\gamma}=\underset{\sim}{\lambda}$, but $\mu(\underset{\sim}{\gamma})$ is differentiable. Specifically, we write

$$T_n(\hat{\underset{\sim}{\lambda}}_n)-\mu(\underset{\sim}{\lambda})=[T_n(\hat{\underset{\sim}{\lambda}}_n)-\mu(\hat{\underset{\sim}{\lambda}}_n)-T_n(\underset{\sim}{\lambda})+\mu(\underset{\sim}{\lambda})]+[T_n(\underset{\sim}{\lambda})-\mu(\underset{\sim}{\lambda})+\mu(\hat{\underset{\sim}{\lambda}}_n)-\mu(\underset{\sim}{\lambda})]$$

and show conditions under which

$$(n/\log\log n)^{\frac{1}{2}}[T_n(\hat{\underset{\sim}{\lambda}}_n)-\mu(\hat{\underset{\sim}{\lambda}}_n)-T_n(\underset{\sim}{\lambda})+\mu(\underset{\sim}{\lambda})]\xrightarrow{\text{wp}\,1}0. \tag{1.1}$$

and

$$\limsup_{n\to\infty}(n/\log\log n)^{\frac{1}{2}}[T_n(\underset{\sim}{\lambda})-\mu(\underset{\sim}{\lambda})+\mu(\hat{\underset{\sim}{\lambda}}_n)-\mu(\underset{\sim}{\lambda})]=1\ \text{wp}\,1, \tag{1.2}$$

for a suitable constant τ. The second goal of this paper is to develop weaker conditions under which either

$$T_n(\hat{\lambda}_n)\xrightarrow{\text{wp}\,1}\mu(\hat{\underset{\sim}{\lambda}})\quad\text{or}\quad T_n(\hat{\underset{\sim}{\lambda}}_n)\xrightarrow{p}\mu(\underset{\sim}{\lambda}),$$

again without assuming $T_n(\underset{\sim}{\gamma})$ is differentiable at $\underset{\sim}{\gamma}=\underset{\sim}{\lambda}$.

Section 2 of this paper establishes conditions for a law of the iterated logarithm, for weak convergence, and for strong covergence when $T_n(\hat{\lambda}_n)$ is a U-statistic with an estimator substituted into its kernel. The results are applied to construct a test with power one for detecting differences in the scales of two populations with common form but unknown locations. The test is constructed in the same vein as tests with power one by Darling and Robbins (1968).

Section 3 illustrates further applications of the results of Sect. 2 to produce convergence results for two important classes of statistics: adaptive M-estimators and leave-one-out cross validation assessment statistics. Section 4 shows how the LIL results of Sect. 2 can be extended to cover other broad classes of statistics via use of the differential. Such extensions are illustrated by considering the class of adaptive L-statistics and a specific member of that class proposed by De Wet and Van Wyk (1979).

2. *U*-statistics with an Estimated Parameter

In this section we consider statistics of the form

$$U_n(\hat{\lambda}_n)=\binom{n}{m}^{-1}\sum_{\alpha\in A} h(X_{\alpha_1}, \ldots, X_{\alpha_m}; \hat{\lambda}_n),$$

where $\hat{\lambda}_n$ is an estimator of λ. The set A is the collection of all subsets of size n from the integers $\{1, \ldots, n\}$ and $h(\cdot\,;\cdot)$ is assumed to be symmetric in its first m arguments. Thus $U_n(\hat{\lambda}_n)$ would be a U-statistic, except that it has the estimator $\hat{\lambda}_n$ substituted into it. (While we do not use boldface notation, both $\hat{\lambda}_n$ and X_i can be vector valued.) Replacing $\hat{\lambda}_n$ with the variable γ, we write

$$U_n(\gamma)=\binom{n}{m}^{-1}\sum_{\alpha\in A} h(X_{\alpha_1}, \ldots, X_{\alpha_m}; \gamma)$$

and

$$\theta(\gamma)=E_\lambda[h(X_1, \ldots, X_m; \gamma)]$$

where the subscript λ indicates the actual parameter value. If λ is known, $U_n(\lambda)$ is a U-statistic estimator of $\theta(\lambda)$. But when λ is unknown, statistics of the form $U_n(\hat{\lambda}_n)$ are often used to estimate $\theta(\lambda)$. If the kernel $h(\cdot\,;\gamma)$ is differentiable as a function of γ at $\gamma=\lambda$, then an expansion of $U_n(\hat{\lambda}_n)$ around $U(\lambda)$ can be used to study the asymptotic properties of $U_n(\hat{\lambda}_n)$. But in many common examples $h(\cdot\,;\gamma)$ does not have this differentiability and hence more general results are needed.

The following theorem establishes (1.1) for $U_n(\hat{\lambda}_n)$ and is the main result needed to prove a law of the iterated logarithm for a U-statistic with an estimated parameter. We shall also use it to establish results for other classes of statistics in later sections.

Theorem 2.1. *Suppose there is a neighborhood $K(\lambda)$ of λ and a constant $k_1>0$ such that, if $\gamma\in K(\lambda)$ and $D(\gamma, d)$ is a sphere centered at γ with radius d such that $D(\gamma, d)\subset K(\lambda)$, then:*

$$E[\sup_{\gamma'\in D(\gamma,d)} |h(X_1, \ldots, X_m; \gamma')-h(X_1, \ldots, X_m; \gamma)|^k]\leqq k_1 d,$$

for $k=1, 2, 3, 4$ and 6.

Suppose also that for some bounded sphere C, $P[(n/\log\log n)^{\frac{1}{2}}(\hat{\lambda}_n-\lambda)\in C$ for all but a finite number of $n]=1$. Then

$$(n/\log\log n)^{\frac{1}{2}}[U_n(\hat{\lambda}_n)-U_n(\lambda)-\theta(\hat{\lambda}_n)+\theta(\lambda)]\xrightarrow{\text{wp}\,1}0.$$

Proof. For $\varepsilon > 0$ and C as stated in the conditions, let $\{C_i\}_{i=1}^{I}$ be a finite collection of spheres such that $\mathrm{rad}(C_i) < (\varepsilon/8k_1)$ and C_i is centered at s_i. The result will follow provided the terms

$$\begin{aligned} Q_{n1i}(s)+Q_{n2i} \equiv \binom{n}{m}^{-1} \sum_{\alpha\in A} [\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s) \\ -\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s_i)] \\ +\binom{n}{m}^{-1} \sum_{\alpha\in A} [\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s_i) \\ -\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda)], \end{aligned}$$

with $\tilde{h}(\cdot\,;\gamma) = h(\cdot\,;\gamma) - \theta(\gamma)$, satisfy

$$P\left((n/\log\log n)^{\frac{1}{2}} \sup_{s\in C_i} |Q_{n1i}(s)| > \frac{\varepsilon}{2} \text{ for inf. many } n\right) = 0$$

and

$$P\left((n/\log\log n)^{\frac{1}{2}} |Q_{n2i}| > \frac{\varepsilon}{2} \text{ for inf. many } n\right) = 0,$$

for $i=1, \ldots, I$. The term $(n/\log\log n)^{\frac{1}{2}} \sup_{s\in C_i} |Q_{n1i}(s)|$ is bounded above by $\frac{\varepsilon}{4}$ $+(n/\log\log n)^{\frac{1}{2}} P_{n1i}$, where

$$\begin{aligned} P_{n1i} = \binom{n}{m}^{-1} \sum_{\alpha\in A} (n/\log\log n)^{\frac{1}{2}} [\sup_{s\in C_i} |\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s) \\ -\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s_i)| \\ -E[\sup_{s\in C_i} |\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s) \\ -\tilde{h}(X_{\alpha_1}, \ldots, X_{\alpha_m}; \lambda+(\log\log n/n)^{\frac{1}{2}} s_i)|]]. \end{aligned}$$

The terms $(n/\log\log n)^{\frac{1}{2}} P_{n1i}$ and $(n/\log\log n)^{\frac{1}{2}} Q_{n2i}$ are of the form

$$U_n = \binom{n}{m}^{-1} \sum_{\alpha\in A} k_n(X_{\alpha_1}, \ldots, X_{\alpha_m}),$$

which would be a U-statistic except for the fact that the function which plays the role of the kernel, $k_n(\cdot)$, depends upon n. For U_n it is possible to define a projection

$$\hat{U}_n = \sum_{i=1}^{n} E(U_n | X_i) - (n-1)\, E[k_n(X_1, \ldots, X_m)],$$

and show, if $E[k_n^2] = O(n^{\frac{1}{2}})$, $n \to \infty$, that $U_n - \hat{U}_n \xrightarrow{\text{wp}1} 0$. The technique is the same as that employed in the development of asymptotic theory for U-statistics (see, for example, Randles and Wolfe 1979, p. 77). Using this method to reduce to the approximating quantities $\hat{U}_{nji}$, $j=1, 2$, it is then straight-forward to show that each $\hat{U}_{nji}$ satisfies

$$\sum_{n=1}^{\infty} E[\hat{U}_{nji}]^6 < \infty,$$

and therefore each term converges with probability one to zero. □

In particular situations where the kernel value differences are bounded in a neighborhood of λ, the following remark simplifies the verification of the conditions of Theorem 2.1.

Remark 2.2. Suppose there exists $M > 0$ such that

$$(1) \quad |h(x_1, \ldots, x_m; \gamma) - h(x_1, \ldots, x_m; \lambda)| \leqq M$$

for every $x_1, \ldots, x_m$ and all γ in some neighborhood of λ. Then if (1) of Theorem 2.1 holds for $k=1$, it holds for $k=2, 3, 4$, and 6.

The following utilizes Theorem 2.1 to give a law of the iterated logarithm for a U-statistic with an estimated parameter.

Theorem 2.3. *Let λ be p-variate, and denote the* ith *coordinate by λ_i, $i=1, \ldots, p$. Assume that $\theta(\gamma) = E_\lambda[h(X_1, \ldots, X_m; \gamma)]$ is differentiable at $\gamma = \lambda$, and denote the vector of partial derivatives of $\theta(\cdot)$ evaluated at the point λ by $\Delta\,\theta(\lambda)$, with* ith *component $\Delta_i\,\theta(\lambda)$, $i=1, \ldots, p$. Assume that for $i=1, \ldots, p$ there exists $k_i(x)$ such that $E[k_i(X_1)] = 0$, $\sigma_i^2 = E[k_i^2(X_1)] < \infty$, and*

$$(n/\log\log n)^{\frac{1}{2}} \left[\hat{\lambda}_{ni} - \lambda_i - \frac{1}{n} \sum_{j=1}^{n} k_i(X_j)\right] \xrightarrow{\text{wp}\,1} 0.$$

Assume that condition (1) of Theorem 2.1 holds and also that $E[h^2(X_1, \ldots, X_m; \lambda)] < \infty$. Then

$$\limsup_{n\to\infty} [(n/2\,\sigma^2 \log\log n)^{\frac{1}{2}} (U_n(\hat{\lambda}_n) - \theta(\lambda))] = 1$$

with probability one, where

$$\sigma^2 = \operatorname{var}[m\,\tilde{h}_1(X_1) + k_1(X_1)\,\Delta_1\,\theta(\lambda) + \ldots + k_p(X_1)\,\Delta_p\,\theta(\lambda)],$$

and

$$\tilde{h}_1(x) = E[h(x, X_2, \ldots, X_m; \lambda)] - \theta(\lambda).$$

Proof. We note that the assumptions imply that there is a bounded sphere C such that

$$P((n/\log\log n)^{\frac{1}{2}}(\hat{\lambda}_n - \lambda) \text{ is in } C \text{ for all but a finite number of } n \text{ values}) = 1.$$

Letting $\hat{U}_n(\lambda)=\frac{m}{n}\sum_{i=1}^{n}\tilde{h}_1(X_i)+\theta(\lambda)$ denote the projection of $U_n(\lambda)$, we write

$$(n/2\sigma^2\log\log n)^{\frac{1}{2}}[U_n(\hat{\lambda}_n)-\theta(\lambda)]=(n/2\sigma^2\log\log n)^{\frac{1}{2}}$$

$$\cdot\left[\frac{1}{n}\sum_{i=1}^{n}(m\tilde{h}_1(X_i)+k_1(X_i)\,\Delta_1\,\theta(\lambda)+\ldots+k_p(X_i)\,\Delta_p\,\theta(\lambda))\right] \quad (2.4)$$

$$+(n/2\sigma^2\log\log n)^{\frac{1}{2}}\left[\sum_{i=1}^{p}\Delta_i\,\theta(\lambda)\right]\left[\hat{\lambda}_{ni}-\lambda-\frac{1}{n}\sum_{j=1}^{n}k_i(X_j)\right] \quad (2.5)$$

$$+(n/2\sigma^2\log\log n)^{\frac{1}{2}}[U_n(\lambda)-\hat{U}_n(\lambda)] \quad (2.6)$$

$$+(n/2\sigma^2\log\log n)^{\frac{1}{2}}[U_n(\hat{\lambda}_n)-\theta(\hat{\lambda}_n)-U_n(\lambda)+\theta(\lambda)] \quad (2.7)$$

$$+(n/2\sigma^2\log\log n)^{\frac{1}{2}}[\theta(\hat{\lambda}_n)-\theta(\lambda)-(\hat{\lambda}_n-\lambda)'\,\Delta\,\theta(\lambda)]. \quad (2.8)$$

The term 2.4 will satisfy the usual law of the iterated logarithm, for the value of σ^2 given in the statement of the theorem, and hence will have a lim sup of one with probability one. The term (2.5) will converge with probability one to zero by assumption. It is well known that (2.6) converges with probability one to zero. Terms (2.7) and (2.8) converge with probability one to zero by Theorem 2.1 and the definition of the differential, respectively. □

We now present weaker conditions which are sufficient to establish convergence in probability or with probability one of a U-statistic with an estimated parameter, when the kernel $h(\cdot,\gamma)$ is not assumed differentiable in γ.

Theorem 2.9. *Suppose that $E[|h(X_1,\ldots,X_m;\lambda)|]<\infty$, and that there is a neighborhood $K(\lambda)$ of λ, such that if $D(\lambda,d)$ is a sphere centered at λ with radius d satisfying $D(\lambda,d)\subset K(\lambda)$, then*

(i) $$\lim_{d\to 0}E[\sup_{\gamma\in D(\lambda,d)}|h(X_1,\ldots,X_m;\gamma)-h(X_1,\ldots,X_m;\lambda)|]=0.$$

Under these assumptions:

(A) *If* $\hat{\lambda}_n\xrightarrow{p}\lambda$, *then* $U_n(\hat{\lambda}_n)\xrightarrow{p}\theta(\lambda)$.

(B) *If* $\hat{\lambda}_n\xrightarrow{\text{wp}\,1}\lambda$, *then* $U_n(\hat{\lambda}_n)\xrightarrow{\text{wp}\,1}\theta(\lambda)$.

Proof. These utilize the type of techniques in Theorem 2.1, but are simpler. See Iverson (1982) for details. □

These results are quite useful in establishing asymptotic results in a variety of settings. We use Theorem 2.9 in an application below and also in the next section to establish consistency of an adaptive M-estimator, and a cross-validation assessment statistic.

Application: Scale Difference Test with Power One

Let $X_1, \ldots, X_n$ be i.i.d. with d.f. $F((x-\theta_1^*)/\eta_1)$ and $Y_1, \ldots, Y_n$ be i.i.d. with d.f. $F(y-\theta_2^*)/\eta_2)$, where $F(t)$ is absolutely continuous, strictly increasing in a neighborhood of zero and satisfies $F(t)=1-F(-t)$ for all t. The population medians, θ_i^*, and scale parameters, η_i, are unknown. For testing

$$H_0\colon \eta_1=\eta_2 \text{ vs. } H_1\colon \eta_1<\eta_2$$

(two-sided alternatives could be considered), we might use

$$V_n=n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi(|Y_j-\text{median}(Y_1, \ldots, Y_n)|-|X_i-\text{median}(X_1, \ldots, X_n)|),$$

where $\psi(t)=1, 0$ as $t>, \leqq 0$. (See, for example, Randles 1984). Let

$$U_n=\binom{n}{2}^{-1}\sum_{i<j}h((X_i, Y_i), (X_j, Y_j); \hat{\lambda}_1, \hat{\lambda}_2),$$

where

$$h(\underline{t}, \underline{s}; \gamma_1, \gamma_2)=\tfrac{1}{2}\psi(|t_2-\gamma_2|-|s_1-\gamma_1|)$$
$$+\tfrac{1}{2}\psi(|s_2-\gamma_2|-|t_1-\gamma_1|),$$

and $\hat{\lambda}_1(\hat{\lambda}_2)$ denotes the sample median of the $n\,X_i$'s(Y_i's). Using Theorem 2.9 it is easily shown that $(n/\log\log n)^{\frac{1}{2}}(V_n-U_n)\to 0$ w.p. 1.

By proceeding sequentially and applying Theorem 2.3 we can construct a test of H_0 vs. H_1 with power one. Let $\varepsilon>0$ and define

$$k_n=(1+\varepsilon)(\log\log n/3\,n)^{\frac{1}{2}}.$$

Taking an initial sample of n_0 observations from each population, we continue sequentially sampling (each time selecting one new observation from both populations) until we first find

$$V_n-\tfrac{1}{2}\geqq k_n, \tag{2.10}$$

whereupon we stop sampling and reject H_0 in favor of H_1. Define N to be the smallest $n\geqq n_0$ satisfying 2.10 and $N=+\infty$, if no such n exists. Note that V_n converges w.p. 1 to $\theta(\theta_1^*, \theta_2^*)>\frac{1}{2}$ under H_1, where

$$\theta(\gamma_1, \gamma_2)=P(|Y_1-\gamma_2|>|X_1-\gamma_1|).$$

Thus,

$$P[\text{type II error}]=P[N=\infty\,|\,H_1]$$
$$=P[V_n-\tfrac{1}{2}<k_n, \text{for all } n\,|\,H_1]=0.$$

In addition, we note

$$P(\text{type I error})=P(V_n-\tfrac{1}{2}\geqq k_n \text{ for some } n\geqq n_0\,|\,H_0)$$
$$=P((V_n-\tfrac{1}{2})(3\,n/\log\log n)^{\frac{1}{2}}\geqq(1+\varepsilon) \text{ for some } n\geqq n_0\,|\,H_0).$$

The conditions of Theorem 2.3 are easily verified for U_n where one notes that the partial derivatives of $\theta(\cdot)$ are zero and $\sigma^2=\frac{1}{6}$ when H_0 is true. The relationship between U_n and V_n shows that this LIL also applies to V_n and therefore, the type I error probability can be made arbitrarily small through selection of n_0.

3. Applications to Adaptive *M* and Cross-Validation Estimators

In this section we apply the theorems of Sect. 2 to obtain general results for two important classes of statistics: adaptive M-estimators and leave-one-out cross-validation estimators. While they illustrate the previous theory, they are also useful in their own right. We begin with M-estimators.

An M-estimator may be defined as a solution T_n of the equation $\sum_{i=1}^{n} \psi(X_i; t) = 0$, where $X_1, \ldots, X_n$ is a random sample with unknown distribution function F. We define an adaptive M-estimator as a solution $T_n(\hat{\lambda}_n)$ of the equation $\sum_{i=1}^{n} \psi(X_i; t, \hat{\lambda}_n) = 0$. The estimator $\hat{\lambda}_n$ is used to achieve some adaptation of the $\psi(\cdot)$ function to the underlying population as estimated by the data. This is an intuitively appealing procedure, since the optimal choice of $\psi(\cdot)$ function varies widely, depending on the unknown population distribution. Moberg et al. (1978) developed an adaptive M-estimator of location, which may be described within this framework. They demonstrated that it performed well when compared with other robust estimators in a Monte Carlo study. The estimator $\hat{\lambda}_n$ served to measure tailweight and skewness in the sample and then select a particular $\psi(\cdot)$ function, from a predetermined finite set, which was suitable for a population with those values. We will use Theorems 2.9 and 2.1 to develop a strong consistency result and a law of the iterated logarithm for a broad class of adaptive M-estimators $T_n(\hat{\lambda}_n)$. Define

$$\theta_F(t, \gamma) = E_F[\psi(X_1; t, \gamma)]$$

and

$$\theta_{F_n}(t, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \psi(X_i; t, \gamma).$$

Theorem 3.1. *Let $T(\lambda)$ be an isolated root of $\theta_F(t, \lambda) = 0$ and assume $\psi(x; t, \gamma)$ is monotone in t. Assume that $\psi(x; t, \gamma) = h(x, \gamma)$ satisfies the conditions of Theorem 2.9 for every t in a neighborhood of $T(\lambda)$. Then $\hat{\lambda}_n \xrightarrow{\text{wp}\,1} \lambda$ implies $T_n(\hat{\lambda}_n) \xrightarrow{\text{wp}\,1} T(\lambda)$.*

Proof. Assume $\psi(x; t, \gamma)$ is nonincreasing in t. Then $\theta_F(t, \lambda)$ and $\theta_{F_n}(t, \hat{\lambda}_n)$ are nonincreasing in t, and for any $\varepsilon > 0$, $\theta_F(T(\lambda) + \varepsilon, \lambda) < 0 < \theta_F(T(\lambda) - \varepsilon, \lambda)$. However, by Theorem 2.9

$$\theta_{F_n}(T(\lambda) + \varepsilon, \hat{\lambda}_n) \xrightarrow{\text{wp}\,1} \theta_F(T(\lambda) + \varepsilon, \lambda)$$

and

$$\theta_{F_n}(T(\lambda) - \varepsilon, \hat{\lambda}_n) \xrightarrow{\text{wp}\,1} \theta_F(T(\lambda) - \varepsilon, \lambda). \quad \square$$

Theorem 3.2. *Assume that each component of the vector $\hat{\lambda}$, namely, $\hat{\lambda}_{ni}$, may be approximated by a sum $\frac{1}{n}\sum_{j=1}^{n} k_i(X_j)$ as in Theorem 2.3. Assume that under the conditions of Theorem 2.1*

$$E[\sup_{(t',\gamma')\in D((t,\gamma),d)} |\psi(X_1;t',\gamma')-\psi(X_1;t,\gamma)|^k]\leqq K_1 d, \qquad k=1,2,3,4, \text{and } 6.$$

Let the partial derivatives $\Delta_i\theta_F(t,\gamma)$, $i=1,\dots,p+1$, exist and be continuous in a neighborhood of $(T(\lambda),\lambda)$, and assume $\Delta_1\theta_F(T(\lambda),\lambda)\neq 0$. Assume that $\theta_F(T(\lambda),\lambda)=0$, $E[\psi(X_1;T(\lambda),\lambda)]^2<\infty$, and that $\psi(x;t,\gamma)$ is monotone in t. Let $T_n(\hat{\lambda}_n)$ be a solution of $\theta_{F_n}(t,\hat{\lambda}_n)=0$ such that $\hat{\lambda}_n\xrightarrow{\text{wp}\,1}\lambda$. Then

$$\limsup_{n\to\infty}[(n/2\sigma^2\log\log n)^{\frac{1}{2}}(T_n(\hat{\lambda}_n)-T(\lambda))]=1$$

with probability one, where

$$\begin{aligned}\sigma^2=\operatorname{Var}\{[-1/\Delta_1\theta_F(T(\lambda),\lambda)][\psi(X_1;T(\lambda),\lambda)\\ +\Delta_2\theta_F(T(\lambda),\lambda)\,k_1(X_1)+\dots+\Delta_{p+1}\theta_F(T(\lambda),\lambda)\,k_p(X_1)]\}.\end{aligned}$$

Proof. Let

$$\begin{aligned} h(t)&=[\theta_F(t,\lambda)-\theta_F(T(\lambda),\lambda)]/(t-T(\lambda)), && t\neq T(\lambda),\\ &=\Delta_1\theta_F(T(\lambda),\lambda), && t=T(\lambda),\\ Z_n&=\Delta_1\theta_F(T(\lambda),\lambda)/h(T_n(\hat{\lambda}_n)), \end{aligned}$$

and

$$\begin{aligned}T_n^*=\Bigg[&\frac{1}{n}\sum_{i=1}^{n}\psi(X_i;T(\lambda),\lambda)+\Delta_2\theta_F(T(\lambda),\lambda)\frac{1}{n}\sum_{i=1}^{n}k_1(X_i)\\ &+\dots+\Delta_{p+1}\theta_F(T(\lambda),\lambda)\frac{1}{n}\sum_{i=1}^{n}k_p(X_i)\Bigg]\cdot 1/(-\Delta_1\theta_F(T(\lambda),\lambda)).\end{aligned}$$

We will show

$$(n/\log\log n)^{\frac{1}{2}}\,k(T_n(\hat{\lambda}_n))[T_n(\hat{\lambda}_n)-T(\lambda)-Z_n T_n^*]\xrightarrow{\text{wp}\,1}0.$$

The above term is bounded by

$$\begin{aligned}&(n/\log\log n)^{\frac{1}{2}}\,|\theta_{F_n}(T(\lambda),\lambda)+\theta_F(T_n(\hat{\lambda}_n),\hat{\lambda}_n)|\\ &+(n/\log\log n)^{\frac{1}{2}}\Bigg|\theta_F(T(\hat{\lambda}_n),\lambda)-\theta_F(T_n(\hat{\lambda}_n),\hat{\lambda}_n)\\ &+\Delta_2\theta_F(T(\lambda),\lambda)\frac{1}{n}\sum_{i=1}^{n}k_1(X_i)+\dots+\Delta_{p+1}\theta_F(T(\lambda),\lambda)\frac{1}{n}\sum_{i=1}^{n}k_p(X_i)\Bigg|.\end{aligned}$$

The second term can be shown to converge with probability one to zero using the multivariate mean value theorem and the assumptions. To show the first term converges, we will use an extension of Lemma 1 of Ghosh (1971) which states that if V_n and W_n are random variables such that $|W_n| \leqq C$ for all n sufficiently large with probability one, for some $C>0$, and for any $\varepsilon>0$, and any y,

$$\begin{aligned} &P(V_n < y-\varepsilon,\ W_n \geqq y, \text{infinitely often}) \\ &\quad = P(V_n > y+\varepsilon,\ W_n \leqq y, \text{infinitely often}) = 0 \end{aligned}$$

then $W_n - V_n \xrightarrow{\text{wp}\,1} 0$. Assume $\psi(x;t,\gamma)$ is nonincreasing in t. For any fixed z, there is, by the implicit function theorem, a neighborhood N of λ on which functions $t_n(\gamma)$ and $t_0(\gamma)$ are uniquely defined for large enough n by the equations

$$z(\log\log n/n)^{\frac{1}{2}} = \theta_F(t_n(\gamma),\gamma) \quad \text{and} \quad 0 = \theta_F(t_0(\gamma),\gamma).$$

These functions have continuous partial derivatives. If $A_n = \{\omega: \hat{\lambda}_n \notin N\}$ then $P(A_n$ occurs infinitely often$)=0$. Define $t_n(\hat{\lambda}_n) = t_0(\hat{\lambda}_n) = T(\lambda)$ for $\hat{\lambda}_n \notin N$. Then, for any z and large enough n,

$$\begin{aligned} &\{\omega: -(n/\log\log n)^{\frac{1}{2}}\,\theta_F(T_n(\hat{\lambda}_n),\hat{\lambda}_n) < z\} \\ &\quad \subseteq \{\omega: -(n/\log\log n)^{\frac{1}{2}}\,\theta_F(T_n(\hat{\lambda}_n),\hat{\lambda}_n) < -(n/\log\log n)^{\frac{1}{2}}\,\theta_F(t_n(\hat{\lambda}_n),\hat{\lambda}_n)\} \cup A_n \\ &\quad \subseteq \{\omega: T_n(\hat{\lambda}_n) < t_n(\hat{\lambda}_n)\} \cup A_n \\ &\quad \subseteq \left\{\omega: (n/\log\log n)^{\frac{1}{2}}\,\frac{1}{n}\sum_{i=1}^{n}\psi(X_i;t_n(\hat{\lambda}_n),\hat{\lambda}_n) \leqq 0\right\} \cup A_n \\ &\quad = \left\{\omega: (n/\log\log n)^{\frac{1}{2}}\,\frac{1}{n}\sum_{i=1}^{n}[\psi(X_i;t_n(\hat{\lambda}_n),\hat{\lambda}_n) - \theta_F(t_n(\hat{\lambda}_n),\hat{\lambda}_n)] \leqq z\right\} \cup A_n. \end{aligned}$$

The result will therefore follow if we show

$$\begin{aligned} (n/\log\log n)^{\frac{1}{2}} &\left[\frac{1}{n}\sum_{i=1}^{n}\left[\psi(X_i;t_n(\hat{\lambda}_n),\hat{\lambda}_n) - \theta_F(t_n(\hat{\lambda}_n),\hat{\lambda}_n)\right]\right. \\ &\left. -\frac{1}{n}\sum_{i=1}^{n}\psi(X_i;T(\lambda),\lambda)\right] \xrightarrow{\text{wp}\,1} 0. \end{aligned} \tag{3.3}$$

The fact that $t_n(\gamma)$ converges uniformly over a neighborhood of λ to $t_0(\gamma)$ as $n\to\infty$, the differentiability of $\lambda_F(t,\gamma)$, and the assumptions concerning $\hat{\lambda}_n$ can be used to show there is a bounded sphere C such that $P[(n/\log\log n)^{\frac{1}{2}}(t_n(\hat{\lambda}_n) - T(\lambda),\ \hat{\lambda}_n - \lambda) \in C$ for all but a finite number of $n] = 1$. This fact and the assumptions imply that (3.3) holds by Theorem 2.1. □

We will demonstrate an application of Theorem 3.2 in Sect. 5.

We now examine a class of statistics which are members of what Stone (1974) termed cross-validation assessment statistics. They are used to estimate the success which a model fitted with all current data will have in fitting future data. Specifically, we examine the class of leave-one-out estimators of performance. Lachenbruch (1967) proposed an estimator in this class to estimate the probability of correct selection in discriminant analysis. The Lachenbruch estimator is less optimistically biased than the resubstitution estimator. Eastment and Krazanowski (1982) and Butler and Rothman (1980) have used similar leave-one-out estimators of performance in principle component analysis and in prediction intervals, respectively. Here we apply theorems from Sect. 2 to establish a law of the iterated logarithm and consistency results for leave-one-out estimators. They are specifically applicable to the Lachenbruch estimator, but details are not included. Broader results for cross-validation assessment statistics, including asymptotic normality, are proved in Iverson and Randles (1986).

Let $X_1, \ldots, X_n$ denote the observed data and $\hat{\lambda}_n$ be the vector of estimated parameters for the fitted model. Let $\hat{\lambda}_{n(i)}$ denote the estimated parameters based on all the observations except X_i. The performance of the fitted model for the i^{th} data point is assessed via

$$h(X_i, \hat{\lambda}_{n(i)}).$$

The results are averaged over all observations to create a leave-one-out performance assessment estimator of the form

$$U_n(\hat{\lambda}_{n(\,)}) = n^{-1} \sum_{i=1}^{n} h(X_i, \hat{\lambda}_{n(i)}). \tag{3.4}$$

This takes the form of a U-statistic of degree one with estimators substituted into the kernel. These statistics differ from those in Sect. 2 because a slightly different estimator is substituted into each term in the sum.

For example, consider the simple regression model

$$Y_i = \beta C_i + \varepsilon_i,$$

where the ε_i's are i.i.d. and the C_i's are known constants. To measure the performance of an estimator $\hat{\beta}_n$ we might use

$$\begin{aligned} & n^{-1} \sum_{i=1}^{n} |Y_i - \hat{\beta}_{n(i)} C_i| \\ & = n^{-1} \sum_{i=1}^{n} |\varepsilon_i + (\beta - \hat{\beta}_{n(i)}) C_i| \\ & = n^{-1} \sum_{i=1}^{n} h((\varepsilon_i, C_i); \beta - \hat{\beta}_{n(i)}) \end{aligned}$$

with $X_i = (\varepsilon_i, C_i)$ and $h(x; \gamma) = |\varepsilon + \gamma c|$.

As the above example illustrates, we will want to assume that $X_1, \ldots, X_n$ are independent but not necessarily identically distributed and that $h(\cdot)$ is not necessarily differentiable in γ. Define

$$\theta_i(\gamma) = E[h(X_i; \gamma)] \tag{3.5}$$

and

$$\bar{\theta}_n(\gamma) = n^{-1} \sum_{i=1}^{n} \theta_i(\gamma). \tag{3.6}$$

Write $\lambda_t(\hat{\lambda}_{nt}, \hat{\lambda}_{n(i)t})$ for the t^{th} component of the p-vector $\lambda(\hat{\lambda}_n, \hat{\lambda}_{n(i)})$, respectively. We assume that

$$\begin{aligned}&\theta_i(\gamma) \text{ is differentiable at } \gamma = \lambda \text{ for all } i\\ &\text{and that the differential is achieved uniformly in } i.\end{aligned} \tag{3.7}$$

If the vector of partial derivatives of $\theta_i(\gamma)$ at $\gamma = \lambda$ is denoted by $\Delta\, \theta_i(\gamma)$, we define

$$\Delta\, \bar{\theta}_n(\lambda) = n^{-1} \sum_{i=1}^{n} \Delta\, \theta_i(\lambda) \tag{3.8}$$

and assume that

$$\Delta\, \bar{\theta}_n(\lambda) \to \Delta\, \theta(\lambda). \tag{3.9}$$

We also assume there is a bounded sphere C_* centered at the origin such that

$$P((n/\log\log n)^{\frac{1}{2}}(\hat{\lambda}_n - \lambda) \in C_* \text{ for all but a finite number of } n) = 1. \tag{3.10}$$

In addition, suppose that for some constant β, $\frac{1}{2} < \beta \leqq 1$, there is a positive integer r satisfying $2r(\beta - \frac{1}{2}) > \frac{5}{2}$, such that

$$E[(\hat{\lambda}_{n(i)t} - \hat{\lambda}_{nt})^{2r}] = O(n^{-2\beta r}) \tag{3.11}$$

uniformly in i for $t = 1, \ldots, p$. We also assume there exists a representation for $\hat{\lambda}_n$, that is, functions $\beta_{ti}(x)$, such that for $t = 1, \ldots, p$,

$$(n/\log\log n)^{\frac{1}{2}} \left[(\hat{\lambda}_{nt} - \lambda_t) - n^{-1} \sum_{i=1}^{n} \beta_{ti}(X_i) \right] \xrightarrow{\text{wp}\,1} 0. \tag{3.12}$$

In addition, suppose there is some neighborhood $K(\lambda)$ of λ and a constant $K_1 > 0$ such that if $\gamma'' \in K(\lambda)$ and $D(\gamma'', d' + d'') \subset K(\lambda)$ for d' and $d'' > 0$, then for every i

$$E[\sup_{\gamma' \in D(\gamma'', d'')} \sup_{\gamma \in D(\gamma', d')} |h(X_i; \gamma') - h(X_i; \gamma)|^k] \leqq K_1(d' + d'') \tag{3.13}$$

for $k = 1, 2, 3, 4$, and 6.

Theorem 3.14. *Suppose that* (3.7)–(3.13) *hold. In addition assume that*

$$n^{-1}\sum_{i=1}^{n} k_{*i}(X_i)$$

satisfies a law of the iterated logarithm (e.g., Theorem 1.10C in Serfling 1980, p. 36) *where*

$$k_{*i}(x)=h(x;\lambda)-E[h(X_i;\lambda)]+\sum_{t=1}^{p}\beta_{ti}(x)\,\Delta_t\,\theta(\lambda), \tag{3.15}$$

and assume that for some $0<m_*<M_*$,

$$m_*\leqq \operatorname{Var}[k_{*i}(X_i)]\leqq M_*$$

for every i. Then

$$\limsup_{n\to\infty}\left(\left(\frac{n^2}{2B_{*n}^2\log\log B_{*n}}\right)^{\frac{1}{2}}\{U_n(\hat\lambda_{n(\,)})-\bar\theta_n(\lambda)\}\right)=1$$

with probability one, where

$$B_{*n}^2=\sum_{i=1}^{n}\operatorname{Var}(k_{*i}(X_i)).$$

Proof. A straightforward adaption of the proof of Theorem 2.3 to the case where the observations are independent but not identically distributed, and the kernel is of degree one, establishes that

$$P\left(\limsup_{n\to\infty}\left(\frac{n^2}{2B_{*n}^2\log\log B_{*n}}\right)^{\frac{1}{2}}\{U_n(\hat\lambda_n)-\bar\theta_n(\lambda)\}=1\right)=1.$$

Let C denote a bounded sphere centered at the origin. With r and β as in (3.11), let $0<\beta_1<(\beta-\frac{1}{2})$ satisfy $2r(\beta-\frac{1}{2}-\beta_1)>\frac{5}{2}$. Then by means of the Borel-Cantelli lemma we see that

$$P(n^{\frac{1}{2}+\beta_1}(\hat\lambda_{n(i)}-\hat\lambda_n)\notin C \text{ for at least one } i, \text{ for infinitely many } n)=0.$$

For $\varepsilon>0$, let $C_*\subset\bigcup_{w=1}^{W}C_w$ where each sphere C_w has a radius bounded by $\varepsilon m_*/K_1$, $w=1,\ldots,W$. Therefore,

$$\begin{aligned}
&P((n^2/2B_{*n}^2\log\log B_{*n})^{\frac{1}{2}}\,|U_n(\hat\lambda_{n(\,)})-U_n(\hat\lambda_n)|>\varepsilon \ \text{ inf often})\\
&\quad\leqq\sum_{w=1}^{W}P\Bigg((n^2/2B_{*n}^2\log\log B_{*n})^{\frac{1}{2}}\,n^{-1}\\
&\qquad\cdot\sum_{i=1}^{n}\{\sup_{s_w\in C_w}\ \sup_{s\in C}|h(X_i;\lambda+(\log\log n/n)^{\frac{1}{2}}s_w+n^{-\frac{1}{2}-\beta_1}s)\\
&\qquad-h(X_i;\lambda+(\log\log n/n)^{\frac{1}{2}}s_w)|\}>\varepsilon \text{ infinitely often}\Bigg).
\end{aligned} \tag{3.16}$$

For $w=1, \ldots, W$, we write

$$\begin{aligned}&(n^2/2B_{*n}^2 \log\log B_{*n})^{\frac{1}{2}} n^{-1}\\ &\qquad \cdot \sum_{i=1}^{n} \{ \sup_{s_w \in C_w} \sup_{s \in C} | h(X_i; \lambda + (\log\log n/n)^{\frac{1}{2}} s_w + n^{-\frac{1}{2}-\beta_1} s)\\ &\qquad -h(X_i; \lambda + (\log\log n/n)^{\frac{1}{2}} s_w)|\}\\ &\quad \equiv U_{nw} = U_{nw} - E[U_{nw}] + E[U_{nw}].\end{aligned}$$

By (3.13), for each w, $E[U_{nw}] < \frac{\varepsilon}{2}$, and the sixth power of $U_{nw} - E[U_{nw}]$ is $O(n^{-3/2})$, and therefore each of the summands in (3.16) is zero. □

Remark. It can be shown that (3.11) will hold with $\beta = 1$ if the $\hat{\lambda}_{nt}$ are U-statistics such that $E[h^{2r}(X_1, \ldots, X_m)] < \infty$. It can also be shown that if $\hat{\lambda}_n = \bar{Y}_n$, a mean of i.i.d. random variables, the result of Theorem 3.14 will still hold if the assumption that the c.d.f. of Y_1 has a finite absolute moment of order $\beta > 2$ replaces the assumption that $E[(\bar{Y}_{n(i)} - \bar{Y}_n)^{2r}] = O(n^{-2r})$ for some integer $r > 0$, such that $r > \frac{5}{2}$.

Applying a slight extension of Theorem 2.9 also yields the following consistency result for leave-one-out assessment statistics.

Theorem 3.18. *Suppose there is an* $M > 0$ *such that*

$$E[\{h(X_i; \lambda)\}^2] \leqq M$$

for all i. Also, assume there is a neighborhood, $K(\lambda)$, *of* λ *and a constant* $M_1 > 0$ *such that for any sphere* $D(\lambda, d)$ *centered at* λ *with radius d, satisfying* $D(\lambda, d) \subset K(\lambda)$, *and any sphere* $D(\gamma, d')$ *with center* $\gamma \in D(\lambda, d)$ *and radius* d', *satisfying* $D(\gamma, d') \subset K(\lambda)$,

$$E(\sup_{\gamma \in D(\lambda, d')} \sup_{\gamma' \in D(\gamma, d')} |h(X_i; \gamma') - h(X_i; \gamma)|) \to 0$$

uniformly in i as d' and d go to zero and

$$E(\sup_{\gamma \in D(\lambda, d)} \sup_{\gamma' \in D(\gamma, d')} |h(X_i; \gamma') - h(X_i; \gamma)|^2) \leqq M_1$$

for every i.

(A) *Suppose for some constant* β, $\frac{1}{2} \leqq \beta \leqq 1$, *and some positive integer r such that* $2\beta r > 1$, *that uniformly in* $i = 1, \ldots, n$,

$$E[(\hat{\lambda}_{n(i)t} - \hat{\lambda}_{nt})^{2r}] = O(n^{-2\beta r}) \tag{3.19}$$

for $t = 1, \ldots, p$. *Then* $\hat{\lambda}_n \xrightarrow{p} \lambda$ *implies*

$$U_n(\hat{\lambda}_{n(\,)}) - \bar{\mu}_n(\lambda) \xrightarrow{p} 0.$$

(B) *If β and r may be chosen as in* (A) *except that $2\beta r \geqq \frac{5}{2}$, such that* (3.19) *holds, then $\hat{\lambda}_n \xrightarrow{\text{wp}\,1} \lambda$ implies*

$$U_n(\hat{\lambda}_{n(\,)}) - \bar{\mu}_n(\lambda) \xrightarrow{\text{wp}\,1} 0.$$

Theorems 3.14 and 3.18 apply directly to Lachenbruch's (1967) leave-one-out estimators of the probability of misclassification in discriminant analysis. Details are found in Iverson (1982). Other applications may be found in Iverson and Randles (1986).

4. Differentiable Statistical Functions

In this section we use the results of Sect. 2 to develop a general method for establishing laws of the iterated logarithm for differentiable statistical functions which depend on an estimated parameter. The use of the differential as an extension device is in the spirit of work by Boos (1977, 1979), Boos and Serfling (1980), and Serfling (1980). Letting $T(F, \gamma)$ denote a real-valued functional defined on $\mathscr{F} \times R^p$, where $\mathscr{F}$ denotes a convex class of distribution functions large enough to include all discrete distributions, we seek to establish conditions for a law of the iterated logarithm to hold for $T(F_n, \hat{\lambda}_n)$, when it may not be assumed that $T(F, \gamma)$ is differentiably smooth in γ. Here F_n is the empirical d.f. of a random sample $X_1, \ldots, X_n$ and $\hat{\lambda}_n$ estimates λ in the sense that $(n/\log\log n)^{\frac{1}{2}}(\hat{\lambda}_n - \lambda)$ is bounded with probability one.

Let $\mathscr{D}$ denote the linear space generated by differences $G-F$ of members of $\mathscr{F}$ and let $\|\cdot\|$ represent a norm on the space $\mathscr{D}$. Suppose there is a function $T(F, \gamma; \Delta)$ defined on $\Delta \in \mathscr{D}$ and linear in Δ, such that

$$T(G, \gamma) - T(F, \gamma) - T(F, \gamma; G-F) = o(\|G-F\|),$$

as $\|G-F\| \to 0$. Then $T(F, \gamma; \Delta)$ is called the differential of the functional T at (F, γ) with respect to $\|\cdot\|$. We say that T is uniformly differentiable over γ in $N(\lambda)$, where $N(\lambda)$ is some neighborhood of λ, if

$$\sup_{\gamma \in N(\lambda)} |T(G, \gamma) - T(F, \gamma) - T(F, \gamma; G-F)| = o(\|G-F\|), \tag{4.1}$$

as $\|G-F\| \to 0$. Letting

$$\mu(\gamma) = T(F, \gamma),$$

we use the decomposition

$$\begin{aligned} &T(F_n, \hat{\lambda}_n) - \mu(\lambda) \\ &\quad = [T(F_n, \hat{\lambda}_n) - T(F, \hat{\lambda}_n) - T(F, \hat{\lambda}_n; F_n - F)] && (4.2) \\ &\qquad + T(F_n, \hat{\lambda}_n; F_n - F) - T(F, \lambda; F_n - F) && (4.3) \\ &\qquad + T(F, \lambda; F_n - F) + \mu(\hat{\lambda}_n) - u(\lambda). && (4.4) \end{aligned}$$

We thus seek conditions under which terms (4.2) and (4.3) will be $o((\log\log n/n)^{\frac{1}{2}})$ with probability one and term (4.4) will satisfy a law of the iterated logarithm.

Letting σ_x denote the d.f. with point mass 1 at x, since the differential is linear in its third argument,

$$\begin{aligned}&T(F, \hat{\lambda}_n; F_n-F)-T(F, \lambda; F_n-F)\\&\quad=\frac{1}{n}\sum_{i=1}^{n} T(F, \hat{\lambda}_n; \sigma_{X_i}-F)-\frac{1}{n}\sum_{i=1}^{n} T(F, \lambda; \sigma_{X_i}-F)\end{aligned}$$

and is thus seen to be the difference between a U-statistic with an estimated parameter and the corresponding U-statistic with the actual parameter value. When $\theta(\gamma)=E[T(F, \gamma; \sigma_{X_i}-F)]=0$, the results of Sect. 2 apply, and can be used to show this term will be $o((\log\log n/n)^{\frac{1}{2}})$ with probability one. If $T(F, \gamma)$ is uniformly differentiable, this property may be used to establish that term (4.2) is $o((\log\log n/n)^{\frac{1}{2}})$ with probability one. We formalize this approach with the following theorem, the proof of which is straightforward.

Theorem 4.5. *Let $T(F, \gamma)$ be a differentiable statistical function, with differential $T(F, \gamma; \Delta)$, such that* (4.1) *is satisfied for some norm $\|\cdot\|$ such that $\|F_n-F\| =O((\log\log n/n)^{\frac{1}{2}})$ with probability one. Suppose that $\mu(\lambda)$ has a differential at λ, and that for $i=1, \ldots, p$ there exists $k_i(x)$ such that*

$$E[k_i(X_1)]=0, \qquad 0<\sigma_i^2=E[k_i^2(X_1)]<\infty,$$

and

$$(n/\log\log n)^{\frac{1}{2}}\left[\hat{\lambda}_{ni}-\lambda_i-\frac{1}{n}\sum_{j=1}^{n} k_i(X_j)\right]\xrightarrow{\text{wp}\,1}0.$$

If $h(x, \gamma)=T(F, \gamma; \sigma_x-F)$ satisfies (1) *of Theorem 2.1,*

$$\theta(\lambda)=E[T(F, \lambda; \sigma_{X_1}-F)]=0 \quad \textit{and} \quad E[T^2(F, \lambda; \sigma_{X_1}-F)]<\infty,$$

then

$$\limsup_{n\to\infty}(n/2\sigma^2\log\log n)^{\frac{1}{2}}[T(F, \lambda; F_n-F)+\mu(\hat{\lambda}_n)-\mu(\lambda)]=1 \text{ w p 1},$$

where

$$\sigma^2=\operatorname{var}[T(F, \lambda; \sigma_{X_1}-F)+k_1(X_1)\,\Delta_1\,\mu(\lambda)+\ldots+k_p(X_1)\,\Delta_p\,\mu(\lambda)],$$

and $(\Delta_1\,\mu(\lambda), \ldots, \Delta_p\,\mu(\lambda))$ denotes the vector of partial derivatives of $\mu(\cdot)$ evaluated at the point λ. If the differential of $\mu(\cdot)$ at λ is zero, it is only required of $\hat{\lambda}_n$ that $(\hat{\lambda}_n-\lambda)=O((\log\log n/n)^{\frac{1}{2}})$ with probability one.

We illustrate an application of Theorem 4.5 by considering L-statistics with an estimated parameter. The asymptotic normality of adaptive L-statistics has been shown by Parr (1982) and Randles (1982).

We represent an L-statistic with a parameter which must be estimated as $T(F_n, \hat{\lambda}_n)$, where

$$T(F, \gamma)=\sum_{j=1}^{m} a_j(\gamma)\, T_j(F, \gamma)$$

and each term $T_j(F, \gamma)$ is one of the following three types

$$\text{Type I: } T_j(F,\gamma)=\int_0^1 F^{-1}(u)\,J_j(u,\gamma)\,du,$$

$$\text{with} \quad J_j(u,\gamma)=\begin{cases}0 & \text{if } u<\alpha_j(\gamma)\\ 1 & \text{if } u\geqq\alpha_j(\gamma).\end{cases}$$

$$\text{Type II: } T_j(F,\gamma)=\int_0^1 F^{-1}(u)\,J_j(u,\gamma)\,du,$$

where $J_j(u, \gamma)$ is continuous in $u\in(0, 1)$.

$$\text{Type III: } T_j(F,\gamma)=F^{-1}(p_j(\gamma)).$$

For example, an adaptive symmetrically trimmed mean may be represented as

$$T(F_n,\hat{\lambda}_n)=\frac{1}{1-2\alpha(\hat{\lambda}_n)}\int_{\alpha(\hat{\lambda}_n)}^{1-\alpha(\hat{\lambda}_n)} F_n^{-1}(u)\,du, \tag{4.7}$$

where the amount to be trimmed off each end, $\alpha(\hat{\lambda}_n)$, is based on considerations such as tailweight of the population which, being unknown, must be estimated by the data. This may be described in the form (4.6) as the sum of two type I terms with $\alpha(\gamma)=\alpha_1(\gamma)=1-\alpha_2(\gamma)$ and $a_1(\gamma)=-a_2(\gamma)=(1-2\alpha(\gamma))^{-1}$. Adaptive trimmed means of this type have been proposed by de Wet and van Wyk (1979).

The factors $a_j(\lambda)$ will not affect the results obtained for terms (4.2) and (4.3) obtained by applying Theorem 4.5 to the individual components $a_j(\hat{\lambda}_n)\,T_j(F_n, \lambda_n)$, as long as

$$(n/\log\log n)^{\frac{1}{2}}\,[a_j(\hat{\lambda}_n)-a_j(\lambda)]=O(1) \text{ w p 1}.$$

However, these terms must be taken into account in showing

$$\sum_{j=1}^{m} a_j(\lambda)\,T_j(F,\lambda;F_n-F)+\sum_{j=1}^{m}[a_j(\hat{\lambda}_n)\,\mu_j(\hat{\lambda}_n)-a_j(\lambda)\,\mu_j(\lambda)]$$

satisfies the law of the iterated logarithm, if an exact value for the lim sup is desired. If individually it may be shown that each of terms (4.4) satisfy a law of the iterated logarithm, then an upper bound for the lim sup of their sum may be found.

Boos (1979) provides suitable norms and conditions on F under which $T(F,\gamma;\Delta)=-\int_{-\infty}^{\infty}\Delta(t)\,J(F(t),\gamma)\,dt$ is a differential for type I and II terms $T(F, \gamma)$. Building on these results, Randles (1982) states conditions on $J(\cdot,\cdot)\,\alpha(\cdot)$, and $F(\cdot)$ under which (4.1) will hold for type I and II terms $T(F, \gamma)$ with respect to these norms. We will not restate these conditions here. It is straightforward to show that $T(F, \gamma; \sigma_x-F)$ satisfies (1) of Theorem 2.1. Therefore Theorem 4.5 may be applied to yield a law of the iterated logarithm for $T(F_n, \hat{\lambda}_n)$.

Theorem 4.5 may also be applied to type III terms. However, less restrictive conditions under which a law of the iterated logarithm will hold can be obtained by applying Theorem 3.2. The adaptive sample percentile may be represented as an adaptive M-estimator with $Y(\cdot)$ function $\psi(x; t, \gamma) = \operatorname{sign}(x-t) - (1-2p(\gamma))$:

$$T(F_n, \hat{\lambda}_n) = F_n^{-1}(p(\hat{\lambda}_n)) = \inf\{x\colon F_n(x) \geqq p(\hat{\lambda}_n)\}$$
$$= \sup\left\{t\colon \sum_{i=1}^{n} [\operatorname{sign}(X_i - t) - (1-2p(\hat{\lambda}_n))] > 0\right\}.$$

In the notation of Sect. 3,

$$\theta_F(t, \gamma) = (2p(\gamma)-2)\, F(t) + 2p(\gamma)(1-F(t)),$$

and therefore

$$\Delta_1\, \theta_F(t, \gamma) = -2f(t) \quad \text{and} \quad \Delta_i\, \theta_F(t, \gamma) = 2\Delta_{i-1}\, p(\gamma), \qquad i = 2, \ldots, k+1,$$

assuming $p(\gamma)$ has partial derivatives $\Delta_i\, p(\gamma)$. It follows by Theorem 3.2 that the adaptive sample percentile $T(F_n, \hat{\lambda}_n)$ will satisfy a law of the iterated logarithm if $F(\cdot)$ has a continuous derivative $f(\cdot)$ in a neighborhood of $F^{-1}(p(\lambda))$ such that $f(F^{-1}(p(\lambda))) > 0$, $p(\gamma)$ is differentiable at λ, and $\hat{\lambda}_n$ may be approximated as in Theorem 2.3.

Finally, as a specific application of material in this section, we prove a law of the iterated logarithm for an adaptive trimmed mean proposed by de Wet and van Wyk (1979). In this development we use the notation and build on the work showing its asymptotic normality found in Sect. 4 of Randles (1982). We also apply the work on differentials for L-statistics proved by Boos (1979). Assume that $1.75 < \lambda < 1.95$, $\int |x|\, dF(x) < \infty$, $\int F^{-1}(u)\, J_*(u)\, du < \infty$, and $F(\cdot)$ is absolutely continuous with a density bounded away from zero in a neighborhood of $F^{-1}(\alpha(\lambda))$ and $F^{-1}(1-\alpha(\lambda))$. Suppose, in addition, that

$$\int q(F(x))\, dx < \infty$$

for some $q \in Q_2$ of Boos. Then because,

$$(n/\log\log n)^{\frac{1}{2}}\, \|F_n - F\|_q = O(1) \text{ w p 1},$$

it follows from the corollary in Sect. 4 of Boos, that $\hat{\lambda}_n$ may be approximated as in Theorem 4.5 and hence that

$$(n/\log\log n)^{\frac{1}{2}}\, [(1-\alpha(\hat{\lambda}_n))^{-1} - (1-\alpha(\lambda))^{-1}] = O(1) \text{ w p 1}.$$

It is straightforward to show (1) of Theorem 2.1 holds, using Remark 2.2. Randles shows that 4.1 holds and that $\mu(\gamma)$ is differentiable.

Applying Theorem 4.5 to each of the two type I terms whose sum is the adaptive *L*-statistic, it follows that

$$\lim_{n\to\infty}\sup (n/2\,\sigma^2 \log\log n)^{\frac{1}{2}}\,[T_n(\hat{\lambda}_n)-T_n(\lambda)] \leqq 2 \text{ w p } 1,$$

provided

$$0<\sigma^2=\int_0^1\int_0^1 (\min(s,t)-s\,t)\,J_*(s)\,J_*(t)\,dF^{-1}(s)\,dF^{-1}(t)<\infty.$$

References

1. Butler, R., Rothman, E.D.: Predictive intervals based on reuse of the Sample. J. Am. Statist. Assoc. **75**, 881–889 (1980)
2. Eastment, H.T., Krzanowski, W.J.: Cross validatory choice of the number of components from a principle component analysis. Technometrics **24**, 73–77 (1982)
3. Boos, D.D.: The differential approach in statistical theory and robust inference. Unpublished dissertation, Florida State University (1977)
4. Boos, D.D.: A differential for *L*-statistics. Ann. Statist. **7**, 955–959 (1979)
5. Boos, D.D., Serfling, R.J.: A note on differentials and the CLT and LIL for statistical functions, with applications to *M*-estimates. Ann. Statist. **8**, 618–624 (1980)
6. Darling, D.A., Robbins, H.: Some nonparametric sequential tests with power 1. Proc. Nat. Acad. Sci. **61**, 804–809 (1968)
7. De Wet, T., Van Wyk, J.W.J.: Some large sample properties of Hogg's adaptive trimmed means. South African Statist. J. **13**, 53–69 (1979)
8. Ghosh, J.K.: A new proof of the Bahadur representation of quantiles and an application. Ann. Statist. **42**, 1957–1961 (1971)
9. Iverson, H.K.: Asymptotic properties of *U*-statistics with estimated parameters. Unpublished Ph.D. thesis, Department of Statistics, University of Iowa (1982)
10. Iverson, H.K., Randles, R.H.: Large sample properties of cross-validation assessment statistics. J. Statist. Plan. Inf. **15**, 43–62 (1986)
11. Lachenbruch, P.A.: An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Ann. Statist. **10**, 462–474 (1967)
12. Moberg, T.F., Ramberg, J.S., Randles, R.H.: An adaptive *M*-estimator and its application to a selection problem. Technometrics **20**, 255–263 (1978)
13. Parr, W.C.: A note on adaptive *L*-statistics. Comm. Statist. – Theor. Meth. **11**, 1511–1518 (1982)
14. Pierce, D.A.: The asymptotic effect of substituting estimators for parameters in certain types of statistics. Ann. Statist. **10**, 475–478 (1982)
15. Randles, R.H., Wolfe, D.A.: Introduction to the theory of nonparametric statistics. New York: Wiley 1979
16. Randles, R.H.: On the asymptotic normality of statistics with estimated parameters. Ann. Statist. **10**, 462–474 (1982)
17. Randles, R.H.: On tests applied to residuals. J. Am. Statist. Assoc. **79**, 349–354 (1984)
18. Serfling, R.J.: Approximation theorems of mathematical statistics. New York: Wiley 1980
19. Stone, M.: Cross-validating choice and assessment of statistical predictions. J. Roy. Statist. Soc. B**36**, 111–147 (1974)
20. Sukhatme, B.V.: Testing the hypotheses that two populations differ only in location. Ann. Math. Statist. **29**, 60–78 (1958)

Received August 1984; in revised form September 21, 1987